USER ASPECTS OF AN AUTOMATIC AID TO TRANSLATION
AS EMPLOYED IN A LARGE TRANSLATION SERVICE

F. Krollmann

Bundessprachenamt

Summary

The present paper describes experience gathered in the employment of an automatic aid to translation on the basis of stored terminology, as practiced by the Language Operations Division of the Bundessprachenamt. The system in question has been in everyday use for over 10 years. The present paper attempts to weigh up the advantages and disadvantages, from the user's point of view, of the various possible ways of outputting computer-stored vocabulary. It investigates the criteria used for selecting and sorting the technical vocabulary in the computer in accordance with the various uses to which this vocabulary is put. It also discusses the wishes of the users as regards the structure of the data stored, In particular the input and output of definitions, and also goes into one aspect of the user's requesting techniques, viz. how to restrict the search to the subject fields he requires.

1.   <u>**PRELIMINARY REMARKS**</u>

That computers can effectively be used as an aid to translation, in particular for storing and processing terminology, is no longer a subject of dispute. Most of the major linguistic services aspire to a terminology data bank, already have one, or have a share in such an instrument. As we pointed out as early as in 1965,[1] a terminology data bank is ideally suited for use as an aid to the translator. Seen from the scientific viewpoint of computational linguistics, the problems posed by this by-product of the as yet unsuccessful attempts at automatic translation may already seem trivial. All known methods are essentially simple search procedures based on the comparison of characters but with more or less sophisticated features. Anybody who has closely followed the more recent scientific congresses in this field will know that the main focus of interest in the design of systems nowadays is turning towards the questions of user feedback and making life easier for the user.

It may therefore be of interest to consider here from the point of view of the users' experience a system which has been fully operational for over 10 years and which, along with Bachrach's system in Luxembourg, is the oldest of its type.

Allow me first to make two qualifications:

- Though the system used at the Bundessprachenamt was conceived as more than just a machine aid to translation and for a wider circle of users, its main user is the Language Operations

---

1) Krollmann, Schuck, Winkler: Herstellung textbezogener Fachwortlisten mit einem Digitalrechner. Beiträge zur Sprachkunde und Informationsverarbeitung, 1965, Nr. 5

Division of the Bundessprachenamt itself. This division of the organization has at its disposal all the technical facilities the system has to offer, from the on-line high-speed printer to operation in conversational mode. Smaller translation services will not have these facilities available to the same extent.

- The Language Operations Division is a large central translation service with nearly 50 revisor groups, each with translators specializing in some specific field, engaged in primarily technical or scientific translation. The translators working in this division nearly all have a thorough knowledge, for the most part acquired within the Bundessprachenamt, of the subject fields in which they work.

For these reasons, the appraisals in the following may only be valid mutatis mutandis for other translation services which do not have such technical facilities or such large numbers of specialized translators at their disposal.

By and large it is true to say that despite initial opposition on the part of the translators, who at first were sceptical towards the computer, the day-to-day work of our translation service would now be inconceivable without the system. The system does not by any means take the place of the translator's thought process; it rather relieves him of routine search operations. Neither has it made the use of other reference material at a higher level - e.g. monolingual technical handbooks - superfluous.

The system has proved its worth as an excellent instrument in simplifying the translator's work, in particular in co-ordinating major translating projects, especially if these are being carried on at a number of different locations. Here I must point out that we have to handle translating jobs of many thousand pages in which the vocabulary used must be uniform down to the last word.

2. <u>THE MACHINE-AIDED TRANSLATION PROCEDURE</u>

2.1. <u>The text-oriented glossary</u>

The machine-aided translation procedure followed in the Bundes-
sprachenamt is based on the use of the computer as a termino-
logy store, whereby the aspect of most importance to the
translator is the text-oriented glossary. Since most of
those present today will be familiar with the procedure, I
shall limit my description of it to the bare essentials.
The translator prepares a list of the technical terms
occurring in his text, ranging from single words to composita
of considerable length, in the order in which they appear in
the text to be translated. With a minimum of effort, i.e.
that entailed in excerpting the problematic terms or passages
and reducing them to dictionary reference form, he thus obtains
the terminological skeleton of the text he has to translate.
The fact that the computer will offer a whole range of possible
translations for some of his search words is by no means a
disadvantage; but more of that later.

The great advantage of the text-oriented glossary is that
the translator receives the answer to his search instructions
in the form of a hard copy printout on which he can then enter
his own comments which will be important to the revisor and
for the subsequent terminological analysis of the completed
glossary. The text-oriented glossary also guarantees a
constant feedback into the data base (new inputs and amend-
ments), since if the computer indicates that a search word
is not contained in its memory, the translator will supply
the missing term which can then be added to the terminology
bank.

For the person revising the translation, this part of the
procedure offers the great advantage that he can see precisely
the stock of terminology from which the translator has made

his final selection and can follow the translator's decision-making process. The disadvantage of the text-oriented glossary is that, for economic reasons, glossaries can, in practice, only be produced in batches. This means that there will always be a delay of a few hours between the translator preparing his list of search questions and receiving his printout from the computer, even if - as in our case - the user has his own ON-LINE printer. If the latter is not the case, and if the search list and printout have to be sent by mail, the time lost will often be so great that the glossary will no longer be of any use, especially if the translation in question has to be completed before a specific deadline. This applies, for example, to the external language services attached to the Bundessprachenamt. It follows, then, that the availability of an ON-LINE high-speed printer is practically a conditio sine qua non if really flexible use is to be made of the text-oriented glossary.

2.2.   Alphabetic lists

In addition to the normal text-oriented glossary, it is, of course, also possible to have the vocabulary requested delivered in alphabetical order. This alphabetic listing can take one of two completely different forms:

2.2.1   Alphabetic printout of text-oriented glossaries

Ready use of text-oriented alphabetic listings is made for the terminological co-ordination of translation projects which have been divided up amongst several translators. The terminology occurring in such translation jobs, which are sometimes of some considerable volume, is normally gathered for the project as a whole rather than merely for the various subject fields covered by the project. Such project-oriented (we call them problem-oriented) alphabetical terminology lists then form the basis for co-ordinating the

use of uniform terminology throughout the project. This is,
of course, of particular importance when such a large-scale
project is not only split up between a large number of the
Division's own translators but translators from the language
sections of enterprises in industry are also engaged in
work on the project. The procedure in this case is that the
text-oriented glossaries produced by the individual trans-
lators are compiled and re-generated to produce a comprehensive
alphabetical list which then serves the translators and
subsequently the revisers as an aid to co-ordination. These
lists can also be stored for future reference on an inter-
mediate storage device.

2.2.2 Alphabetic subject-field glossaries

These are an aid to translation in frequent demand, used
primarily by the smaller external language services attached
to the Bundessprachenamt. They take the form of bilingual
alphabetical lists of technical terminology extracted from
the master terminology bank on the basis of predetermined
subject codes. There is, however, considerably less demand
for subject-field glossaries within the Language Operations
Division of the Bundessprachenamt itself, as the restriction
to one or to just a few subject fields means that terminology
which is present in the computer but not coded as belonging
to the subject fields in question will not be printed out.
Subject-field glossaries of this type, which can often attain
a considerable volume, will, in my opinion, decline in im-
portance with the introduction of the COM procedure.

2.3 COM (Computer-Output-Microfilm) procedure

The COM procedure makes it possible to output the entire
terminology bank or any part thereof on microfilm cards,
thereby making it available to even the most remote trans-
lator who has no direct link to the computer. These micro-
films are relatively cheap to produce. Due to the fact that
a very large master terminology bank can be accommodated

on a relatively small number of microfilm cards (approx. 22 pages = 1 card), it is possible to supply the user with relatively up-to-date inventories at reasonably frequent intervals. Apart from the initial cost of purchasing the required microfilm readers, this system eliminates the high level of expense entailed by maintaining a direct link to the computer. In this respect, the COM procedure is the ideal instrument for small translation services "latching on" to larger systems. The translators in the Bundessprachenamt themselves use this procedure for shorter translation jobs. However, for us, having direct access to the computer, the COM procedure has some disadvantages, with the result that we normally use it only in the search for isolated technical terms.

It is much easier, in the case of large translation jobs, to prepare a text-oriented request list of the terms sought than to resort to the use of the microfilm reader which, though technically simpler, is inordinately time-consuming when the number of terms sought is high (and these are scattered throughout the alphabet).

The most serious disadvantage of the COM procedure is, however, that no hard copy is produced, with the result that communication amongst the translators working on the same translation job and between translator and revisor is severely impeded. It becomes almost impossible to co-ordinate the terminology used in major translation projects.

A further, though not necessarily crucial drawback of the microfilm system is that the translator receives an inventory which is never quite as up-to-date as the latest status of the master terminology bank in the computer.[2]

---

2) A definitive appraisal of the COM procedure is not yet feasible, as the procedure has only been in operation for a few months. Changes to the manuscript may become necessary between the submission date (15.11.76) and the actual Conference in May.

2.4. <u>Operation in conversational mode</u>

Operation in conversational mode via a display screen is another relatively new method of obtaining access to the computer terminology bank. Experience has already shown that the translator himself will hardly make use of this facility. Because of the expense involved, it will not be possible to equip every translator with a display screen. Consequently, the translator will as a rule have no practice in employing the procedure and will require the assistance of an intermediary.

Nevertheless, conversational operation is the ideal aid for terminologists and lexicographers as it enables them to acquire insight into the dynamic computer inventory and thus to manipulate this data base selectively.

In any event, it should be borne in mind that, as things stand at present, a list in hard copy form is still the best output medium for the translator. Output via a display in conversational mode, on the other hand, is less suited to the translator's purposes.

3. <u>CHOICE OF VOCABULARY</u>

The vocabulary selected for input into the terminology data bank reflects the differing requirements of the various users. These requirements may vary considerably, depending on whether the data bank is to be used, for example, primarily as a tool for a translation service, in dictionary or glossary production, or for other scientific tasks (e.g. thesaurus compilation). A different procedural approach will therefore be required for each of these objectives.

Very broadly speaking, I should like to assume in the following that there are two main approaches, of which the most varied permutations may exist.

When compiling dictionaries and special subject glossaries,
a systematic approach is called for in collecting vocabulary.
For the object of a dictionary or special subject glossary
is to present the vocabulary included as an extract from
the total word stock of the language in question. Here it
is most important that the basic vocabulary and the general
concepts of that language should be collated as fully as
possible as the basis on which to build up an ever more
specialized vocabulary of ever more limited and infrequent
use. The borderline as to whether or not any term should
be included in the dictionary is drawn at the point where
that term is used so infrequently as to cease to appear
economically worthwhile to the compiler of the dictionary.

The other extreme is the pragmatic, selective approach which
is called for when a terminology data bank is to be created
primarily for the purposes of a technically oriented trans-
lation service. A technical translator with well-founded
knowledge and many years of experience in his field will
not normally have to ask the computer to supply him with
the basic terminology of that field. A translator spe-
cializing in aeronautical engineering, for example, will
not have to look for the translation of the word "under-
carriage".

Whereas, when the dictionary concept is systematically
applied, work starts at the bottom level and proceeds
upwards, being cut off at the point where the low frequency
of a term makes its inclusion uneconomical, the technical
translator's interest in the data bank takes exactly the
opposite course: the more general and frequently used a
word is, the less it is likely to be of interest to him,
as he is familiar with it anyway. Consequently, when the
latter approach is used, the terminology store will generally
be filled up, though not unsystematically, more from an

ad hoc point of view dictated by the users' real needs.
Neologisms will be more interesting than terms which are
"old hat". A great many (about 4/5) of the terms in the
terminology bank will be composite, which again means
that the computer will have to accept entries of various
lengths.

These contrasting aspects also entail contrasting personnel
requirements for organizations running terminology data
banks. Whereas it may be quite adequate to employ qualified,
experienced lexicographers in the collation, evaluation,
processing and storage of vocabulary when systematically
compiling a <u>general</u> dictionary and providing one does not
delve too deeply into specialized terminologies - this
vocabulary is adequately defined in conventional mono-
lingual dictionaries (e.g. Webster), so there is no need
for subject specialists - real technical terminology can
only be extracted and processed by a team made up of the
expert in the field in question and the terminologist/
lexicographer. [3]

The functions of the technical expert and the lexicographer
in such a team are predetermined from the start. The lexico-
grapher (or terminologist) is responsible for ensuring that
the terminology is presented in lexicographically pure form,
that the expert is not confronted with linguistic in-
accuracies in his terminological work, that linguistic un-
ambiguity is maintained at all times by comparing the
vocabularies of various technical fields, and that homo-
geneity is maintained within the technical field in question.

---

3) Cf. Wersig: Probleme und Verfahren der Terminologiearbeit, in:
   Fachsprachen, Terminologie - Struktur - Normung, published by
   Beuth-Verlag, 1976, pp. 43 et seq.

Responsibility for the technical accuracy of the terminology,
on the other hand, must be vested clearly and exclusively
in the technical expert. It is for this reason that inde-
pendent compilers of technical dictionaries are nearly
always first and foremost experts in the technical field
concerned and lexicographers only as a secondary function.
It is inconceivable for a terminologist to attempt, perhaps
for reasons of linguistic puritanism, to bring about a
linguistic reformation by ignoring the usage of a whole
technical world and creating new terms not accepted by
that technical world, merely because he considers his own
terminology linguistically more correct, logical and well-
founded. No technical jargon will allow itself to become
the victim of pedantry.

Unlike common language, technical terminology often goes
beyond the terminologist's/lexicographer's ability to check
it. While the lexicographer dealing with common vocabulary
will have a wide range of reference material at his disposal
and this reference material takes a long time to become
obsolete, such material is not available to the same extent
for checking up on technical terminology, unless this
terminology is of a very basic nature. It is this lack of
access to reference material, particularly in the case of
the more recently established technologies, which must be
compensated for by the expert knowledge of the specialist
in the field. It is also the expert's task to ensure that,
when the system comes into everyday use, the data base is
kept constantly up to date. Terminology fed into the computer
some years ago may well have become obsolete; the technical
expert coming across such deficiencies, e.g. in a text-
oriented glossary, will rectify the situation and input
new terminology and/or change obsolete terms.

4.   <u>DATA STRUCTURE; DEFINITIONS</u>

An essential question of interest to the user is that of the contents and scope of the data record. That this must contain source language and target language equivalents, source symbols, language symbols etc. goes without saying. But opinions are divided as soon as it comes to sorting terminology into subject fields. I personally recommend caution in this respect. To use too fine a mesh of subject codes in splitting up terminology for use in machine-aided translation leads to a high redundancy level and increases the danger of coding errors.

A further point of dispute was and still is whether or not to include definitions and/or sample contexts in the data entry. Of course, terminology work without some clarification of the concepts behind the terms involved is inconceivable. And a knowledge of the conceptual contents of any term is indispensable to the person prescribing that term. The question is whether it is essential to supply the translator with definitions for all the terms on his initial request list, which often contains hundreds of entries.

If, when approaching the system with his search questions, the translator is searching for hundreds of single-word or composite terms in one batch, he should not at first be supplied with the corresponding definitions and sample contexts, especially not in the "hard" sciences. This generalization requires immediate qualification:

- Of course, definitions and sample contexts should be
  kept available in background storage for recall at any
  time. In our case, the source code for each term formerly
  referred us to reference documentation in which the de-
  finition of that term could be found if required. Nowadays
  computer technology is so advanced as to make it possible

to keep explanatory comments, too, in a computer memory,
to which direct access can be gained at short notice.

- Naturally, this does not exclude the possibility of in-
cluding a very concise explanation or qualification,
generally in the form of a qualifying expression in
parenthesis following the source-language or target-
language term.

- The "soft" sciences (social, legal and administrative
sciences) abound in terms likely to require definition.

If I emphasise that qualifying expressions are present in
a background memory, this means that they are not supplied
to the translator with his printout but only later, on
specific request. This part of our procedure is at present
still in the planning phase. The translator, however, is
more interested in receiving the target-language equivalent
of his source-language term and, in the case of the majority
of terms from the "hard" sciences, he is able to select this
equivalent without the aid of a definition. Considering the
various types of output previously described, to output
qualifying expressions to every single search term would
normally result in huge amounts of noise. In the case of
the COM and hard copy output procedures, this would inflate
the number of microfilm cards and printed pages beyond belief;
in conversational operation it would prolong inordinately
the length of time required for the search.

5. <u>REQUESTING TECHNIQUES; RESTRICTION TO CERTAIN SUBJECT FIELDS</u>

A further common difference of opinion between terminologists
and translators arises from the question whether, in the
search for target-language equivalents, the search request
should be restricted to a limited number of subject fields,
so as not to be swamped by a flood of target-language
equivalents. From the user's viewpoint, however, this

redundant information is not unwelcome, even desirable.
The classic example of an argument against restriction is
the request for a translation of the word "control", a
term for which the computer will output a wide range of
equivalents. But that is just what the translator needs
in this case. Any translator will have ten translations for
"control" in the back of his mind, but none of these may
satisfy him; he is looking for the eleventh or the twelfth
translation, the one that is "on the tip of his tongue"
but refuses to come out. In this case he uses the computer
as he would a normal dictionary.

In the majority of cases, however, where an equivalent is
sought to a true, composite technical term, the computer
will offer only a few equivalents anyway. If the translator
were to restrict his request to a few fields he would exclude
a priori potential sources of helpful information or would
neglect to make allowance for sources of error. To give you
a few examples:

- The lexicographers engaged in coding the terminology
  according to its subject field are only human and as
  such cannot be expected to know or suspect all the
  possible subject fields in which a certain technical
  term is likely to occur; all categorizing systems have
  their weak points. If a term is allocated by coding to
  only some of the subject fields in which it actually
  occurs, the translator will receive the answer "MISSING"
  if he restricts his search to one or more of the other
  fields.

- The translator may well be interested in "taking a look
  over the garden fence" to see how a technical term is
  translated in a related field.

- We have a special problem in the form of the various
  problem-oriented special data bases, some of which have

not been subject-field coded at all. Checks performed on text-oriented glossaries on the basis of French source-language texts revealed, for example, that over 50 % of all positive and correct answers given by the computer on the text-oriented glossaries were coded as problem-oriented terminology and not to any subject field, even when the texts analysed had nothing to do with the projects form which the terminology was taken.

As a well-known documentation specialist [4] pointed out recently at a congress, "pure-born" technical terms are, in any case, few and far between. Considering the degree to which the various scientific disciplines are inter-related, the translator must expect to find a considerable amount of "alien" terminology in any technical text, irrespective of the actual specialized field primarily concerned.

There is still much more to be said on the wishes of the translator as the user of the computer system, for example as regards the planned expansion of the system to include cross references and references to synonyms. Unfortunately, the short time available here today does not permit me to go into visions of the future.

---

4) Zimmermann, D.: Sprachwissenschaft und Terminologiearbeit,
          a paper read on 18.10.76 in Düsseldorf before
          the Arbeitsgemeinschaft für Rationalisierung.