IBM Research

RC 5624
(#24297)
9/15/75
Communications

29 pages

ON-LINE GENERATION OF TERMINOLOGICAL DIGESTS IN LANGUAGE TRANSLATION: AN AID IN TERMINOLOGY PROCESSING

E. O. Lippmann

IBM Thomas J. Watson Research Center
Yorktown Heights, N. Y. 10598

Yorktown Heights, New York

San Jose, California

Zurich, Switzerland

RC 5624
(#24297)
9/15/75
Communications

29 pages

ON-LINE GENERATION OF TERMINOLOGICAL DIGESTS IN LANGUAGE TRANSLATION: AN AID IN TERMINOLOGY PROCESSING

E. O. Lippmann

IBM Thomas J. Watson Research Center
Yorktown Heights, N. Y. 10598

ABSTRACT: Evaluation of technical and scientific translations dealing with complex subject matter has shown that a) the majority of errors made by the translators involve terminology, and b) the translators spend a great deal of their time searching for the correct equivalents of technical-scientific terms to be translated. This paper describes a technique of generating terminological digests speedily on terminals connected to a computer in order to overcome these impediments and aid the translator in streamlining the translation production process. A terminological digest represents the glossarial framework of a translation, a unique dictionary constructed automatically for the text to be translated. The user can produce a terminological digest by invoking the appropriate program on his terminal, entering on the keyboard the terms he wishes to have looked up. All terms entered are immediately retrieved from an up-to-date scientific-technical dictionary and provided with target language equivalents and other pertinent information. At the user's option, the dictionary entries may be presented singly, as a list in the order of entering the terms (e.g., the order in which they occur in the text to be translated), or as an alphabetically-sorted list. These lists may be displayed, typed out, or printed and saved as "minidictionaries" for a particular field.

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication elsewhere and has been issued as a Research Report for early dissemination of its contents. As a courtesy to the intended publisher, it should not be widely distributed until after the date of outside publication.

Copies may be requested from:
IBM Thomas J. Watson Research Center
Post Office Box 218
Yorktown Heights, New York 10598

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

THE PROLIFERATION OF SPECIALIZED TERMINOLOGY

As has been emphasized in a variety of works on translation [1,2,3,4,5,6], mere knowledge of the general vocabulary and the grammars of two or more languages do not necessarily enable an educated person to translate scientific, economic, technical or legal literature. To translate specialized material of this type correctly, an acquaintance with the subject matter and a command of the proper terminology is required. For the professional translator, one of the most exasperating aspects of translating modern texts on a variety of subjects is the inordinate amount of searching and learning time he has to expend to evaluate special terms employed by the authors of source language documents. It has been claimed that up to 60% of a conscientious translator's work time is consumed in tracking down proper terminological information [7].

Not only is the volume of new terminology increasing rapidly, but also the looseness of using the technical vocabulary is growing among various specialists in the field [8]. Added to the proliferation of unabbreviated specialized terminology must be the ever-enlarging usage of acronyms and initialisms, particularly in information processing, which frequently cannot be decoded even by experts in a given field without the aid of a proper dictionary [9]. The number of acronyms formally collected in the U.S. has now swelled to nearly 103,000 terms [10] and is continuing to grow [11,12,13,14]. However, even the most up-to-date printed dictionary cannot maintain a rate of speed parallel with the burgeoning growth of specialized terminology [15]. That dictionaries are obsolete is especially evident in computer systems terminology [16], which is proliferating at a rate which could perhaps be compared to the proliferation of higher-level languages in the programming field.

It is, of course, true that as the use of higher-level languages can effectively assist non-data-processing professionals in communicating with a computer in their own professional jargon [17], so special terminology can make communication for a scientist or technologist more efficient and convenient when interacting with fellow specialists [18]. And just as the development of programming languages is likely to go forward [19], one would expect a continual growth of new terminology, given the rate at which technological concepts are developed. In fact, Jean Sammet's statement concerning one of the major reasons for the proliferation of programming languages can be applied virtually unchanged to the terminological explosion [17: 310]: "Some of the causes and motivations behind the development of these languages rest in quirks of human nature rather than technological progress or lack thereof.

Thus, as long as people find it fun to develop languages, as long as they want something which is specifically tailored exactly to their needs, and as long as they are going to find picayune faults with the existing languages, there is very little that technical progress can do to reduce the number of languages".*

Thus, although the translator may be familiar with the scientific or technical subject matter and its fundamental terminology, the number of terms which take on different shades of meanings, cover new concepts when used by various authors, or are outright neologisms, may tend to confound the translator in his attempts to conscientiously determine the correct translation [20,21,22]. Compounding this confusion is the fact that at the rate at which new terms are coined to communicate technological development, even the best technical and scientific bilingual or multilingual dictionaries are out of date by the time they appear in print [15,22,23,24,25,26].

IMPROVING THE ACCESS TO SPECIALIZED TERMINOLOGY

To overcome the impediments in scientific and technical translation arising from the dispersion and inaccessibility of terminology, a variety of suggestions have been made [7,8,27,28,29,30,31,32] which may be summarized as follows:

1. Constructive efforts should be made by closer coordination between the terminology boards of the various disciplines, technical associations and laboratory information services to standardize jargon appearing in printed form.

2. A coherent approach should be established to ensure that all meanings of a scientific/technical term are recorded in the dictionary with reference to the field to which the meaning applies.

3. The rate at which special terms are to be added to existing monolingual glossaries and bilingual or multilingual dictionaries should be increased, eventually contributing to a reduction in the translator's overall search time for specialized terminology.

---

* Of the large number of synonymous computer technology terms, only one example is given for "the function to combine object modules to produce a single program," as used by the data processing community: linkage editor, link editor, builder, winder, loader, linking loader, relocatable loader, linkage loader, linking (relocatable) loader, collector, job loader [16].

4. The latest dictionaries and reference manuals should be at the translator's immediate disposal so that the proper terms can be looked up speedily and conveniently.

5. The translator should be able to rapidly obtain a terminological digest of a text to be translated, i.e., a list of searched-for terms extracted from the dictionary in order of textual occurrence and/or alphabetical occurrence.

This paper does not address itself to objectives 1 and 2, viz. closer coordination between the parties to standardize jargon and to ensure that all meanings of technical terms are documented. Such endeavors are the domain of the technical experts and technical writers. Obviously, the mere availability of computers does not constitute a remedy for controlling the jargon explosion or for solving human communications problems. However, the opportunity to interact with computers on-line in a terminal-oriented environment does provide the potential for finding effective solutions to objectives 3 through 5 without undue emphasis on the willingness of the technologists to document and standardize their specialized terminology expeditiously.

Approaches to meeting objectives 3 and 4 have been described in detail elsewhere [33,34,35,36,37]. In summary, objective 3 can be met by giving the user access to a time-shared computer system supporting data bases and dictionary maintenance programs to allow bilingual or multilingual dictionary generation and updating. Objective 4 involves dictionary lookup and browse programs for presenting dictionary entries in hardcopy or video format on appropriate terminal devices or on high-speed printers capable of producing high quality copy. Concomitant requirements would entail (1) context editing, making it easy both to change stored translation texts and dictionaries and to input new ones, and (2) formatting, to produce a variety of professional-looking translation and dictionary layouts.

The approach which is described on the following pages involves objective 5 and is oriented toward reducing the manual searching and sifting time which the translator requires to determine the proper terminology in a translation, and thereby toward increasing his productivity. The approach deals with the semi-automatic generation of terminological digests of texts to be translated, a method whose basic ideas can be traced back to the work on text-related glossaries by the West German Bundessprachenamt [38].

The Bundessprachenamt (Übersetzerdienst der Bundeswehr) analyzed the types of errors in technical-scientific translations and found, among other things, that the rate of mistranslated and untranslated specialized terminology increased proportionately to the increase in the technical difficulty level of source texts. On the greatest difficulty level, terminological errors accounted for 62.1% of the translation error total (where the range of error types included such categories as orthographical mistakes, punctuation, wrong German preposition, inflectional errors, English word order, text inaccuracies and omission of information). Concurrent time and productivity studies were conducted indicating that translators using dictionary lists with special source/target language terms which were exclusively related to the technical-scientific texts to be translated could reduce the error rate by approximately 40% and increase their productivity by over 50% as compared to their colleagues who had well-equipped conventional technical-scientific libraries and the consultation of their colleagues at their disposal [39].

GENERATING TERMINOLOGICAL DIGESTS

As indicated above, a terminological digest is a list of terms extracted from a main dictionary in the order of textual occurrence or alphabetical occurrence, i.e., the desired terminological framework of a text to be translated. The text to be processed may be of arbitrary length. Figure 1 represents an example of a text portion to be translated. Figure 2 shows the automatically-produced terminological digest of this text in the order in which the desired terms occur, and Figure 3 shows the terminological digest in alphabetic order.

The user has at his disposal a terminal, either a typewriter or a video display unit (Figure 4), which may be connected over regular telephone lines to a computer. After having turned on the terminal and, if required, dialed up the computer and made the connection, production of a terminological digest is achieved by first invoking the appropriate program and then entering the terms one wishes to have looked up at the keyboard.

If the source text is stored in machine-readable format in the system, as is the case for text produced by a variety of text-processing systems, it may also be displayed and examined by rolling it up or down ("scrolling") on the surface of a display screen. However, depending on the translation job, material to be translated may only be obtainable in non-machine-processable format, in which case it cannot be stored in the machine. Moreover, even though

source text may be available in machine processable format, it may not be accessible on a particular computer system with dictionary files, because of lack of storage space or because of computer installation policy. In fact, secretarial copy aids for holding manuscripts may make manual page flipping competitive with, and perhaps even more economical than, scrolling of machine-readable source text.

Temporary lack of a terminal, or assignment considerations between translators and typists, may call for separation of the task of identifying special terms for terminological digest production and of entering these terms in the system. For example, the translator may wish to encircle the desired terminology in the source text and submit it to a typist who may input these terms on-line or off-line, by terminal or typewriter (e.g., an MC/ST or MT/ST).

All terms entered are immediately looked up by the program in a comprehensive dictionary, which may be oriented toward a special subject area, and provided with target language equivalents and other germane information. At the user's option, the terms may be presented singly (i.e., immediate display of a term in dictionary context as soon as the term is entered or selected), as a list in the order of entering (normally the order in which they occur in the text to be translated), or as an alphabetically-sorted list. The terminological digests may also be saved and used repetitively, automatically edited as any other translation document [33], directed to offline output devices for high-speed printing or punching, or transmitted by telecommunications to other terminals or computers.

The dictionary storage organization is designed to cope with the growth potential of multilingual dictionaries, where an extremely large number of entries may eventually be accumulated. Details of this organization and of the associated dictionary lookup procedure are described elsewhere [33]. During execution of a single dictionary lookup, the desired source language term is compared with a table of source language terms whose corresponding dictionary entries occur at fixed intervals in the dictionary file. Dependent upon a high/low/equal compare result, appropriate routines are called to move the desired entry into a storage buffer for a video display unit or a typewriter terminal. Unless the buffer is completely occupied, adjacent entries are retrieved and packed into it until it is filled. This dictionary excerpt is then flashed onto the display screen. If a video display unit is not used, the entry is printed out on the typewriter.

For a terminological digest consisting of a list of entries, the same dictionary lookup procedure is used except

that the system delays the search until all terms to be looked up have been entered. At this point the lookup procedure is applied iteratively to each term of the entered collection in order to obtain the corresponding dictionary entry. When this process has been completed, the resulting entries are displayed, typed out or printed as a group.

Depending upon the translation task or the size of the digest, the user can cause immediate digest display, type-out on his terminal, or printout on a high-speed printer, or he can save it as a file for future translation work and possible additional terminological and statistical investigations. At his option, he may also cause the system to generate an alphabetically-sorted digest.* This could be especially helpful if more than one translator works on one large text, each concentrating on sections for which terminological digests in text order and a global alphabetically-sorted digest encompassing the entire text (Figure 5) are generated. Although different sections of the same text are translated by different translators, consistency of terminology is maintained by usage of the machine dictionary, ensuring that the same terms are always translated in the same way [40]. A specialist in an editorial function may decide that some editing of the terminological digest is required before it is used by the individual translators. For example, a manual on unit record machines in electronic data processing might contain, among other things, the terms "card stacker" (English/German dictionary equivalents: Kartenablage, Ablagefach), "vertical line" (English/German dictionary equivalents: Senkrechte, Vertikale), and "level" (English/German dictionary equivalents: Ebene, Ordnung, Stufe, Pegel, Niveau). All of those translations may be valid within the same text. However, the specialist may conclude

a) that "Ablagefach" must be edited out of the terminological digest to maintain uniformity in equipment terminology,

b) that although either "Senkrechte" or "Vertikale" could be edited out, they may be used interchangeably by the translators because of their complete unambiguity in German, and

c) that "Ebene, Ordnung, Stufe, Pegel, Niveau" must be left untouched, since each translation may have to be used even within a small stretch of text, so that the choice must therefore be left to the discretion of the translator. By using the automatic editing facility [33], changes to the terminological digests can be made extremely rapidly.

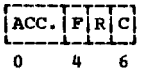
---

* see Appendix for memory layout and sort mechanism in terminological digest generation.

KEY IN STORAGE

For purposes of protection and recording of references and changes, main storage is divided into blocks of 2,048 bytes, each block having an address that is a multiple of 2,048. A control field, called "key in storage", is associated with each block of storage.

The key in storage has the following format:



The bit positions in the key are allocated as follows:

Access-Control Bits (ACC): Bits 0-3 are matched against the four-bit protection key whenever information is stored, or whenever information is fetched from a location that is protected against fetching.

Fetch-Protection Bit (F): Bit 4 controls whether protection applies to fetch-type references: a zero indicates that only store-type references are monitored and that fetching with any protection key is permitted; a one indicates that protection applies to both fetching and storing. No distinction is made between the fetching of instructions and of operands.

Reference Bit (R): Bit 5 normally is set to one each time a location in the corresponding storage block is referred to either for storing or for fetching of information. This bit is associated with dynamic address translation.

Change Bit (C): Bit 6 is set to one each time information is stored into the corresponding storage block. This bit is associated with dynamic address translation.

The key in storage is not part of addressable storage. The program can explictly place information in all seven bits of the key by SET STORAGE KEY, and the contents of the key can be inspected by INSERT STORAGE KEY. Additionally, the instruction RESET REFERENCE BIT provides a means of inspecting the reference and change bits and of setting the reference bit to zero.

PROTECTION

The protection facility is provided to protect the contents of main storage from destruction or misuse caused by erroneous or unauthorized storing or fetching by the program. It provides protection against improper stor
storing and f
improper fetchi

PROTECTION ACTION

When protection
access, the key in s
the protection key
request for storage
permitted only when
matches the protecti
said to match when the
of the key in storage
protection key or wh
is zero. A fetch is
keys match or when
storage is zero. The
summarized in the tal
tion Action."

Conditions	
Bit 4 of Key in Storage	Key Relat
0	Mat
0	Mis
1	Ma
1	Mi

Explanatio

Match	Th th tc pr
Yes	Acc
No	Acc fetc not progr conte tion a

Summary of Protec

When the access
by the CPU, and
protection key of
as the comparand
the CPU occupies b
PSW. When the
channel, and prote
tection key assoc
tion is used as t
tion key for an I
in bit positions
word (CAW) and i
0-3 of the chanr
as a result of

Figure 1: Example of a text portion to be translated.

key in storage: Speicherschluessel [S/370] [SYST]

protection: Schutz, Protektion, Beschuetzung; Schutzzoll [LEG]

reference: Hinweis, Bezugnahme Nachschlagen; mit Verweisungen versehen;
durch Verweisungen finden

main storage: Hauptspeicher [SYST]

block: Block [SYST]; Satzblock [SYST]; blocken [SYST]; blockieren;
in Bloecke formen

address: Adresse; Ansprache; adressieren; anreden

multiple: Vielfaches; vielfach

control field: Kontrollfeld [S/370]; Sortierfeld [SYST]

key in storage: Speicherschluessel [S/370] [SYST]

storage: Speicher [SYST]; Speicherung; Lagerung; Lagermiete [COM]

bit: Bit [SYST], Binaerziffer [SYST]; Bohrspitze [MECH]; Schluesselbart;
kleines Stueckchen

allocate: zuordnen, zuteilen, anweisen

access-control bit: Zugriffs-Steuerungsbit [S/370]

match: abgleichen [SYST]; verbinden, paaren, paarweise verbinden;
passend verbinden [MECH]; Gleiche(r, s), Zusammenbringen

protection key: Schutzschluessel [S/370] [SYST]

fetch: Abruf [SYST]; abrufen [SYST]; abholen

fetch-protection bit: Abrufschutzbit [S/370] [SYST]

apply: zutreffen; anwenden, verwenden; auftragen [MECH]; bewerben

fetch-type
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

indicate: anzeigen, angeben, andeuten

store-type
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

fetch: Abruf [SYST]; abrufen [SYST]; abholen

store: speichern [SYST]; lagern, aufspeichern; Speicher [SYST];
Lager, Magazin

instruction: Instruktion [SYST]; Anweisung, Belehrung; Lehre

operand: Operand (Parameter in einer Instruktion);
operand specification = Spezifikation fuer einen Operanden

reference bit: Hinweisbit [S/370] [SYST]

refer: hinweisen, verweisen, sich beziehen; sich wenden

associate: in Verbindung stehen; assoziieren, vereinigen

dynamic address translation: dynamische Adressumsetzung [S/370] [SYST]

change bit: Veraenderungsbit [S/370] [SYST]

dynamic address translation: dynamische Adressumsetzung [S/370] [SYST]

addressable: adressierbar

storage: Speicher [SYST]; Speicherung; Lagerung; Lagermiete [COM]

SET STORAGE KEY
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

contents: Inhalt

inspect: pruefen, untersuchen, beaufsichtigen

INSERT STORAGE KEY
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

RESET REFERENCE BIT
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

protection: Schutz, Protektion, Beschuetzung; Schutzzoll [LEG]

main storage: Hauptspeicher [SYST]

store: speichern [SYST]; lagern, aufspeichern; Speicher [SYST];
Lager, Magazin

fetch: Abruf [SYST]; abrufen [SYST]; abholen

protection: Schutz, Protektion, Beschuetzung; Schutzzoll [LEG]

Figure 2: Terminological digest of a text portion to be translated into German.

access-control bit: Zugriffs-Steuerungsbit [S/370]

address: Adresse; Ansprache; adressieren; anreden

addressable: adressierbar

allocate: zuordnen, zuteilen, anweisen

apply: zutreffen; anwenden, verwenden; auftragen [MECH]; bewerben

associate: in Verbindung stehen; assoziieren, vereinigen

bit: Bit [SYST], Binaerziffer [SYST]; Bohrspitze [MECH]; Schluesselbart;
kleines Stueckchen

block: Block [SYST]; Satzblock [SYST]; blocken [SYST]; blockieren;
in Bloecke formen

change bit: Veraenderungsbit [S/370] [SYST]

contents: Inhalt

control field: Kontrollfeld [S/370]; Sortierfeld [SYST]

dynamic address translation: dynamische Adressumsetzung [S/370] [SYST]

fetch: Abruf [SYST]; abrufen [SYST]; abholen

fetch-protection bit: Abrufsschutzbit [S/370] [SYST]

fetch-type
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

indicate: anzeigen, angeben, andeuten

INSERT STORAGE KEY
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

inspect: pruefen, untersuchen, beaufsichtigen

instruction: Instruktion [SYST]; Anweisung, Belehrung; Lehre

key in storage: Speicherschluessel [S/370] [SYST]

main storage: Hauptspeicher [SYST]

match: abgleichen [SYST]; verbinden, paaren, paarweise verbinden;
passend verbinden [MECH]; Gleiche(r, s), Zusammenbringen

multiple: Vielfaches; vielfach

operand: Operand (Parameter in einer Instruktion);
operand specification = Spezifikation fuer einen Operanden

protection: Schutz, Protektion, Beschuetzung; Schutzzoll [LEG]

protection key: Schutzschluessel [S/370] [SYST]

refer: hinweisen, verweisen, sich beziehen; sich wenden

reference: Hinweis, Bezugnahme Nachschlagen; mit Verweisungen versehen;
durch Verweisungen finden

reference bit: Hinweisbit [S/370] [SYST]

RESET REFERENCE BIT
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

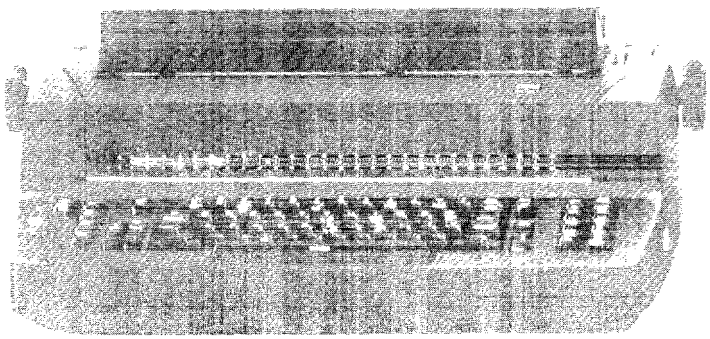
SET STORAGE KEY
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

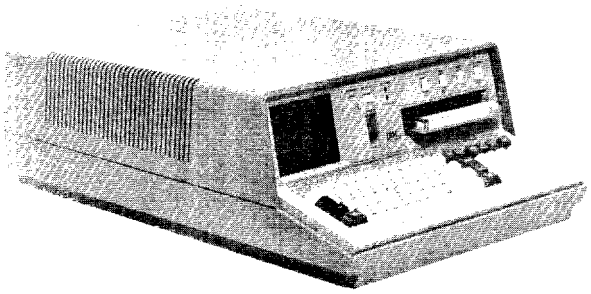
storage: Speicher [SYST]; Speicherung; Lagerung; Lagermiete [COM]

store: speichern [SYST]; lagern, aufspeichern; Speicher [SYST];
Lager, Magazin

store-type
*** FEHLT IM WOERTERBUCH (IST ABFRAGE FALSCH BUCHSTABIERT?) ***

Figure 3: Alphabetically-sorted terminological digest of a text portion to be translated into German.





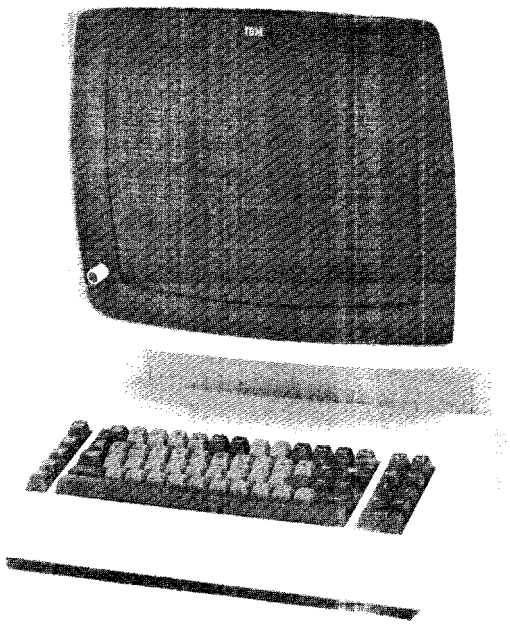


Figure 4: Basic types of typewriter and video display terminals.

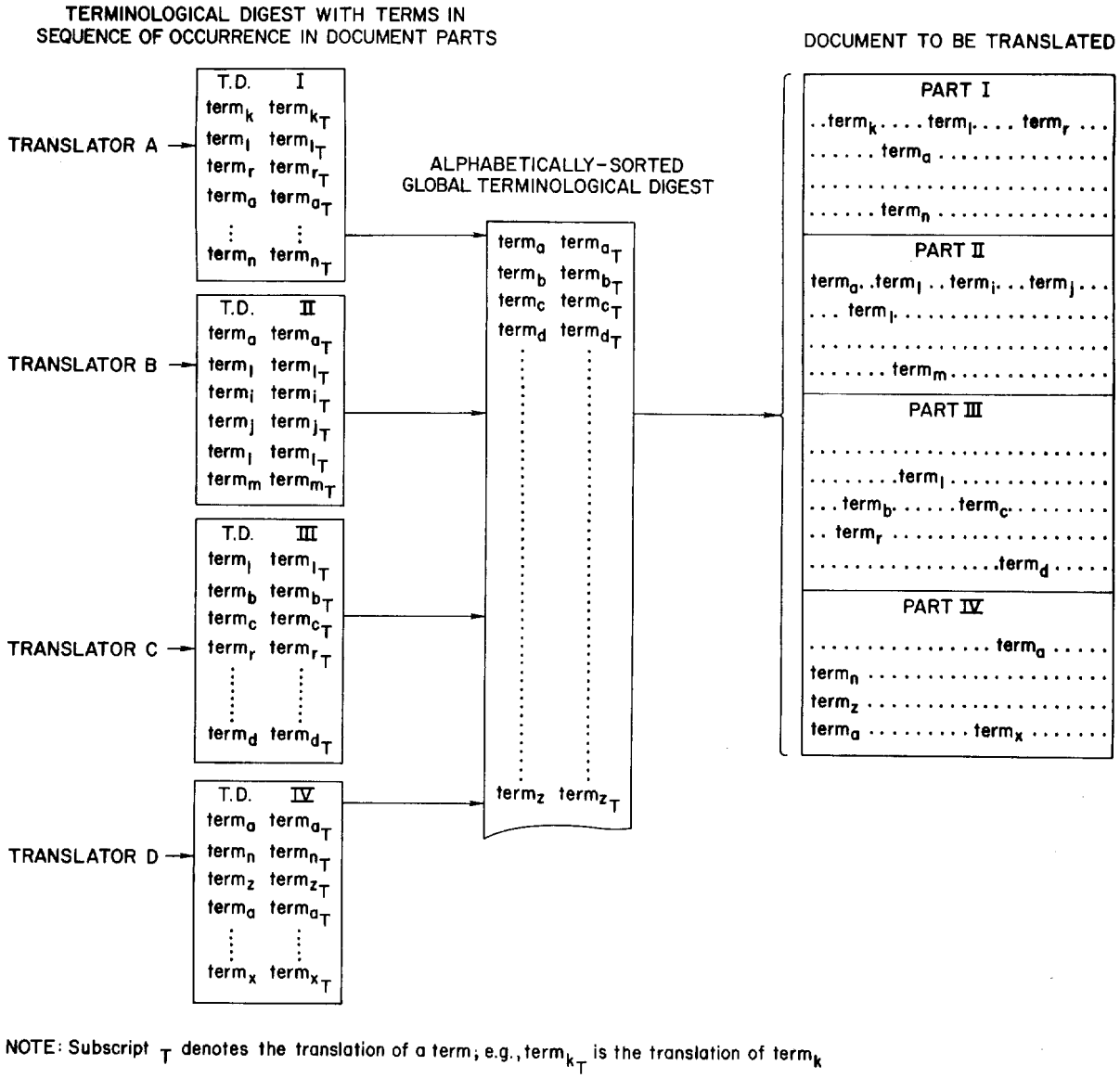


Figure 5: Team of translators working cooperatively on a large translation task, using terminological digests.

ON-LINE OPERATING SIMPLICITY

Producing a terminological digest by computer online should be maximally simple in order to adapt to the needs of an inexperienced computer user, such as the translator, interpreter, terminologist, lexicographer, editor, or typist. Figure 6 illustrates how a user can employ a video display terminal to generate a terminological digest. The system allows the user various options as to how to proceed, prompting him to enter his choice prior to terminological digest generation. In Figure 6, after having selected a sorted version of the terminological digest, the user has begun entering the desired terms, working from a printed text document.

If the source text is available in machine-readable format, it can also be automatically retrieved and displayed on the screen. Figure 7 shows how a portion of the English source text of a manual, which was originally produced by automatic typesetting, is displayed on the video screen (first 22 lines of display). The user can move the entire text of the manual up or down on the screen, in fact viewing the text as through a window, looking for terms whose translation he wishes to know. Whenever such a term occurs, the user enters it via the keyboard, at which time it is displayed at the left of the bottom line of the screen and the term within the text is brightened (i.e., displayed in double intensity, verifying to the user that it has been selected for terminological digest generation). The user can continue entering as many terms as desired, thereby creating the terminological skeleton of the text to be translated. Hitting the ENTER key of the keyboard an extra time causes production of the terminological digest, which can then be displayed (Figure 8A) or printed on an attached typewriter (Figure 8B) or a high-speed printer (Figure 8C).

Of course, the user need not display the text to be translated on the video screen, if he prefers to work from its printed version. Moreover, a text may not be available in machine-readable format, or the user may employ a typewriter terminal where it may be inefficient to print out portions of the text intermittently and then enter the terms to be looked up.

During terminological digest generation the user may flip back and forth between his input and the already-entered terminological digest terms, e.g., to scan for possible similarities or redundancies of terms. He may also delete terms already selected.

user---->
system-->
user---->
system-->
user---->

```
lookup TA
TYPE 'NO' IF YOU DON'T KNOW HOW THIS LOOKUP WORKS. OTHERWISE PROCEED WITH OPTIO
NS:
no
NOTE THESE OPTIONS....
TYPE 'SINGLE' IF EACH TERM IS TO BE LOOKED UP SINGLY."TYPE 'ALLTERMS' IF ALL TE
RMS ARE FIRST TO BE LOOKED UP AND THEN PRINTED OUT COLLECTIVELY."TYPE 'FILE' IF
 ALL TERMS ARE TO BE LOOKED UP AND SAVED IN A FILE.
TYPE 'SORT' IF ALL TERMS ARE TO BE ALPHABETICALLY SORTED, LOOKED UP AND THEN PR
INTED OUT."TYPE 'SORTFILE' IF ALL TERMS ARE TO BE ALPHABETICALLY SORTED, LOOKED
 UP AND THEN SAVED IN A FILE."AN EXTRA CARRIAGE RETURN TERMINATES THIS LOOKUP.
sort

edit
omit
parameter list
positional
filetype
filemode
verification
record length
pad_
                                   Benutzer-Eingabe          1  1
```

Figure 6: Snapshot of a video display screen during on-line generation of a terminological digest (current system environment is indicated to the user in double brightness on right-hand bottom line).

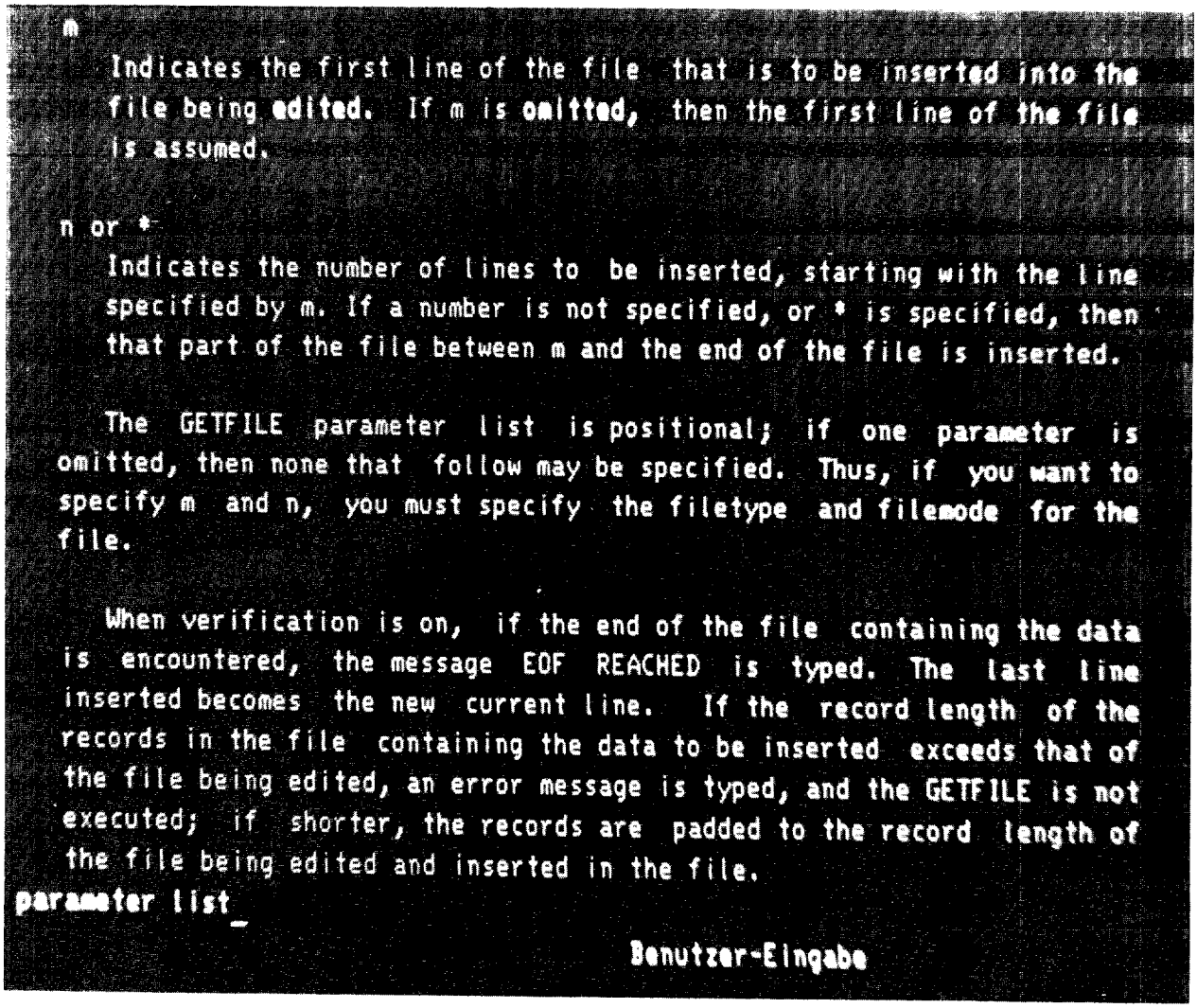


Figure 7: Display of a text portion of a computer manual with term "parameter list" being entered.

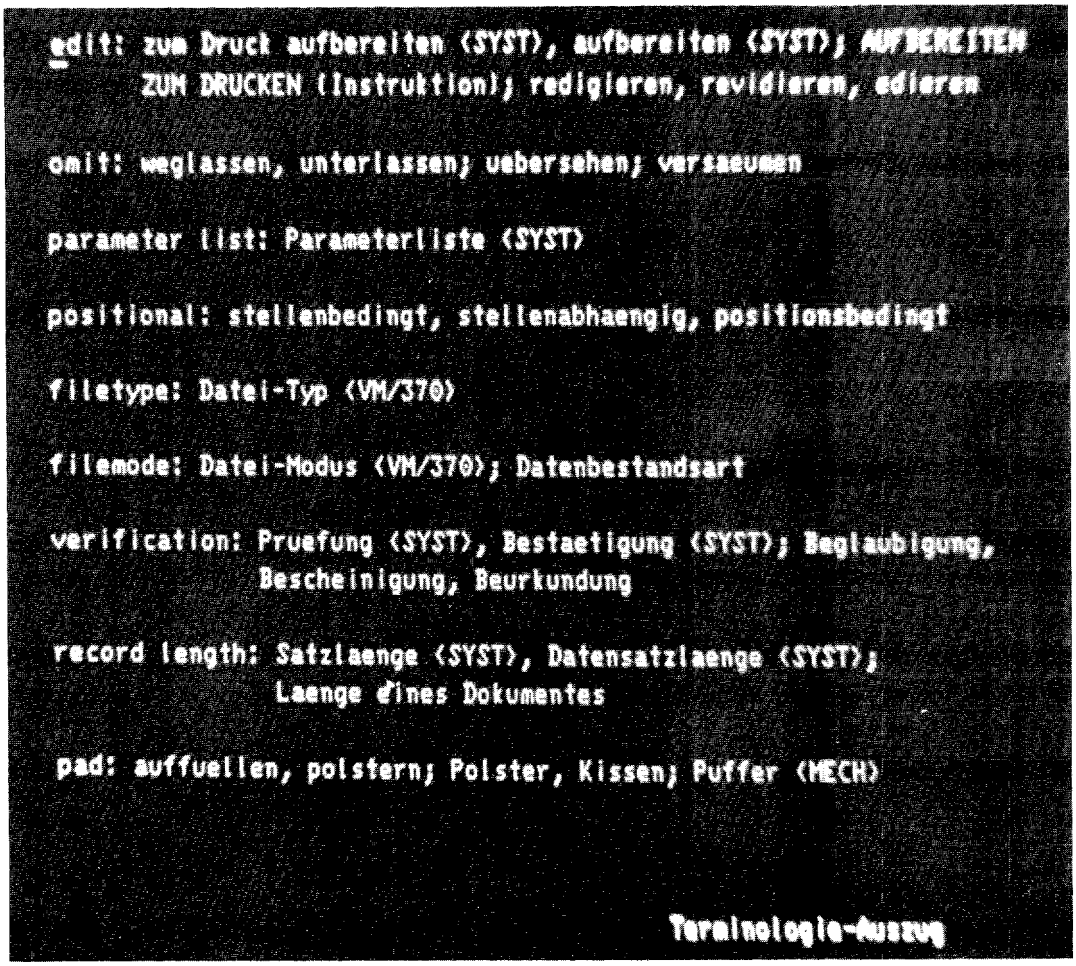


Figure 8A: Display of a portion of an English-German terminological digest showing terms in order of textual occurrence.

edit: zum Druck aufbereiten [SYST], aufbereiten [SYST]; AUFBEREITEN ZUM DRUCKEN (Instruktion); redigieren, revidieren, edieren

omit: weglassen, unterlassen; uebersehen; versaeumen

parameter list: Parameterliste [SYST]

positional: stellenbedingt, stellenabhaengig, positionsbedingt

filetype: Datei-Typ [VM/370]

filemode: Datei-Modus [VM/370]; Datenbestandsart

verification: Pruefung [SYST], Bestaetigung [SYST]; Beglaubigung, Bescheinigung, Beurkundung

record length: Satzlaenge [SYST], Datensatzlaenge [SYST]; Laenge eines Dokumentes

pad: auffuellen, polstern; Polster, Kissen; Puffer [MECH]

Figure 8B: Portion of an English-German terminological digest shcwing terms in order of textual occurrence, typed out on CMC/ST.

edit: zum Druck aufbereiten <SYST>, aufbereiten <SYST>; AUFBEREITEN ZUM DRUCKEN (Instruktion); redigieren, revidieren, edieren

omit: weglassen, unterlassen; uebersehen; versaeumen

parameter list: Parameterliste <SYST>

positional: stellenbedingt, stellenabhaengig, positionsbedingt

filetype: Datei-Typ <VM/370>

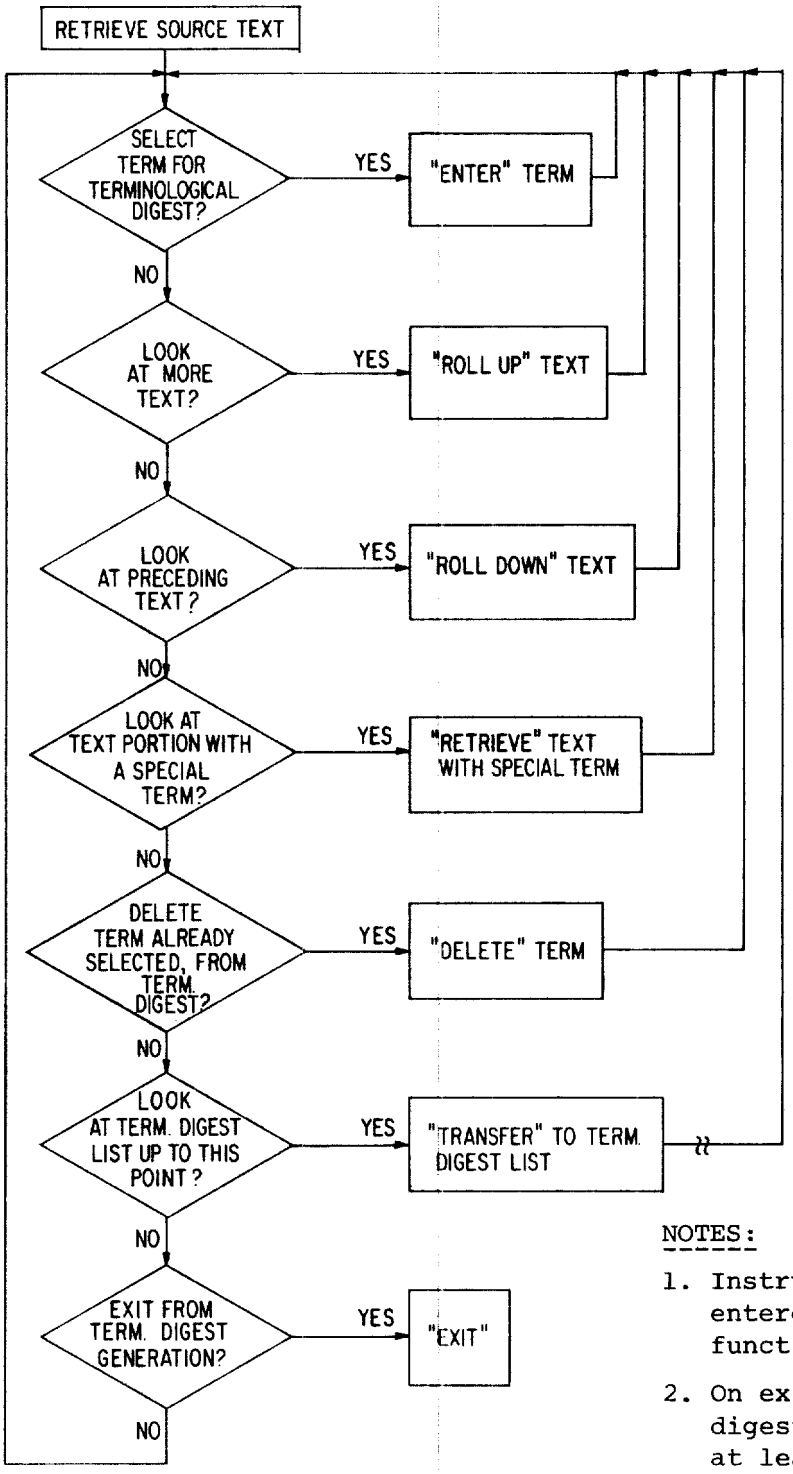
filemode: Datei-Modus <VM/370>; Datenbestandsart

verification: Pruefung <SYST>, Bestaetigung <SYST>; Beglaubigung, Bescheinigung, Beurkundung

record length: Satzlaenge <SYST>, Datensatzlaenge <SYST>; Laenge eines Dokumentes

pad: auffuellen, polstern; Polster, Kissen; Puffer <MECH>

Figure 8C: Portion of an English-German terminological digest showing terms in order of textual occurrence, printed on an IBM 1403 Printer.



NOTES:

1. Instructions in quotes may be entered by pressing a program function key.
2. On exit, a terminological digest will be produced if at least one term has been selected.

Figure 9: Flow of control of a terminological digest generation process.

If he has at his disposal a display terminal with program function keys, which allow communicating instructions to the system in a most simple way by merely pressing the key, the user may depress one of six keys for:

1. Moving text upward on the screen;

2. Moving text downward on the screen;

3. Flipping screen "pages";

4. Retrieving a screen "page" with a typed term (or string of characters);

5. Deleting a term already entered in the terminological digest list;

6. Transferring to the terminological digest to examine the terms thus far entered.

Having transferred to the terminological digest environment, the program functions are symmetrical, i.e., the keys have the same meaning for the terminological digest list except that pressing the sixth key will transfer the user back to the point in the text where he left off before transferring to the terminological digest list. The user may also enter any of the six instructions on the keyboard rather than hitting the appropriate program function key; this is mandatory if he uses a keyboard without such keys. A typed instruction must be preceded by a > sign to signal the system that this is not a term to be looked up. Figure 9 represents the flow of control of a terminological digest generation process for a source text accessible through the system.

From a human factors engineering point of view, the keyboard of the display terminal, which provides the main contact of the user with the computer system, basically operates like a typewriter but, in several ways, offers greatly improved performance for a translator or editor who may be a non-typist. The keys react immediately to being pressed by the user and so quietly that someone directly adjacent to him may not be audibly aware of the operation. Visual communication is made easy by the display of large-sized characters within sufficient context on the screen. Since typing on the keyboard "prints" characters on the screen instead of on paper, error correction is greatly improved. By moving the cursor* to the error (or any characters to be changed) and keying in the correct

---

* a movable underscore marking the position on the screen that the character entered from the keyboard will occupy

characters, previous information is overlaid by the new data. Changing information in the middle of text can also be quickly performed with the aid of the cursor: for example, when data are inserted or deleted in the middle of text, the immediately following data are automatically shifted forward in the case of insertions, and automatically contracted in the case of deletions. Moreover, using the ERASE INPUT key causes the information just typed (but not yet entered) to be blanked, while using the CLEAR key immediately clears the entire display without causing a disturbance within the system.

INTERPRETATION AND ON-LINE GENERATION OF TERMINOLOGICAL DIGESTS

Since, at the user's option, the translation of a queried term may be immediately displayed as soon as the term is entered, the system could very well be used by interpreters during consecutive as well as simultaneous interpretation. As described in [41: 154],

> It is generally believed that an interpreter cannot consult reference works or colleagues as a translator can. This is only partially true. I have always found it useful to have pertinent dictionaries (whether general or specialized) in the booth, which can be consulted by one's boothmate or by the interpreter himself. Since an unknown or vaguely known word is likely to crop up more than once, it should be looked up in one of three ways; (a) by the boothmate immediately after it occurs; (b) by the interpreter during his rest period; or, if (a) is not possible and recourse to (b) would mean risking a second or even third encounter with the problematical word, then (c) by the interpreter himself while he is still interpreting. Boothmates, too, may be queried with an inquisitive look and a note can often be slipped to a coworker in a neighboring booth. Such consultations should of course never interrupt the flow of the interpretation. An interpreter should also be alert to difficult words which a boothmate or other colleague may have to interpret and should therefore not hesitate to slip appropriate notes at the proper time. This is often the case when one of the interpreters has specialized knowledge and can sometimes even anticipate the appearance of a difficult word or expression with amazing accuracy.

Thus, a display terminal, which has great flexibility with regard to siting and no air-conditioning requirements,

could be a valuable aid for an interpreter who has to consult special dictionaries extremely rapidly for difficult terms encountered during a discourse. Because on-line lookup is generally faster than flipping pages [42], such queries need not infringe upon the flow of the interpreter's speech. In addition, all terms looked up can be simultaneously saved by the system in a file representing, in effect, the terminological "minutes" of a meeting. Therefore, it is conceivable that rapid on-line lookup of specialized terminology may enhance the process of interpretation and increase the fidelity of the interpreter's work.

Whenever possible, interpreters are urged to collect all documentation required for interpretation before every conference and scrutinize this material for specialized terminology to prepare its translation. Moreover,

> just as the translator, the interpreter should build up a glossary of technical terms both for his own use and that of his colleagues. These can be circulated among the interpreters at the end of each session and are of course kept for future conferences. Delegates can also usually be queried after the session on difficult terms [41: 155].

Thus, most interpreters and translators have their private terminology lists in addition to the official reference material. The system and equipment described will permit the user to rapidly input, update and display terminology lists and dictionary excerpts via the display unit and/or typewriter-terminal.

ACKNOWLEDGMENT

The author would like to express his appreciation to Warren J. Plath and Mark Pivovonsky for their comments and criticisms, which were invaluable in the preparation of this paper. A similar note of gratitude is due to William I. Bertsche, President of the American Translators Association, as well as to Professor Etilvia Arjona, Monterey Institute of Foreign Studies, California, whose interest has greatly contributed to the write-up of this work.

REFERENCES

[1] D. L. Gold, "On Quality in Translation," babel, vol. XVIII, No. 1/1972, pp. 10-12.

[2] D. L. Gold, "On Quality in Translation II", babel, vol. XVIII, No. 4/1972, pp. 29-30.

[3] J. Maillot, "Terminologie et traduction," Meta Journal des traducteurs, vol. 16, No. 1-2, March-June 1971, pp. 75-81.

[4] E. A. Nida,, Toward a Science of Translating, E. J. Brill, Leiden, Netherlands, 1964, pp. 145-155.

[5] H. W. Sinaiko and R. W. Brislin, "Evaluating Language Translations: Experiments on Three Assessment Methods," Journal of Applied Psychology," vol. 57, No. 3, 1973, pp. 328-334.

[6] H. W. Sinaiko, "Verbal Factors in Human Engineering: Some Cultural and Psychological Data," Ethnic Variables in Human Factors Engineering, A. P. E. Chapanis, ed., The John Hopkins Press, Baltimore, 1975, pp. 159-177.

[7] K. H. Brinkmann, "Uberlegungen zum Aufbau und Betrieb von Terminologie-Datenbanken als Voraussetzung der maschinenunterstützten Übersetzung," Nachrichten für Dokumentation, vol. 25, No. 3, June 1974, pp. 99-105.

[8] R. N. Basu, "Barriers to Effective Communication in the Scientific World," IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, vol. PC-15, No. 2, June 1972, pp. 30-33.

[9] R. C. Moser, Space-Age Acronyms, Abbreviations and Designations, IFI/PLENUM, New York, 1969.

[10] E. T. Crowley and R. C. Thomas, Acronyms and Initialisms Dictionary, Gale Research Company, Book Tower, Detroit, Michigan, 1973.

[11] E. T. Crowley and R. C. Thomas, New Acronyms and Initialisms, (Supplement to [9]), Gale Research Company, Book Tower, Detroit, Michigan, 1974.

[12] E. T. Crowley and R. C. Thomas, New Acronyms and Initialisms, (Supplement to [10]), Gale Research Company, Book Tower, Detroit, Michigan, 1975.

[13] E. Pugh, Second Dictionary of Acronyms & Abbreviations, Archon Books, The Shoe String Press, Inc., Hamden, Connecticut, 1974.

[14] R. de Sola, Abbreviations Dictionary, American Elsevir Publishing Company, Inc., New York, 1974.

[15] E. A. Lacy, "Special Dictionaries for the Electronics Engineer," IEEE TRANSACTIONS ON ENGINEERING WRITING AND SPEECH, vol. EWS-6, No. 1, September 1969, pp. 31-35.

[16] A. P. Sayers, Ed., Operating Systems Survey, The COMTRE Corp., Auerbach Publishers, New York, 1971.

[17] J. E. Sammet, "An Overview of programming languages for specialized application areas," AFIPS Spring Joint Computer Conference, AFIPS Press, Montvale, New Jersey, vol. 40, 1972, pp. 299-311.

[18] J. M. Lufkin, "Generalization and Interpretation of Science and Technology," IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, vol. PC-15, No. 4, December 1972, pp. 108-111.

[19] F. B. Thompson and B. H. Dostert, "The future of specialized languages," AFIPS Spring Joint Computer Conference, AFIPS Press, Montvale, New Jersey, vol. 40, 1972, pp. 313-319.

[20] J. A. Bachrach, "An Experiment in Automatic Dictionary Look-up," The Incorporated Linguist, vol. 13, No. 2, April 1974, pp. 47-49.

[21] U. Förster, "Der Sprachberatungsdienst," babel, vol. XVIII, No. 2/1972, pp. 24-38.

[22] R. Herzog, "Die Anwendung computer-linguistischer Methoden bei der Kompilation von Fachwörterbüchern," Beiträge zur Linguistik und Informationsverarbeitung, No. 18, July 1970, pp. 26-40.

[23] J. A. Bachrach and L. Hirschberg, "Une troisième version du DICAUTOM," Actes de la 2ème conférence internationale sur le traitement automatique des langues, Grenoble, 23-25 août 1967.

[24] T. Longyka, "Technical Translation and Industrial Terminology," The ATA CHRONICLE, vol. II, No. 1, January 1973, pp. 6-9.

[25] C. J. Hyman, German-English English-German Electronics Dictionary, Consultants Bureau, New York, 1965.

[26] J. Horn, Computer and Data Processing Dictionary and Guide, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1966.

[27] R. N. Basu, "A Jargon Explosion," New Scientist, vol. 29, March 24, 1966, p. 788.

[28] National Academy of Sciences and National Academy of Engineering, SCIENTIFIC AND TECHNICAL COMMUNICATION - A Pressing National Problem and Recommendations for Its Solution, National Academy of Sciences, Washington, D. C., 1969.

[29] R. A. Evans, "But That Word Really Means...," IEEE TRANSACTIONS ON RELIABILITY, vol. R-23, No. 3, August 1974, p. 129.

[30] M. Paré, "Computerized Multilingual Word Banks Can Provide Terminological Assistance to International Standards Organizations," UNESCO Symposium on International Cooperation in Terminology, INFOTERM, Vienna, April 1975.

[31] F. Kertesz, "How to Cope with the Foreign-Language Problem: Experience Gained at a Multidisciplinary Laboratory," Journal of the American Society for Information Science, vol. 25, No. 2, March-April 1974, pp. 86-104.

[32] W. Horn, "Übersetzungsverfahren mit textbezogener Abfrage," Der Sprachmittler, vol. 8, No. 1, January 1970, pp. 17-20.

[33] E. O. Lippmann, "An Approach to Computer-aided Translation," IEEE TRANSACTIONS ON ENGINEERING WRITING AND SPEECH, vol. EWS-14, No. 1, February 1971, pp. 10-33.

[34] E. O. Lippmann and W. J. Plath, "Time-sharing and Computer-aided Translation," THE FINITE STRING, vol. 7, No. 8, October 1970, pp. 1-4.

[35] J. Schulz, "Le système TEAM, une aide à la traduction," META Journal des traducteurs, vol. 16, No. 1-2, March-June 1971, pp. 95-104.

[36] J. P. Vinay, "Utilisation électronique de la Banque de mots," META Journal des traducteurs, vol. 16, No. 1-2, March-June 1971, pp. 95-104.

[37] R. Dubuc, "TERMIUM: System Description," META Journal des traducteurs, vol. 17, No. 4, December 1972, pp. 201-219.

[38] F. Krollmann, "Linguistic data banks and the technical translator," META Journal des traducteurs, vol. 16, No. 1-2, March-June 1971, pp. 117-124.

[39] F. Krollmann, H. Schuck and U. Winkler, "Herstellung textbezogener Fachwortlisten mit einem Digitalrechner - ein Verfahren der automatischen Übersetzungshilfe," Beiträge zur Sprachkunde und Informationsverarbeitung, No. 5, January 1965, pp. 7-30.

[40] K. Gingold, "A User's Guide to Inferior Translations," The ATA CHRONICLE, vol. IV, No. 5, May 1975, pp. 6-7.

[41] D. L. Gold, "On Quality in Interpretation," babel, vol. XIX, No. 4/1973, pp. 154-155.

[42] D. Fredericksen and L. Power, "A Query System for Reviewing On-line Manuals," Computing Center Newsletter, IBM T. J. Watson Research Center, Yorktown Heights, New York, vol. 8, No. 4, March 10, 1975, pp. 37-39.

[43] IBM Systems Reference Library, IBM System/360 Time Sharing System Assembler Programmer's Guide, Form No. GC28-2032, IBM Corporation, Time Sharing System Programming Publications, Kingston, New York, 1972, pp. 157-159.

APPENDIX: SORT MECHANISM AND MEMORY LAYOUT OF TERMS IN TERMINOLOGICAL DIGEST GENERATION

This section describes the essentials of the memory layout and access scan of the terms used for terminological digest generation, whether performed in order of textual occurrence or in alphabetically-sorted order. When the terms selected for the terminological digest come into the computer, they are available in the original sequence, and, after (optional) application of a sort routine, in alphabetically-sorted sequence in the same memory area as well. They may then be retrieved in either sequence, looked up in a dictionary according to the procedure mentioned in [33] and outputted in soft or hard copy as a digest for the user.

Figure 10 represents a section of the memory with terms after they have entered the system. The terms, which can be of practically unlimited length, are placed adjacent to each other in main memory as they come in, according to a variable-length storage scheme. Each term is preceded initially by a length code and an empty memory cell. If the terms are to be sorted alphabetically, the sorting process inserts in each such empty cell a pointer (the "chain pointer") indicating the position of the next term in alphabetical sequence. If a terminological digest is requested in textual term order, this pointer is irrelevant: The terms are scanned sequentially in memory 'from left to right' and the dictionary search is performed on each term. If a terminological digest is requested in alphabetical order, the terms are first "sorted" by chaining them one by one in alphabetical sequence; as each new term is processed, the partially formed chain is scanned from the beginning in order to determine where the term is to be inserted. After the sort procedure, the terms are accessed through their chain pointers in alphabetical sequence for the dictionary search. The flow chart (Figure 11) describes this sort/comparison mechanism for variable-length terms.

The sort process uses, in addition to the pointer attached to each term, five "global" pointers to keep track of the terms being worked on. These pointers are referred to on the flowchart as FRESPNT, SAENTRY, GRVALAD, TEMPS, and PREVOLD.

Looking at a cycle of the sort process, assume a start-address pointer (SAENTRY) points to a partial sorted chained sequence and a high-value pointer (GRVALAD) points to the latest, i.e., alphabetically highest, term in that sequence. The term pointed at by the high-value pointer has already been chained to the current term (pointed at by

FRESPNT), which is yet to be examined. At this moment it is not yet known whether the current term is alphabetically higher or lower than the highest-valued term of the chain: if it is higher it will remain where it is in the chain; if it is lower it will be rechained. The current term is first compared against the term pointed at by SAENTRY. If the current term is lower, it is attached to the front of the chain, so that SAENTRY will point to it from now on. Otherwise, the whole chain is scanned beginning with the start-address pointer and each of its terms is compared against the current term. A next-term pointer (TEMPS) is used to scan the chain and a previous-term pointer (PREVOLD) trails the next-term pointer by one item, making it possible to "insert" the current term into the chain when the next term is not lower than the current term. If the "not-lower" comparison is not encountered until the end of the chain is reached, it is certain to be encountered when eventually the current term is compared to itself. The end-of-chain condition is detected by the current term having an empty chain pointer field. In such a case, the current term remains chained where it is, but the high-value pointer is advanced to point to the current term. Whatever course the comparison has taken, the program then chains the next unexamined term in memory to the last term of the chain, which is pointed at by the possibly updated high-value pointer.* The program is now ready to repeat the sort cycle.

A tight comparison mechanism for arbitrarily long comparands (terms) is attached to the sort function, making it possible to compare up to 2^{32} characters (although such long terms would certainly never be encountered).

The advantage of employing the above memory layout and sort mechanism for terminological digest generation is fourfold:

1. Terms to be sorted need not be limited or have fixed length.

2. Since the "sorting" is actually done by pointer-updating only, the data (terms) to be sorted are not moved in memory.

---

* The advantage of chaining a record to the sorted sequence before that record is examined is that the check for the end-of-chain condition can be coded outside the inner comparison loop, thus making the loop faster. (In addition, great efficiency and compactness of the comparison/sort routine in machine code is maintained by the use of assembler-language programming.)

3. The terms remain in their original order in the memory area, enabling the user to optionally fetch the terms in original or in sorted sequence, by retrieving the terms sequentially or by threading through the chain pointers, respectively.

4. If sorting is done in virtual memory, merging phases are not required, since merging is replaced by the automatic system-paging mechanism[43].

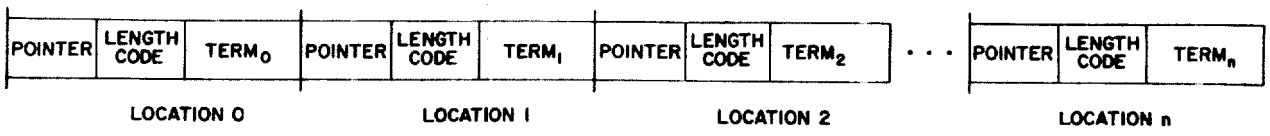


Figure 10: Conceptual memory section containing terms for a terminological digest.

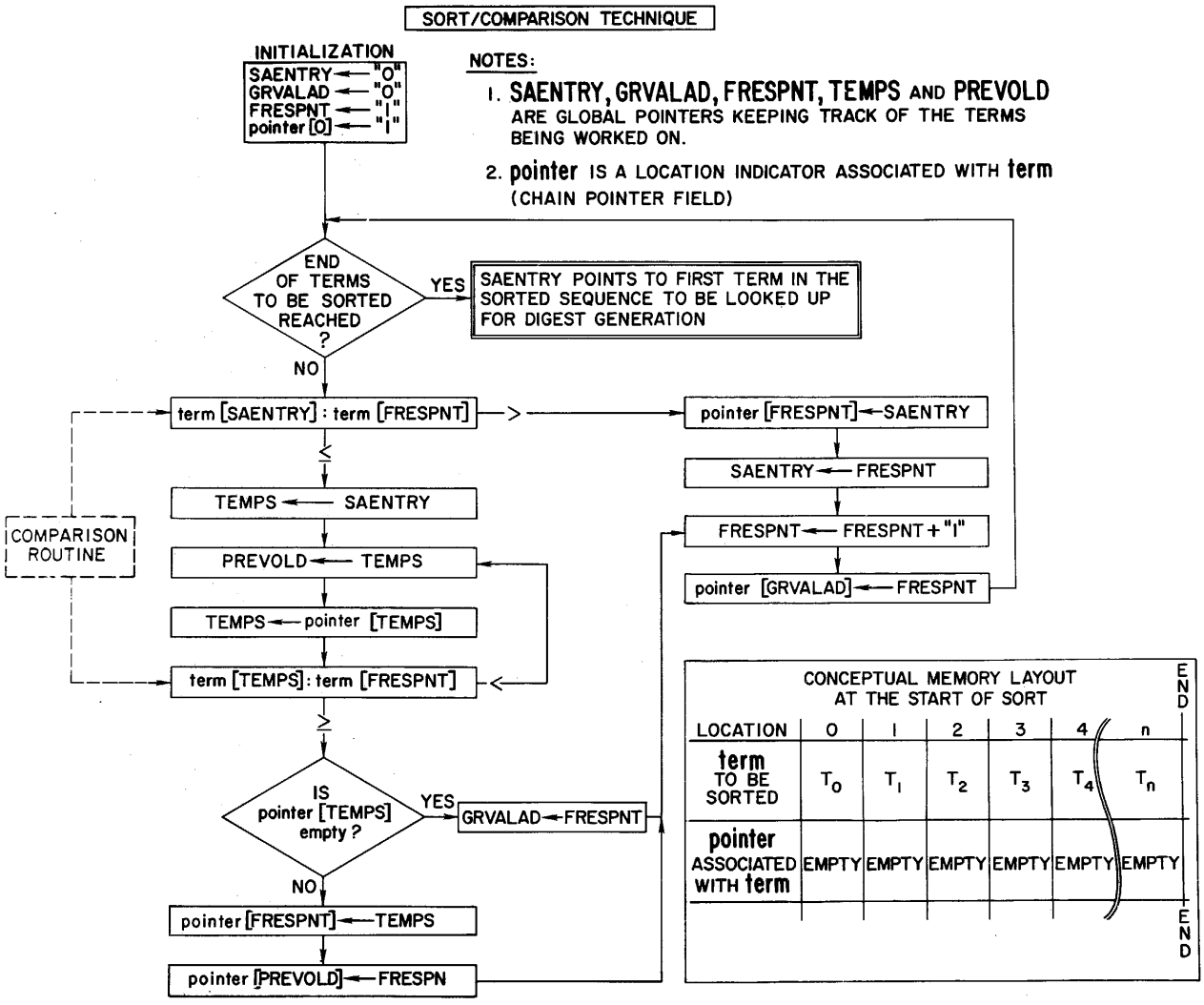


Figure 11: Sort/comparison technique.

NAME BEDRICH CHALOUPKA

INSTITUTION XONICS, INC., McLEAN, VIRGINIA

Use following space for an abstract or summary of your project

Title of Project MACHINE TRANSLATION

The Xonics Machine Translation is a fast and efficient system for the translation of Russian into English and for translation of other languages which have similar grammatical features. This system is representative of the philosophy that effective Machine Translation is one which simulates the activities of the human translator.

The computer programs which make up the system are written in the PL/1 language. They will operate on an IBM 360 or 370 computer. The entire system uses less than 100,000 bytes of computer memory for operation. Translation can be done in three different modes:

(a) Batch - for translation of large quantities of text.

(b) Sentence-by-sentence - for translation of abstracts and short articles.

(c) Interactive - for translation utilizing teleprocessing.

The system contains supporting programs for updating dictionaries. Both translation and updating of dictionaries can be done through teleprocessing.

NAME John Chandioux

INSTITUTION TAUM Project, University of Montreal

Use following space for an abstract or summary of your project

Title of Project Meteo Weather Forecast Translation

Meteo is an automatic system for the translation of weather forecasts from English into French. Public forecasts for the whole of Canada are directly sent to the system via communications network. The sentences accepted by the system do not need to be edited or revised. The remaining sentences are extracted by an interactive editor and displayed on a screen terminal for translation by a human translator. Meteo has been operating on an experimental basis 24 hours a day since December 1975 in parallel with the translation bureau. It will be fully operational by May of 1976 and presently produces 30,000 words per day The actual translation time spent is over 1,000 words per minute and estimated costs all inclusive are two cents per word.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME CHANDIOUX John

INSTITUTION TAUM, Université de Montréal

Use following space for an abstract or summary of your project

Title of Project Leibnitz, Multilingual system

Leibnitz is an international cooperation between computer translation centers interested in a multilingual system. Several european groups, the TAUM project from the Université de Montréal and a Brazilian group are presently working on this project. Most parts of the system are being written in one of the three languages made available by the CETA in Grenoble. The first one is the ATEF language, a string/tree transducer for dictionary look-up and morphological analysis. The second one is CETA and is a tree manipulating language for both transfer and generation. The last one is a tree/string transducer to be completed sometime in summer of 76.

Each group is either working on the design of an analyzer or generator for a specific language or on the transportability of the available formalisms. Research is presently under way on French, German, English, Italian, Portugese and Russian. English analysis is done by the TAUM team which is presently experimenting with a parser written in REZO its own version of Wood's Augmented Transition Networks. All particpating groups have agreed on a normalized tree representation for the output of analyzers and input of generators in order to minimize problems in the design of transfer components. The first part of the system is expected to be operational within two years.

NAME Major Lynn M. Hansen

INSTITUTION Foreign Technology Division (FTD)

Use following space for an abstract or summary of your project

Title of Project FTD Machine Translation

FTD has been utilizing machine translation since September of 1963 when the IBM MARK II system was installed at Wright-Patterson Air Force Base. The current FTD machine translation system became operational in July 1970. Since that time nearly constant improvement has been made through a series of external optimization contracts and in-house update efforts.

Three major machine translation products are produced at FTD: the unedited machine translation (MTO) with original graphics merged onto the computer printout; the preliminarily or partially edited translation (PET) which includes only those editorial changes necessary to insure the technical accuracy of the translated text; and the finished translation (MTF) which has been completely edited with proper syntactical changes and then retyped in camera-ready format.

The bulk of the machine translations at FTD deals with S&T subject matter; therefore, FTD glossaries and lexicographic routines are basically scientifically oriented. However, the system as it now exists will provide quite adequate indicative translations of almost any material.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME Fred C. Hutton

INSTITUTION Union Carbide Corp., Nuclear Div., Oak Ridge, TN

Use following space for an abstract or summary of your project

Title of Project Georgetown University MT System Usage at Oak Ridge, Tennessee

Ten years' experience in running the programs on the IBM 7090 is described. The present system, reprogrammed for the IBM 360, is described and capabilities of the system are set forth. An example of the use of the language invented by A. F. R. Brown (SLC for "Simulated Linguistic Computer") used in the preparation of the dictionary and linguistic routines, will be presented.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME Erhard O. Lippmann

INSTITUTION IBM T. J. Watson Research Center, Yorktown Heights, New York

Use following space for an abstract or summary of your project

Title of Project Experimental On-Line Computer Aids for the Human Translator

An exploratory computer-aided translation system is being developed which basically consists of storage and retrieval operations carried out on line with a computer during the time in which a translation is produced. The system is not programmed to simulate the human translator by producing automatic translations. Rather, the user can call upon the computer's resources as needed in the translation process to shorten the delay between the initiation of a translation and its finished version. A combination of terminals, computer devices, and software is used to perform functions which have habitual human counterparts of a mechanical nature, e.g., dictionary look-up, dictionary updating, creation of terminological digests (i.e., test related mini-dictionaries), semi-automatic editing generation of cross reference files, text statistics, printing and lay out, and automatic combination insertion, deletion, or duplication of text.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME Automated Language Processing Project; Dr. Eldon G. Lytle, Director

INSTITUTION Brigham Young University, Provo, Utah

Use following space for an abstract or summary of your project

Title of Project Automated Language Processing Project

The Project emphasizes the refinement of computer-assisted translation, as opposed to fully automatic translation, and has devised for this purpose techniques of man-machine interaction which utilize the human for those aspects of the translation task requiring human intelligence and the computer for those aspects of the translation task which can be managed mechanically. Junction Grammar, a new theory of language structure which captures linguistic universals hitherto unknown, serves as the basis for the system.

Phase I of the development (now operational) provides computer editing, file management, and dictionary lookup. Phase II of the development provides computerized analysis, transfer, and synthesis of sentence structure (implementation 1978-79). Proto-type systems are designed for translation from English to Spanish, French, German, and Portuguese, but the method is equally adaptable to any combination of source and target languages.

The primary sponsor of BYU ALP is the Church of Jesus Christ of Latter-day Saints (Mormon), which annually translates approximately 17,000 pages of material into more than fifty (50) languages. It is planned that dictionary lookup and linguistic processing will initially be accomplished at a large central installation. The output of this processing will then be forwarded on "floppy" disks to regional translation centers around the world where residual aspects of the translation and printing task will be accomplished with the aid of mini-computer work stations.

The Project has a staff of 12 full-time and 18 part-time researchers.

NAME Roger C. Schank

INSTITUTION Yale University

Use following space for an abstract or summary of your project

Title of Project Computer Understanding of Text

Research at Yale centers around the building of computer programs that will understand stories. Two program are currently being developed, SAM and PAM.

SAM is composed of the following

1) an analyzer that maps English into a deep conceptual representation.
2) a script applier that uses its knowledge of contexts to supply missing or or implicit inferences about a situation.
3) a memory that finds references for things that it knows about in a text so as to bring its knowledge to bear on the text.
4) a generator that reads information provided to it by (1), (2), and (3) and states that information in English, Chinese, Russian, Dutch or Spanish.
5) a question answerer that interacts with the script applier to answer questions about an input text.

SAM is capable of mechanical translation, automatic summary and paraphrase and question-answering about texts in domains that it has knowledge about.

PAM is like SAM except that it does not have a script applier but instead has a more general mechanism that to infer the goals and intentions of the actors in the stories it hears.

Both of these programs are beginning approaches to the problem of computer understanding.

NAME Robert J. Shillman, Ph.D.

INSTITUTION M.I.T.

Use following space for an abstract or summary of your project

Title of Project Optical Character Recognition Based on Phenomenological Attributes

A theory of character recognition has been proposed and a methodology has been developed which is expected to yield a machine algorithm that will equal human performance in the recognition of isolated, unconstrained, handprinted characters. The methodology is based on the study of ambiguous characters, characters that can be assigned two letter labels with equal probability, rather than on letter ardhetypes. A description of the underlying representation of each of the 26 upper case letters of the English alphabet was obtained through analysis of ambiguous characters which were generated for this purpose. The descriptions are in terms of an abstract set of invariants, called functional attributes, and their modifiers. The relationship between the physical attributes, derived from physical measurements upon a character, and the functional attributes is given by a set of rules called Physical to Functional Rules. Three different techniques for determining these rules through psychophysical experimentation have been tested, and the particular rule for the attribute LEG has been determined. The remaining rules can be obtained in a similar fashion, and the combined results are expected to provide the basis for a machine algorithm. We are currently investigating the Physical to Functional Rules for the remaining attributes and are also interested in the way in which the rules are to be combined.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME Robert F. Simmons

INSTITUTION University of Texas, Austin, Texas

Use following space for an abstract or summary of your project

Title of Project TEXT INFORMATION SYSTEMS

A developmental program is proposed to create a socially useful system that will integrate several existing natural language processing procedures into a robust, transportable, General Text Understanding System for eventual use in applied information centers. The proposal is comprised of seven tasks: 1. Continued development of quantified case predicate forms of conceptual memory structure. 2. Integration of question answering and problem solving procedures. 3. Development of a human-aided, multi-pass, text-to-memory compiler. 4. Generation of natural language outputs for summaries, abstracts, expansions, translations, etc. 5. Generation of special purpose text teaching materials. 6. Implementation of natural language dialogue capabilities. 7. Development of a textword management system for linguistic analysis, retrieval and lexicon development.

The work will be accomplished on a DEC10 to enhance the transportability and communication of documentation for the resulting system.

NAME Dr. Peter Toma, President and Chairman of the Board

INSTITUTION LATSEC, Inc. and World Translation Center, Inc.

Use following space for an abstract or summary of your project

Title of Project SYSTRAN

After having developed the SERNA, AUTOTRAN, and TECHNOTRAN machine translations, I felt that the advent of third generation computers provided the long-awaited opportunity to develop a large-scale, yet fast and economical, systematically planned, unified, universal system. That system is SYSTRAN, whose name is an acronym formed in 1964 from "systems translation."

SYSTRAN is a fully operational machine translation system which can be installed at any IBM 360/370 site within hours. It has been used by the Air Force (translating 15 million words a year) since 1970 and by NASA since 1973. SYSTRAN is fully automatic, requires no human intervention nor pre-editing. It translates between the following language pairs at a speed of 300,000 words per hour: Russian-to-English, English-to-Russian, English-to-French, German-to-English, and Chinese-to-English. We term it a universal translation system because of this and because of the ease with which new translation capabilities can be added.

SYSTRAN's success is due to its strong and very flexible software frame, which allows the immediate implementation and testing of linguistic hypotheses, as well as universality in handling natural languages. Moreover, its special macro language allows linguists to program their own rules. The system can be modified or expanded to any limit at any time. It can never become a "black box."

The complete SYSTRAN package includes all utility programs, dictionary creation and update subsystems, source language analysis programs and target language generation programs, as well as programs for development of frequency listings, concordance materials etc. There are separate dictionaries for stem entries and idiomatic expressions (which are also entered in stem form). Because lexical items are entered in stem form, and because of a complex cross-referencing system, it is necessary to enter any lexeme only once, accompanying it with paradigmatic set information.

Source language analysis programs begin at homograph resolution proceeding through establishment of immediate constituents (IC's), to establishment of syntactic relationships of IC's and establishment of clause types and clause boundaries. Semantic analysis is used not only in selecting proper target language meaning equivalents, but also in establishing certain syntactic relationships.

Target language generation includes structural transformations, synthesis of target language word forms including insertion of auxiliaries, prepositions, etc., and rearrangement within clauses to achieve standard target language word order.

Further development of pronoun translation and prepositional functi

NAME William S-Y Wang

INSTITUTION University of California, Berkeley

Use following space for an abstract or summary of your project

Title of Project Project on Linguistic Analysis

Research on machine translation from Chinese to English under the direction of William S-Y Wang was carried on at the project on Linguistic Analysis (University of California, Berkeley) during the period 1967 to 1975. During the early part of the effort, System I was developed which includes: a) CHIDIC: A Chinese to English machine dictionary of about 80,000 entries (60 percent physics, 30 percent biochemistry, and 10 percent general), and b) Monolithic grammar of about 4,000 rules (context-3, phrase-structure rules). In 1973, two factors caused redesign of the approach toward the development of System II. One, the grammar had become so cumbersome and ad hoc that its effectiveness as well as its potential for improvement were curtailed. Second, the sponsor requested conversion of the system from CDC machines to IBM machines. In response to these factors, System II is designed along the lines of "structured programming" (i.e., it is built on self-contained program modules). It is also designed to be machine-independent, so that it can be implemented at different computer installations.

Efforts in research and development have been aimed at an operational system. We have experimented with numerous trial sentences as well as several "live" texts (from articles of 3,000 characters in length) and have accumulated machine texts of over 560,000 characters. System II is incomplete, lacking especially the machine-editing of output to conform to those morphological features absent in Chinese but required in English.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME Yorick Wilks

INSTITUTION Dept. of Artificial Intelligence, Univ. of Edinburgh, UK.

Use following space for an abstract or summary of your project

Title of Project An AI Approach to MT

The present system takes in paragraphs of English on line and outputs paragraphs of French. It is very small with a vocabulary of about 5-600 word senses, but that is very large for a project of this sort. By "this sort" I mean projects that aim for some deep semantic representation of the input language and from which the translation is produced. There is no separable syntactic stage in this work; the text is fragmented (into clause and phrase length pieces) by the program, and semantic structures are attached directly to these. These semantic structures are called templates and correspond to "mini-assertions." That is to say, the program seeks to display the input as a sequence of mini assertions. These templates are constructed out of formulas, already available from a dictionary, for the word sense of the input. Each word sense has a formula for it, and much of the work in the program is ascertaining what is the correct word sense (and so correct formula) for an input word. The formulas are tree structures built up out of different types of semantic primitive. A formula has internal rules operating on these semantic primitives that enable it to express the meaning of the corresponding word sense. Once the templates have been formed up, various kinds of inference rules operate on them, to produce deeper semantic representations, so as to resolve remaining ambiguities of word sense, prepositions or pronoun reference. When a clear single temple representation has been obtained, a French representation can be generated from it.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME MICHAEL ZARECHNAK

INSTITUTION Georgetown Univeristy School of Languages and Linguistics

Use following space for an abstract or summary of your project

Title of Project Georgetown University General Analysis Techniques-(GAT)-Simulated Linguistic Computer(SLC)

The Georgetown University Russian-English System is running on IBM 360/70 .CPU time for 2000 words @ 9 seconds. The texts translated include scientific,technological,and economic materials.
M.Zarechnak in close cooperation with the linguistic research staff.
The linguistic statements are coded in symbolic language designed by Dr. A.Brown ('SLC'-Programming Language). Input/output is in Assembler language.
A dictionary entry contains a split or unsplit Russian stem, grammatical coding, lexical number, and English part. The clustered entries are recognized through special local operations when the calling signals occur within the sentence under processing.
Syntactic analysis is partly based on morphosyntactic markings and partly on semantic coding.
Users: Primarily scientists at ORNL. Users' comments essentially favorable.
The unddited translation is used primarily for information purposes, although in a few instances, the translations were post-edited when the user requested it.
The quality of the present translation is the same as it was in 1964.
No linguistic improvements were inserted in the system although there are some linguistic programs ready to be inserted.
The semantic level will be added. Its underlying procedures are based on the semantic collocational and colligational distributional patterns, as observed in the real corpora, with such generalization as these corpora would suggest. It is hoped that after large corpora will be described both semantically and analytically, then some theories might be developed and tested deductively for the improvement of the next MT cycle. Each sentence is scanned from the left to the right, and from right to left at least forty times, following a path of certain priority-based strategies. All these scannings in both directions are grouped into four levels: word recognition, syntagmatic,syntactic, and synthesis of English. Some parts of the synthesis are independent of the Russian input.
Size of the dictionary: @50,000 stems.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME S. C. Loh

INSTITUTION Chinese University of Hong Kong

Use following space for an abstract or summary of your project

Title of Project Chinese University Language Translator (CULT)

The Chinese University Language Translator (CULT) is a Chinese-English computer-translating system which is unique in that it utilizes pre-editing of the source text as opposed to post-editing of the target text. The system is essentially a "pragmatic" one, in that the rules for handling complex strings previously requiring pre-editing are introduced as needed. CULT is made up of four modules: Dictionary look-ups, Syntactic Analyzer, Semantic Analyzer, and Output. Among these, the most limited is the Semantic Analyzer, which seems to rely more heavily on pre-editing than the other modules.

CULT is currently being used to translate two Mainland scientific journals, ACTA Mathematica Sinica and ACTA Physica Sinica. The computer, however, is not a person, which means it cannot experience the feeling of the original text and is not of much use for translating literary ~~material~~ essays. Nevertheless, scientific and some non-scientific works are also within the scope of its capabilities.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5

NAME Jim Mathias

INSTITUTION CETA (U.S.A.)

Use following space for an abstract or summary of your project

Title of Project Chinese-English On-Line Retrieval

The CETA (Chinese-English Translation Assistance) group is building a machine readable dictionary file for use in on-line retrieval and for development of dictionaries and indexes for use of human translators. The experimental on-line retrieval system can store an unlimited number of entries. The current file of 640,000 machine readable entries is divided into approximately 110,000 general entries, 10,000 colloquial entries, and 500,000 scientific and technical Chinese-English entries. The experimental system designed for an IBM 360 illustrates the facility of computer storage, retrieval, and display of Chinese characters and Roman alphabet as well as other scripts. It also illustrates the facility of computer techniques for indexing Chinese characters and special adaptability for synthesizing Chinese queries to search telecode sorted files.

NAME Friedrich Krollmann

INSTITUTION Federal Bureau of Languages

Use following space for an abstract or summary of your project

Title of Project FRG Translation Aid System

Germany's Federal Bureau Computer Translation Aids System contains over 700,000 foreign language (English, French, Russian, and Portuguese)-German entries of a technical and scientific nature. These entries can be accessed in a number of different ways depending on the needs of the user. Thus, the programming of the system allows for more specialized foreign language-German glossaries and lexical concordances, as well as linguistic analysis and frequency counts on the technical vocabulary of a given language.

Approved For Release 2008/03/03 : CIA-RDP80T00294A001200010014-5