Looking back at a year of German-English MT with Logos

Klaus E. Tschira

SAP GmbH, Walldorf, Federal Republic of Germany

The usefulness of a machine translation system can only be evaluated in the light of the particular features of the texts and the translation environment. After a discussion of these factors, the effectiveness of human translation (HT) is compared with that of machine translation (MT) using the Logos system. For the particular environment and quality requirements of this first German Logos customer, HT is still more effective than MT with high-quality post-editing. Samples are given in an Appendix.

INTRODUCING SAP

SAP GmbH is a software company with its head office in Walldorf, Federal Republic of Germany. During its first twelve years of existence, SAP has specialised in the development, marketing, installation and maintenance of comprehensive and integrated software systems covering fields such as

- financial accounting
- assets accounting
- cost accounting
- purchasing
- stock management and store location management
- invoice checking
- production planning and control
- order processing and distribution
- personnel management.

The above applications are available in various versions for

leading mainframe computers and several database/data communication systems. There are some 450 installations of SAP systems, most of them in German-speaking countries. The main reason why there are only a few in other countries is the fact that the system documentation used to be available only in German. In 1982, SAP decided to prepare for expansion into non-German-speaking markets by having the documentation translated into English.

THE TEXT MATERIAL

Type, structure, lifetime of texts

The texts to be translated make up several thousand pages of documentation. There are some rather technical texts such as data descriptions, programming handbooks and installation manuals, while others are purely application-oriented (e.g. functional descriptions); the largest and most difficult part of the documentation, however, deals with the representation of data structures and functions to be found in the application areas, mentioned above, within the software system. These user's guides and organisation manuals are inevitably written in a mixture of application terminology and data processing terminology. The texts themselves are usually highly structured with multiple levels of indentation, and embedded tables and graphic representations.

The manuals are being continuously revised and amended to match the expanding scope of functions and to reflect the enhancements of the software systems; eventually, after some years, a particular manual may be completely rewritten.

Problems of terminology and style

With texts on topics from the field of engineering and the natural sciences, we usually find a large number of strictly defined terms; in most cases, these terms are strictly applied to concepts which have been agreed upon internationally. This is also true for data processing terms, but problems begin when one talks about embedding the application systems into different data processing environments (i.e. operating system, data or database management system, and data communication system). Here we find diverging and at the same time overlapping sub-terminologies, sometimes due to different concepts, but perhaps also cherished for competitive reasons.

Terminology issues become complicated in business-oriented texts, for a number of reasons:

- business terminology lacks, at least partly, strictness of definition and usage; many companies have their own house terminology, even for common concepts.
- there are some common concepts that should be translated differently for English and American readers.
- there are some national concepts which are not known in countries of the target language; they deserve special treatment.

Our documentation is usually written by authors who are expert systems analysts and/or programmers, but not usually expert writers. Hence, many texts leave something to be desired as far as style and accuracy of expression is concerned ('Well, everyone knows what I mean by that'). As most of the proofreading was originally done by the authors themselves, the shortcomings came to light, for the most part, only when we were faced with the task of translation.

Quality requirements

SAP does not intend to become known as 'that software house with the "Germish" documentation'. Our manuals are read thoroughly once, for the purpose of becoming familiar with the software systems, but are afterwards used repeatedly for reference. While this type of usage is primarily for information purposes only, we feel that style is nonetheless an important factor in 'getting the message' to the reader. Failure to achieve rapport immediately leads to a decrease in acceptance, as well as an increase in costs for trouble-shooting and maintenance.

Hence high-quality translation is required for 90 per cent of the texts. In fact, some texts have been rated higher in the English translation than their German original.

TRANSLATION
Environment

Because of the increasing volume of texts, the permanent process of revision and amendment, and the possibility for co-ordinated feedback between authors and translators, our first decision was in favour of in-house translation. At the same time, we installed a Wang OIS word processing system to replace our previous, out-of-date text processing system.

Wang introduced us to Logos, and we quickly decided to give it a try, becoming the first Logos customer in Germany; we installed the Wang and Logos systems in December 1982.

All texts to be translated are already available as text documents on the WP system.

<u>Human translation (HT)</u>

SAP's translation staff comprises three professional translators (one of whom is a native speaker of English). They were carefully inducted in October 1982.

As outlined above, terminology is difficult to research, but the number of terms used is not too large and is easy to remember. Proper transfers are established by teamwork - in order to find the terms least ambiguous and most acceptable to the international readership - by using several bilingual and monolingual dictionaries and also by reading English and American textbooks on the topics concerned. Practically no time is spent during actual translation on looking up pre-established terms. The most efficient way for HT is to translate directly into the word processing system, typing over the German original text with the English translation. In this way, no additional effort for text formatting is required since all indents, tabs, blank lines and page-breaks are directly taken over into the target text.

The speed of human translation, excluding terminology research and questioning of authors, etc., but including proofreading by the translator and text formatting in fit-to-print quality, is roughly two pages per hour.

<u>Machine translation (MT) and post-editing (PE)</u>

Being the first German Logos user, we were quite prepared to encounter some initial difficulties, and indeed there were some. We are glad to state that all of them were analysed immediately, and most of them were also removed within a reasonable time.

During early 1983 attention was focussed on operational stability of the system and on building up the dictionary. Later attention shifted to improvement in the quality of the generated language and the preservation of any text formats.

The quality of the MT system must be assessed in the light of the following six aspects:

(i) <u>Dictionary maintenance</u>. The time required per entry is usually under one minute and hence negligible as compared with the time used for terminology research. ALEX, the 'Automatic LEXicographer' provides excellent user guidance

by giving samples and explanations for all syntactic and semantic specifications necessary. Hence no special linguistic experience is required for dictionary maintenance; ALEX can be handled well by 'interested and willing laymen'.

(ii) <u>Pre-editing</u>. In the early versions, a little pre-editing of format was required. Headings - embedded in the text or grouped in a table of contents - were to end with a full stop and the total input text was to end with a page-break. These tasks are ridiculously simple, but require scanning through the source text one extra time, provided the punctuation can be assumed to be correct. Fortunately, this requirement was removed completely. End-of-line, colon and semicolon are now functionally equivalent to the full stop.

Another type of pre-editing is still required: source texts must be de-hyphenated. This is closely related to the fact that, on the Wang OIS system, the normal hyphen, the dash and the syllabification hyphen are represented by the same character; by the way, this is also the reason for inadequate handling of the syllabification hyphens on the Wang WPs.

(iii) <u>Recognition of source-text constructs</u>. In April 1983 I performed some systematic tests on recognition of various forms of verbal phrases (with or without, or with optional, single/double objects). Most of the sentence patterns from Wahrig's <u>dtv-Wörterbuch der deutschen Sprache</u> (dtv Dictionary of the German Language) were interpreted satisfactorily. Exceptions were often found to be rather old-fashioned or poetic constructs. Phrases embedded in a normal sentence using parentheses or hyphens (such as this phrase) are not always recognised correctly and sometimes result in distorted output text or text with little meaning.

(iv) <u>Quality of generated language</u>. This, of course, is the most interesting and most important aspect. First tests with a small volume of general text showed surprisingly good results. Also, during the initial training of our translators, some thirty pages of a more general text were processed; a few of them are said to have been so good that little or no post-editing was required because of the text quality. But generally speaking, quality was strongly dependent on the style of the source-text authors. Most of these difficulties stemmed from different habits of using source-text constructs that were not correctly interpreted in the early versions (see above).

Another source of difficulty lies in missing dictionary entries. Here I do not mean words that are not in the dictionary at all. Of course one cannot expect sentences to be translated correctly when words in essential phrases are

missing, and semantic analysis is thus seriously impaired; not only are the missing words then left untranslated, but also the transfer of phrases with known words is of poorer quality. This situation can be bypassed by running the 'New Word Search' and adding any missing words to the dictionary before attempting MT. But the same symptoms (nonsense or, better, 'low-sense' output text) may also arise with words which are in the dictionary, but which have completely different meaning and semantic implications. Such words obviously cannot be brought to light by 'New Word Search'. The resulting low-sense output requires careful inspection to detect the reasons for the poor quality, which might alternatively result from an inability to analyse the syntactic structure of the source text. Since a translator's first objective is to produce translations, and not to improve a tool, we passed many samples of low-sense output onto the Logos support staff. It was remarkable to see the quality improvements brought about by subsequent new releases and considerable dictionary amendments. Now (at the end of 1983) we observe an output quality which we estimate to be good enough for preliminary information purposes, without post-editing, for about sixty to eighty per cent of the text (see the raw translation in the Appendix; major nonsense is produced only where the full stop in the abbreviation 'sog.' of 'sogenannt' (= 'so-called') is wrongly interpreted as marking the end of a sentence).

Unfortunately, the reader is very easily misled by this apparent good quality and underestimates the post-editing time still required to produce high-quality text from the MT output (see below). Major improvements in the area of proper positioning of words have been announced and we are eagerly looking forward to them. We expect additional benefits from the possibility of entering multi-word phrases into the dictionary, which is also announced for late 1983/early 1984.

(v) <u>Preservation of text formats</u>. This issue is of little interest to the linguist, but is extremely important to the translator as it eventually accounts for a considerable, and especially frustrating, part of the post-editing time. Consider Table 1, which shows some progress with respect to text formatting.

Texts, such as this one, which contain embedded tables with text arranged in columns cannot be translated effectively since the columnar structure cannot be formally represented on a Wang OIS. Tables may be excluded from translation by using '+-' at the beginning and end of the text concerned ('translation on' and 'translation off'). The text between the two '+-' indicators is transported, unchanged, from the source language input to the target language output. It is

	early Logos releases (1983)	later Logos releases (1983)
end-of-sentence indication	full stop only	end-of-line, full stop, colon, semicolon
option of leaving parts of text untranslated	not provided	use '+-' to indicate start and end of text not to be translated
multiple levels of tabs and indents	completely ignored	preserved in summer release but containing errors again in autumn release
multi-sentence paragraphs	separated into multiple paragraphs containing one single sentence	preserved

Table 1.   Progress of text formatting with Logos

most easily translated manually by typing over the correctly formatted source text.

(vi)   <u>Post-editing</u>.   Let   us   assume   that   the   present difficulties   with   multiple   tabs   and   indents   will   be   overcome   as they   were   before.   Then   post-editing   is   required   only   to improve   the   language   quality   and   to   correct   any   ambiguities either   already   contained   in   the   source   text   or   arising   in   the target   text.   If   translations   are   made   for   internal   use   and for   information   purposes   only,   post-editing   might   be   omitted or   reduced   to   a   low   minimum   in   the   near   future.   In   order to   produce   high-quality   translations   of   texts   dealing   with complex   subject-matter,   however,   extensive   post-editing   is required.   On   the   first   reading   of   the   MT   text,   one   is   often misled   into   underestimating   the   time   required   for post-editing,   because   one   tends   to   neglect   the   number   of trivial   juxtapositions   of   words   or   phrases.   Another   limiting factor   is   -   for   complex   texts   -   the   double   task   of understanding   both   source   and   target   texts   during

post-editing before actually matching their meanings mentally. For HT, there is only the act of understanding the source text. In tests we have carried out, the times taken to post-edit MT texts and not-quite-perfect HT texts to a high quality have been equal.

Throughput

In a small translation department with a heavy workload, one cannot have benchmarks too often and one cannot repeat them regularly with the next release of the MT system. So maybe our experience is not statistically relevant. Moreover, one should keep in mind the nature of texts and the high-quality requirements outlined above. For one type of text (data description manuals), we observed an increase in throughput of 30 per cent:

$$\text{human translation: post-editing} = 10{:}7$$

Texts of this type typically contain many phrases used repeatedly throughout the text and have comparatively low complexity. For more complex and formatted texts like that in the Appendix, the times required for HT and for the post-editing of MT texts by translators are close to equal; PE is slightly faster if performed by the author of the source text, provided he/she has enough language competence (authors refer less often to the source text). Unfortunately, it cannot be predetermined quickly and easily whether a given text is more favourable for MT and PE or for HT. HT times, however, can be estimated more accurately, so that when under pressure our translators prefer HT.

HOPES AND PLANS FOR THE FUTURE
Enhancements to be expected

Every effort should be made to improve the quality of the target-language texts generated. This is a task for the Logos team, as far as proper arrangement of words, etc. is concerned. It is also a task for the Logos user, as far as source-text ambiguities are concerned, if production of source text can be controlled. Formats should be preserved exactly. Tools for faster rearrangement of words or phrases should be provided (Ian M. Pigott has told me that the European Commission in Luxembourg has some ingenious glossary functions). Improvement of hyphenation (see above) in Wang WPs could eliminate the need for pre-editing (de-hyphenation).

Style of source text

It is highly advantageous to control the style of source texts, if possible. This does not at all postulate an oversimplified or artificial style. The same stylistic sins slow down both HT and MT/PE. Moreover, I have a strong feeling that the same sins also reduce the comprehensibility of texts for the ultimate reader. There are some ambiguities which typically occur in the German language, e.g. some adverbial phrases may refer to the verb or to its object, giving two possible interpretations which often have very different meanings. Being familiar with the subject-matter, the author, the reader, and possibly the human translator choose the right meaning unconsciously, but Logos (or any other MT system) and the uninformed reader presumably do not. Hence, by focussing on problems of translation, and especially on MT, I feel that I have gained a better feeling for clarity of style. A style guide should be designed, possibly in the form of a programmed instruction manual. Such a manual could be used to motivate and train original writers to write texts in a more understandable fashion and, at the same, write them in such a way that they are less ambiguous to natural and artificial intelligence.

Rearrangement of work flow

With an increasing quality of MT output, the limiting factors for HT as well as for PE speed will increasingly be:

- the style of the source text;
- the time required for the human translator or the post-editor to understand the source text.

Considering the special conditions at SAP (outlined above), we have decided to

- have the translators participate in some of the training given to new systems analysts and programmers;
- integrate the translators into the process of formulation of the original text or at least the process of revision.

Thus, the job description of a translator will be enlarged and enriched to cover comprehensibility of documentation in both the source and the target languages.

Terminology research and dictionary maintenance would remain the translator's responsibility. With the use of an improved MT system, translators would normally post-edit MT output. In times of extraordinary peak loads or bottlenecks due to illness or vacancies for translators, PE - but not

HT - could also be performed economically by some authors with sufficient language competence.

CONCLUSIONS

For texts up to an intermediate level of complexity, for texts with a large, but strictly applied, volume of terminology, and for medium quality requirements, post-editing of Logos MT output may be faster than HT. For information purposes, very rapid post-editing may be performed. For texts with a high level of complexity, such as interdisciplinary texts, for texts with comparatively few technical terms, and for high quality requirements, human translation, post-editing of MT texts and revision of poorer human translation take approximately the same time.

Further investment in the quality of MT output and in controlling the production of source text - wherever possible - is necessary in order to reach the break-even point. There are some indications that such improvements may be expected. Controlling the source-text production will give increased return on investment if and when translations into several target languages are made from the same source text.

AUTHOR

Klaus E. Tschira, SAP GmbH, Max-Planck-Strasse 8, Postfach 1348, D-6909 Walldorf, Federal Republic of Germany.

APPENDIX

The Appendix contains two pages of a complex, formatted German source text, followed by four translations of it:

- source text (ST);
- unrevised Logos machine translation (MT);
- MT post-edited by German translator (PEMT1);
- MT post-edited by author of source text (PEMT2);
- human translation by native speaker (HT).

	Stand	Form	Blatt-Nr.
Systembeschreibung		*ST*	*1*

German Source Text

Im Geschäftsleben werden die Geschäftsvorfälle regelmäßig durch
Belege dokumentiert, wie etwa Buchungsbelege, Aufträge,
Bestellungen. Die in ihnen enthaltenen Daten werden - im Gegensatz
zu den Stammdaten - als "Bewegungsdaten" bezeichnet.

In herkömmlichen EDV-Anwendungs-Systemen wurden oft die Daten
eines Geschäftsvorfalls getrennt bearbeitet und gespeichert. In
den SAP-Systemen vermeiden wir diese unnatürliche Aufteilung und
speichern einen Geschäftsvorfall statt dessen einheitlich als sog.
"Dokument", um dem Sachbearbeiter jederzeit einen Überblick über
den gesamten Vorfall zu ermöglichen.

Je nach dem Typ eines Dokuments (wie z.B. Buchungsbeleg,
Bestellung, Auftrag) enthält es sehr unterschiedliche Daten. Allen
Typen von Dokumenten jedoch ist gemeinsam die Unterteilung in
Kopf- und Positionsdaten.

Im folgenden beschränken wir unsere Betrachtungen zunächst auf
Dokumente vom Typ "Buchungsbeleg". Alle Buchungsbelege haben einen
einheitlichen Belegkopf. Ihm untergeordnet sind stets mehrere
Belegpositionen. Jede Belegposition repräsentiert eine Buchung auf
ein bestimmtes Konto. Sie enthält demnach stets mindestens die
folgenden Daten:

- das zu bebuchende Konto
- den zu buchenden Betrag
- eine Kennzeichnung der Art der Buchung, die wir
  "Buchungsschlüssel" nennen.

Die Buchungsschlüssel können so gewählt werden, daß sie

- alle praktisch vorkommenden Arten von Buchungen umfassen und
- in Verbindung mit Konto-abhängigen Steuerdaten die
  eindeutige Festlegung sämtlicher Zusatzdaten erlauben, die
  im Zusammenhang mit der betrachteten Buchung anwendbar sind.

Die Zusammenfassung aller überhaupt in Buchungspositionen
vorkommenden Daten führt zur Bildung einer fiktiven maximalen
Buchungsposition. Die Felder aller realen Buchungspositionen sind
Teilmengen der Felder der (fiktiven) maximalen Buchungspositionen.

Systembeschreibung	Stand	Form ST	Blatt-Nr. 2

Eine feldweise Steuerung des Auftretens der verschiedenen Felder ist glücklicherweise nicht nötig. Es gibt *Gruppen* von Feldern, die stets gemeinsam auftreten oder nicht auftreten und die wir deshalb zu sog. "Feldgruppen" zusammenfassen können. Tatsächlich genügt es, die oben erwähnte Steuerung (in Abhängigkeit von Konto und Buchungsschlüssel) auf Feldgruppen anstatt auf Felder zu beziehen, was eine erhebliche Vereinfachung bedeutet.

Die am Beispiel von Buchungsbelegen verdeutlichten Strukturen lassen sich in ähnlicher Weise auch bei anderen Belegtypen (wie Bestellungen und Aufträgen) aufzeigen. Durch Verallgemeinerung gelangen wir zu folgendem Schema der Belegverarbeitung und -speicherung:

- Kopf und Positionen eines Dokuments werden gemeinsam und vollständig erfaßt und stets gemeinsam gespeichert.

- Die Positionsdaten können verdichtet werden, so daß in jedem konkreten Einzelfall nur die tatsächlich anwendbaren Feldgruppen gespeichert werden. Hierbei entstehen ungeheuer viele Varianten.

- Für die Verarbeitung werden die Daten der verschiedenen Belegpositionen stets in einem einzigen einheitlichen und festen Format, der (fiktiven) maximalen Belegposition, bereitgestellt.

- Die Konvertierung einer Position zwischen der verdichteten Form (die für die Speicherung besonders wirtschaftlich ist) und der expandierten Form (die eine bequeme Verarbeitung ermöglicht) geschieht durch allgemeine, anwendungsneutrale Routinen, die in der Schnittstelle zur Datenbank-Verwaltung enthalten sind (siehe Abb. 1).

Das dargestellte Prinzip der Belegverarbeitung bietet für die Finanzbuchhaltung entscheidende Vorteile:

- Aus dem Belegzusammenhang kann jederzeit der gesamte Geschäftsvorfall überblickt werden.

- Die belegweise Erfassung ermöglicht einige Konsistenz-Prüfungen. Die bekannten Probleme, die beim zeitlich getrennten Buchen von Geschäftsvorfällen in Form von Einzel- und Sammelbuchungen auftreten, können vermieden werden; die belegweise Verarbeitung bildet demnach die Grundlage für eine jederzeit formal selbstkonsistente Buchhaltung.

	Stand	Form	Blatt-Nr.
Systembeschreibung		*MT*	*1*

unrevised
LOGOS output

In the business-life, the business transactions are documented
regularly by document/vouchers, as for example posting documents,
orders, orders. The data included in them are designated as
movement data - in contrast to the master data -.

The data of a business transaction were often processed and stored
separately in conventional EDP-application-S-systems. In the
SAP-systems, we avoid this unnatural breakdown and store a
business transaction instead of which than sucked uniform. At any
time, to make possible document, about the person in charge an
overview of the entire occurrence.

According to the type of a document (such as, e.g., posting
document, order, order), it includes very different data. The
subdivision into header-and item data is common to all types of
documents however.

In following we limit our considerations first on of document of
the type posting document. All posting documents have an uniform
document header. Several document items are subordinated to
it/him always. Every document item represents a booking on a
certain account. It/she accordingly includes the following data
always at least:

- That to BEBUCHENDE account
- The amount to be posted
- A qualification of the type of the booking which we of
  posting keys call.

The posting keys can be selected in such a way that it

- Include all practically occurring taking of bookings and
- Permit the unambiguous determination of all additional data
  in connection with account-dependent control data, which in
  conjunction with the regarded booking, are usable.

The summary of all data occurring actually in booking-items
conducts to develop a fictitious maximum booking-item. The fields
all physical booking-items are subsets of the fields of the
(fictitious) maximum booking-items.

	Stand	Form	Blatt-Nr.
Systembeschreibung		*MT*	*2*

A field-by-field control of the occurrence of the different fields is luckily not necessary. There are groups of fields which occur always commonly, or occur not and which we sucked therefore too. Field-groups are able to combine. In fact, it suffices to relate the control mentioned above (depending on account and posting key) to field-groups instead of on fields what a considerable simplification means.

The structures clarified as exemplified by of posting documents could be demonstrated in similar manner also with other document types (as orders and order). By the generalization, we were successful to the following scheme of the document handling and storage:

- Header and items of a document are input commonly and completely and always stored together.

- The item data can be condensed so that, in each, only the field-groups usable in fact are stored to actual individual case. With this, many variants result tremendously.

- For the processing, the data of the different document items are readied, always in a single uniform and fixed format, the (fictitious) maximum document item.

- The conversion of an item between the condensed form (which is especially economic for the storage) and the expanded form (which what is to be a comfortable processing) takes place

The represented principle the document handling offers advantages decisive in favor of the financial accounting:

- At any time, the entire business transaction can be surveyed from the document/voucher-context.

- The documentwise input makes possible a few consistency-checks. Occur the well-known problems which business transaction in the case of the posting separated within a given time of/by in the form of individual-and collective postings, can be avoided; the documentwise processing accordingly forms the basis for an at any time formally self-consistent accounting.

	Stand	Form	Blatt-Nr.
Systembeschreibung		*PEMT₁*	1

Post-edited MT
(by German Translator)

In everyday business, the business transactions are regularly
documented by vouchers, as for example posting vouchers, customer
orders, purchase orders. The data included in these vouchers is
called movement data, as opposed to the master data.

In conventional EDP application systems, the data of a business
transaction was often processed and stored separately. In the SAP
systems, we avoid this unnatural breakup and store a business
transaction in a uniform way as a so-called "document". The user
is thus given an overview of the entire transaction at any time.

Depending on the type of a document (such as posting voucher,
purchase order, customer order), the document contains very
different data. The classification into header and item data,
however, is common to all types of documents.

In the following, we will limit our explanations to documents of
the type "posting vouchers". All posting vouchers have a uniform
document header. Several document items are always subordinate to
the document header. Every document item represents an entry to a
particular account. It thus always contains at least the
following data:

- the account to be entered
- the amount to be posted
- an indication of the type of posting which we call the
  "posting key".

The posting keys can be selected in such a way that they

- include all existing types of posting which actually occur
- allow the unique allocation of all additional data which can
  be applied with the posting concerned, in connection with
  account-dependent control data.

The summary of all data actually occurring in posting items leads
to the definition of a fictitious maximum posting item. The
fields of all real posting items are subsets of the fields of the
(fictitious) maximum posting item.
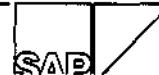
Systembeschreibung	Stand	Form PEHT 1	Bak-Nr. 2

Fortunately, a control of the occurrence of the different fields
(field by field) is not necessary. There are groups of fields
which always occur together or do not occur and which we can
therefore combine to so-called "field groups". In fact, it is
sufficient to relate the control mentioned above (depending on the
account and posting key) to field groups instead of to particular
fields; this would simplify the matter considerably.

The structures explained with the example of posting vouchers can
also be shown in a similar manner with other document types (such
as purchase orders and customer orders). A generalization gives
us the following pattern for the processing and storing of
documents:

- Header and items of a document are entered together and in
  their complete form and are also always stored together.

- The item data can be condensed so that in each individual
  case only those field groups ar stored which will acutally
  be used. This gives a tremendous number of various
  possibilities.

- For processing purposes, the data of the various document
  items are always made available in a single, uniform and
  standard format, i.e., the (fictitious) maximum document
  item.

- The conversion of an item between the condensed form (which
  is especially economical for storing purposes) and the
  expanded form (which makes a comfortable processing
  possible) is made by means of general routines, independent
  of the application. These routines are included in the
  interface to the data base management (see figure 1).

The principle of document processing, as explained above, offers
decisive advantages for financial accounting:

- The entire business transaction can be grasped from the
  document context at any time.

- Some consistency checks are made possible since the
  documents are entered one by one. The well-known problems
  which occur when business transaction are posted at
  different times in the form of single and collective
  postings can be avoided; the procedure of processing
  document by document thus forms the basis for an accounting
  system which is formally self-consistent at any time.

time: 70 min

Stand	Form	Blat-Nr.

Systembeschreibung _____ *Text*) _____ *FEHT 2*   -1

Post-edited MT Source Text
(by author of

Business transactions are regularly based on documents, such as
vouchers, purchase orders, sales orders. The data included in
them are designated as "movement data" in contrast to the master
data.

In conventional EDP application systems, the data of a business
transaction were often processed and stored separately . In the
SAP systems, we avoid this unnatural breakup and instead store a
business transaction in a unified way as a so-called "document" in
order to provide the person in charge with an overview of the
entire business transaction at any time.

According to the type of a document (such as, e.g., voucher,
purchase order, sales order), it includes very different data.
However, the subdivision into header and item data is common to
all types of documents.

In the following, we limit our considerations to documents of the
type "voucher". All vouchers have an uniform document header.
There are always several document items subordinate to it. Every
document item represents an entry to a particular account.
Accordingly, it always includes at least the following data:

- The account to be posted to.
- The amount to be posted.
- A qualification of the type of posting which we call
  "posting key".

The posting keys can be selected in such a way that they

- include all types of postings occurring in practice and
- permit (in conjunction with account-dependent control data)
  the unique determination of all additional data that are
  applicable in connection with the posting entry in question.

The combination of all data actually occurring in posting entries
forms a fictitious maximum posting item. The fields in all real
posting items are subsets of the fields of the (fictitious)
maximum posting item.

	Stand	Form	Best-Nr.
Systembeschreibung		*PEHT2*	*2*

A field-by-field control of the occurrence of the different fields is fortunately not necessary. There are groups of fields which do occur always together, or do not occur and which we can therefore treat commonly as a so-called "field group". In fact, it suffices to apply the control mechanism mentioned above (depending on account and posting key) to field groups instead of fields which means a considerable simplification .

The structures illustarted by the example of vouchers could be demonstrated in a similar manner also with other document types (as purchase orders and sales orders). By generalization, we may deduct the following scheme of the document handling and storage:

- Header and items of a document are entered together and completely and are always stored together.

- The item data can be compressed so that, in each individual case, only the field groups really applicable are stored. Many variants of compressed item layouts are generated this way.

- For processing, the data of the different document items are always provided in a single uniform and fixed format, the format of the (fictitious) maximum document item.

- The conversion of an item between the compressed form (which is especially economic for external and internal storage) and the expanded form (which enables a comfortable design of processing logic) is achieved by general and application-independent routines contained in the interface to the data base management (see Figure 1).

The principle of document handling outlined above offers decisive advantages in the field of financial accounting systems:

- The entire business transaction can be reconstructed from the voucher context at any time.

- Data input document by document makes possible several consistency checks. The well-known problems that occurr when business transactions are posted separately in the form of individual and collective per-day posting entries, can be avoided; hence, document processing forms the basis for an accounting system formally self-consistent at any time.

post-editing time
  (source author) 55 *min*
review by translator 10 *min*

	Stand	Form	Blatt-Nr.
Systembeschreibung	*kor*	*HT*	*1*

Human Translation by native speaker

In the business world, business transactions are recorded on a
regular basis in the form of documents, such as book-keeping
vouchers, customer orders, orders. The data contained in these
documents -- in contrast to the master data -- are termed as
"movements data".

In conventional EDP application systems, the data of a business
transaction have often been processed and stored separately. In
the SAP systems we avoid this unnatural splitting and, instead,
store a business transaction as a so-called "document" in order to
make it possible for the user to have a complete view of the
entire transaction at any time.

A document can contain very different data, depending on what type
it is (e.g., book-keeping voucher, order, customer order). Common
to all types of documents, however, is the subdivision of data
into header data and item data.

In the following, we will restrict our discussion for the time
being to documents of the type "book-keeping voucher". All
book-keeping vouchers have a uniform document header. Subordinate
to this there are always several document items. Each document
item represents a posting to a particular account. Accordingly,
it contains at least the following data:

- the account to which the posting is to be made
- the amount to be posted
- an identification of the type of posting, which we call
  "posting key".

The posting keys can be chosen in such a way that they

- include all types of postings which actually arise, and
- in connection with account-dependent control data, allow
  unique definition of all additional data which can be used
  in connection with the posting in question.

The grouping together of all data which occur in posting items
leads to the formation of a fictitious, maximum posting item. The
fields of all real posting items are partial quantities of the
fields of the (fictitious) maximum posting items.

	Stand	Form	Blatt-Nr.
Systembeschreibung		*HT*	*2*

Fortunately, it is not necessary to control, field by field, the occurrence of the different fields. There are groups of fields which always occur together or do not occur, and which we can therefore group together as so-called "field groups". Indeed it is enough to relate the above-mentioned control (dependent on account and posting key) to field groups instead of fields; this simplifies the matter considerably.

The structures illustrated by the example of book-keeping vouchers can also be pointed out in a similar manner for other document types (such as orders and customer orders). Through generalization we come to the following pattern for document processing and storage:

- Header and items of a document are recorded together and completely, and always stored together.

- The item data can be compressed so that in each actual case only those field groups are stored which can actually be used. Here there is an incredible number of variations.

- For processing purposes, the data of the different document items are always provided in a single, uniform and standard format, i.e., the (fictitious) maximum document item.

- The conversion of an item from the compressed form (which is especially economical for storage) to the expanded form (which makes for comfortable processing) is done by general, application-neutral routines which are contained in the interface to the data base management (see Fig. 1).

The above principle of document processing offers decisive advantages for financial accounting:

- It is possible to have a clear view of the entire business transaction at any time from the context of the document.

- Entry of each single document (i.e., document by document) allows certain consistency checks to take place. The known problems, which occur when posting business transactions at separate times in the form of single and collective entries, can be avoided; processing document by document thus forms the basis for a book-keeping system which is formally self-consistent at any time.

time: 60 min