TYPES OF TRANSLATION AND TEXT FORMS IN THE
ENVIRONMENT OF MACHINE TRANSLATION (MT)

J.C. Sager

Centre for Computational Linguistics
University of Manchester Institute
of Science and Technology

Human translation consists of a number of separate steps
which begin with the identification of the text type,
the purpose and intention of the text, the subject area,
etc. As there are types of texts there are also types
of translation, which do not necessarily match directly.
Since the human and machine translation processes differ
so must the criteria which determine translatability.
What criteria are relevant for MT and can they be derived
from observations of the human effort?

INTRODUCTION

This paper is basically concerned with efficiency of translation
seen as one aspect of the effectiveness of interlingual communication.
To this end we are mainly concerned with the product of translation
and the adequacy of this product for its intended purpose. This
product is, however, diverse, the diversity being the result of
varying emphasis on the separate steps of the total complex process.
In recent years this process has become variously assisted or
amenable to assistance by information technology. A previous Aslib
conference was concerned with the role of text processing in
translation, this conference concentrates on the assistance offered
by MT systems. In order to assess the right point of intervention,
the degree of machine assistance or even the possibility,
acceptability and cost-effectiveness of the replacement of the
human product by machine output in particular cases, we have to
analyze the various steps of the human process. We also have to
relate the adequacy of the product to the cost of the mediation
which should correspond both to the use made of and the significance
attributed to this communication by the originator and the recipient
of the message thus mediated.

We therefore start from two ends: our experience, however limited,
of MT, and an analysis of human translation in its dual sense of
process and product. We accept that various forms and systems of
machine translation exist, and are fully used for limited objectives.
We postulate that under certain conditions MT offers a viable
alternative, partial alternative or supplement to human translation.
Some systems are constructed with a controlled language and thus
satisfy the limited expectations for which they have been planned.
Experience with other forms of MT has shown that some systems
produce different types of output (Qualities?) according to certain
properties of the input texts. We can assume that an identification
of these properties may produce criteria for the selection of texts
suitable for optimal processing. The question is whether these

positive properties can be associated with readily identifiable
text categories or even whether a 'grammar' of such properties can
be constituted which may lead either to acceptable writing rules
for existing text categories or alternatively to the selection of
texts which conform to this 'grammar'.

Since MT is seen here as one of the possibilities of interlingual
mediation - used in parallel with human translation - we must
also identify the diverse types of translation produced by man and
machine and attempt to correlate them to types of texts suitable
for one or another form of mediation. It is, of course, possible
that different types of translation may be applicable to the same
types of texts and vice versa. For machine translation this
approach is clearly tentative and speculative, but it should be
possible to indicate areas of expectation and construct
appropriate experiments to observe and direct research in MT in this
light. The immediate objective is to provide orientation for MT
work which is based on optimizing efforts and results so as to avoid
the disappointments of the past which resulted from too broad a
basis of research and too indiscriminate claims for the applicability
of MT. The still widespread scepticism about the practical value
of MT reinforces this pragmatic and problem-oriented approach.

In the same way as we distinguish different types of human
translation, we can distinguish various forms of MT, defined here
as follows:
Full MT: machine output without human intervention which goes
directly to the end user;

Controlled language MT: the input text written according to
certain rules which produces output of similar characteristics;
the output used directly by the end user;

Partial MT or Machine-assisted MT: human editing applied before,
during or after machine processing; the subtypes Pre-edited MT,
Interactive MT and Post-edited MT may occur in mixed form.

These types have yet to be subdivided by function.

EVALUATION AND COMPARISON

Most comparisons between human and machine translation are
misleading, others are simply unhelpful. If we consider them as
two different methods of achieving similar objectives we may keep
human translation separate from machine translation, each requiring
its different types of mediator. The source text for translation
may be the same, since all texts are in principle susceptible to
human translation, but we are now trying to identify text categories
particularly suitable for machine translation processing. Comparison
of the target language product is gratuitous and has led to much
dispute. MT currently aims at two products: output suitable for
human post-editing, and immediately usable output which is clearly
recognized as different in type from human translation. Nevertheless,
by whichever methods one chooses to describe this information
transfer between natural languages, the final objective of the two
processes is the same, and comparisons will and must be made by the
end user on the adequacy of the product for its purpose and on
the cost-effectiveness of the operation.

A great deal of time has been spent on the comparative evaluation of the quality of MT. Much of this effort was misdirected because it was based on fallacious principles. There is no single standard for human translation with which MT output can be compared, and "standard" has been expressed as an abstract notion of quality which in practice does not exist. Quality of translation is as elusive a notion as quality of the source text and has to be replaced by the criterion of adequacy for the intended purpose which is applicable to various types of translation, and comparisons are only valid if the specific purposes and cost and time factors are comparable.

Adequacy comprises basic requirements such as fidelity or accuracy of transfer of information and intention, stylistic appropriateness and intelligibility, most of which cannot be globally assessed in a single operation as this would require an expert in the subject matter who knows the languages involved in the transfer, i.e. a human evaluator who does not need the translation and who would therefore not be a genuine user of the product.

Equally important factors are reading time by the end user where types of reading have to be considered, e.g. texts written for quick reading, cursory scanning, single reading, repeated reading, as well as the intended range of readers and the size of reader-ship.

Economic factors of production cost are difficult to relate: there is insufficient information available on the true cost of MT system development, there are no data on running cost and maintenance, and, of course, there is little information on volume of translation carried out. Cost of production criteria alone are likely to favour full or controlled language MT, especially if the reading time for the end user is either not significantly longer than that for human translation or irrelevant to consideration. But we do not know whether reading skills of full or controlled language MT can be developed to allow quicker reading, and the time factor of speed of production of MT, i.e. the time a reader has to wait for a translation to be produced, may also be significant. There is in addition the factor of multilingual MT output which we know from TITUS heavily influences cost calculations.

In partial MT comparisons are only valid if edited MT is to serve as a full substitute to human translation and must take into account the entire process of MT, including maintenance of dictionaries. Various standards of MT editing may also produce new subtypes of translation which will further complicate comparative evaluations.

Methods of assessment are therefore as difficult to establish as criteria of evaluation. The most complete survey of existing thought on this subject has been produced by Georges van Slype as a result of a workshop organized by the Commission of the EC as part of its First Multilingual Action Plan. The Translation Bureau of the Federal Government of Canada in Ottawa is sponsoring a study of methods of evaluation of translations, a preliminary report of which appeared in the summer issue of META.

In the context of this conference the factor of acceptability of MT output to post-editors is important, but we must consider the effect on productivity of post-editors in the long term.

Experience in this field is limited at present and fully trained
post-editors may add to our experience during this conference.
The experience with translators and revisors in the Commission
of the EC was on the whole negative, but the type of source text
and consequently the types of translation used as samples may have
been inadequate.

In general, editing of MT involves a variety of different skills
and functions which are not comparable to revision of human
translation. It can be performed, before, during or after
machine processing of texts. It is generally assumed that editing
is suitably performed by translators specially trained for the
job. This claim has yet to be investigated and training patterns
will have to be established. Systems designed for human intervention
in the actual process or as part of the overall process of trans-
lation must take into account existing skills and training facilities
or alternatively specify realistic job descriptions for these tasks.

There is no doubt that translators have a role to play in the
performance of machine-assisted translation, in the post-editing
of MT output from systems designed for this type of human inter-
vention and possibly also in the selection of source texts for
one or another form of machine processing. Whether translators
can make a direct contribution to the actual development of MT
systems is yet to be established.

Generally therefore, comparisons between MT and human translation
are inevitable but should be made with great caution and only in
areas where comparison is useful and productive. Evaluation of
MT can only partly be made in relation to human translation as
other factors also have to be considered.

TYPES OF TRANSLATION

Translation is a process of mediation applied to existing texts
which are variously modified to generate derived texts. This
concept of 'derived text' taken over from information science
is useful here because it implies the continued existence of the
original text, the possibility of envisaging different degrees
of modification and, hence, a different function for the new
product or products. By extending the concept we can also redefine
the role of the translator as that of a language and communication
mediator comparable to technical writers, abstractors, precis
writers, etc - in practice many translators perform such a double
role. We can also widen the concept of translation to comprise
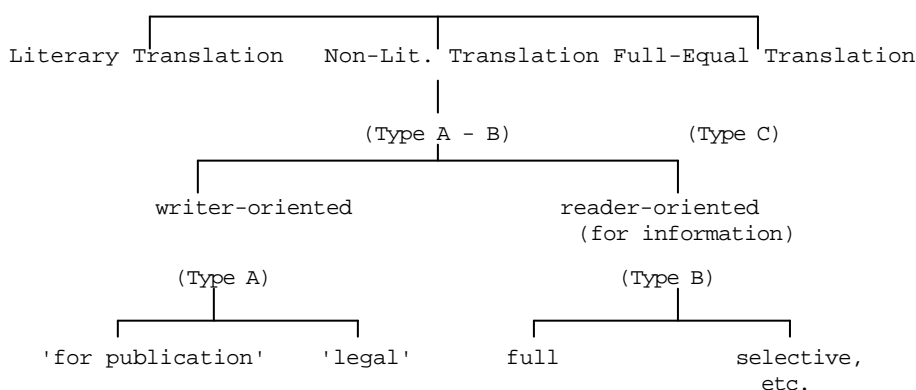three categories of modifications to texts:
a) those which are inevitable for linguistic and pragmatic reasons -
   the traditional notion of translation and the main concern
   of translation theory;

b) those which are deliberate and planned according to the use
   intended for the derived text - the practical types of
   translation such as full, selective, abstract, gist; and, of
   course,

c) those which are accidental; i.e. the errors produced through
   negligence or misunderstanding, to which we have to add those
   unforeseen by MT designers.

The type of mediation to be applied is determined by the status
of the derived text in relation to the original and this status is
generally fixed beforehand either by the writer of the original
text or by the reader of the derived text, who in certain cases
may act in agreement or as a single user. According to its status
the derived text is considered as:

Type A:   - a full substitute, as if it were a new original with
            an independent status, possibly to be subjected to further
            derivation. This is the type usually specified by the
            writer, and produced by the translator in the absence
            of detailed specifications. This type is normally
            considered as the only genuine translation and all
            translations are usually evaluated by the criteria
            applicable to this type.

Type B:   - an alternative to the original with the same or different
            functions. This category is often specified in some
            detail by the informed reader and may require two forms
            of derivation simultaneously, i.e. linguistic transfer
            and adaptation, reduction, or selection of some sort.
            In this case the translator simultaneously works as an
            information scientist. This type of translation is
            probably the most widely used mode and offers scope
            for many subtypes some of which may be suitable for
            machine or machine assisted translation.

Type C:   - a full equal, coexisting with the original, so that after
            the process of mediation the distinction between original
            and derived text disappears. This category is frequent
            in international organizations and some legal documents
            but is otherwise not heavily in demand.

     Any derived text is commissioned and the translator receives
a request for mediation with a more or less detailed specification
as to its purpose and can adjust his work accordingly. Such requests
come either from writers of texts who want to send messages in
another language or from readers who have received a message they
cannot understand. We have characterized type A as writer-oriented
and type B as reader-oriented; type C is neutral in this respect.
These two positions are also represented, however roughly, in the
conventional categorizations of translations as "for publication"
and "for information", in that the former is specified for wider,
often unknown, use and the latter for a limited number of, often
known, readers and for specific uses which may be further qualified
as annotated, reduced, selective, etc. This distinction is well
exemplified by letters. Outgoing correspondence is writer-directed
and requires a fully independent foreign language text; incoming
correspondence is very often translated in gist form only, because
the reader expects the translator to select the relevant information
for him.

If we relate this general categorization of translations to the
various types established by the profession, we find confirmation
of these major types and further subdivisions within them, which
cannot be fully discussed here. We also find some correlation
between types of translation and broad categories of texts and
subjects and can thus tentatively construct a mixed hierarchy
of translation types, as follows:

```
        ┌─────────────────┬────────────────────┐
Literary Translation    Non-Lit. Translation  Full-Equal Translation
                           │
                        (Type A - B)              (Type C)
                ┌──────────┴────────────┐
          writer-oriented          reader-oriented
                                    (for information)

             (Type A)                  (Type B)
        ┌───────┴───────┐          ┌───────┴────────┐
   'for publication'  'legal'     full           selective,
                                                    etc.
```

We have made no attempt to characterize literary translation which is
a well-established category of no particular interest to us here.
This grouping/categorization is preliminary and has to be developed
further to include types of MT output with different functions from
those listed here.

The process of human translation has been divided by Bathgate (1980)
into the following stages:

   (SL text) tuning - analysis - understanding - terminology -
         - restructuring - checking - discussion (TL text)

This 'operational' model claims to represent the different activities
undertaken by the human translator. They overlap and may alter in
their order. 'Tuning' is explained as the preparation any writer
needs to adjust to the task in hand and is thus part of the reading
process for the production of derived texts. 'Discussion' with a
subject specialist is a desirable part of the revision process.

Simpkin (1980) proposes a simpler model which, however, largely
corresponds to the operational model and will serve our purposes.

Reading - Research - Dictionary Look-up - Translation - Revision
                                            dictation   Editing
                                                        Proofreading

It is obvious that the time required for each stage varies with the
type of translation and the degree of specialisation of the human
translator. We note that translation Type A is heavy on research,
whereas a full translation of Type B usually requires less overall
time, especially research, revision and editing.

According to the amount of re-writing conditioned by cultural
requirements, Type A may lend itself to machine-assisted translation.
The full translation of Type B may, in some cases, be replaceable
by full MT if readers find factors such as immediate availability
of the translation and lower cost outweighing the advantages of the
time saved in reading a selective translation. Selective translation
requires highly skilled mediators who may be too expensive and not

readily available. Type C is often carried out by teams of trans-
lators and legal experts but they, too, may find machine-assisted
translation helpful as it may suggest consistency of terminology
and phrasing analogous with previously translated documents.

From this analysis we also derive a number of other observations.
Some translations are not translations in the conventional sense,
but are concerned with the production of derived texts, i.e.
versions into which a number of non-linguistically motivated
changes are introduced by omission, addition, substitution or
transposition which arise from particular conventions and special
user requirements.  This re-writing type of translation has to
be identified as it does not readily lend itself to MT treatment.
Alternatively MT may produce new low-cost types of translation
which may satisfy a demand for information hitherto neither
identified nor met.

It would be useful to quantify the types of translation for any
large-scale user group in order to be able to estimate whether
it is economically viable to develop purpose-specific MT systems
for them.

TEXT FORMS

Translation types are applied to identifiable text categories. When
we write a text or envisage the translation of a text we first
identify the text form and adjust our subsequent behaviour
accordingly, the tuning process defined by Bathgate above. We read
a newspaper differently than a specific report and we also
identify a text for translation as requiring this or that form
just as a writer knows that he has to write, e.g. instructions
for use and that these have a different form than e.g. a sales
brochure for the same product. The text form signals the intention
of the author. As soon as we have identified a text we have certain
expectations of tone, mode of expression, deliberate redundancy,
presentation of facts and arguments, even of layout into paragraphs
and columns. The presence of these features assists our under-
standing, their absence may lead to a misinterpretation of the
intention of the message or even the content.

The characteristics of particular text forms complement subject
or vocabulary specifications so that we know that a certain subject
field is usually represented by a limited number of text forms.
We know that different linguistic communities, even the closely
related European ones, have different categories of text forms
and varying conventions even for those we consider parallel, such
as business letters, scientific articles or contracts.

As readers we are thus more or less familiar with a number of text
forms and can therefore understand them; as writers or translators
we are better acquainted with a smaller number of forms and know
the norms which govern the permitted variations. The injunction
or practice of translators to translate only into their mother
tongue is principally determined by the greater familiarity
with text forms and their permitted variations that they have
in their mother tongue.

While there are monolingual style sheets and writing manuals for a
number of text forms, the field of text grammars has only recently
been opened up by linguistics. On the interlingual level there are
a few contrastive grammars but they are largely concerned with

phrases and sentences. No contrastive grammar has yet been written
for text forms, and skills in translating or writing in a foreign
language are therefore largely acquired by observation and
experience.

While in theory several types of translation may be applicable to
a given text form, in practice there are established patterns of
translation which are often institutionalized and divided into
outgoing and incoming material. Successful translation departments
and freelance translators have analyzed their customers' needs
and offer the services considered most appropriate to each
circumstance. They know from experience which subject field and
text forms require regularly a certain type of translation or
mediation.

There are a number of questions which require further study.
How, or by what features do we recognise a text form and can such
a process be automated?

What are the adjustments we make before reading and translating?
We understand a selective or gist translation only if we know
what text form it comes from. The knowledge that we are dealing
with a letter, a contract, a leader or an article is essential
for interpreting the content of a selective translation and for
attributing the correct intention to the message. Equally, it *is*
clear that we must know how a letter is written in a target
language before we can produce it. The source language original
does not provide us with sufficient guidance; in fact, a glance
at previous correspondence and a list of the points to be made
in a letter may be enough to produce a satisfactory translation.

What is the correlation between subject field, text forms and
translation types? Are there other areas with sufficient volume
of demand beside the one identified by TITUS for such a subject,
text form and translation type specific system? Our increased
capacity for machine analysis of texts - a by-product of MT
research - may indeed assist in the collection of the necessary
data for the production of the monolingual and contrastive text
grammars we need for improvements and further sophistication
of MT.

CONCLUSION

In conclusion, and at the beginning of this conference which
will suggest many new ideas, I should like to draw attention
to the implications of all those observations for MT evaluation
and future MT development.

Different forms of mediation lead to new types of products (of
translation). These products must be assessed for their adequacy,
as the end user is rarely in a position to do so.

As there are different types - not qualities - of human translation,
there are different types of MT, each designed for specific
purposes. The present diversity of approaches to MT development
should therefore be reflected in types of output for established
or recognized uses.

For the practical purposes of assessing adequacy of translation, the division of texts by subject fields is of limited use and has to be supplemented by divisions into text forms and types of translation suitable for optimal information transfer.

Types of translation relate to forms of texts with their own grammars which are preferential subsets of general usage with some subject—specific forms established by norms of usage.

The lexical and syntactic choices made by the skilled human translator on the basis of his recognition of the subject, the text form and the translation type must be replicated in MT development either by allowing for human intervention, subdivision of dictionaries and grammars or by separate systems. Evaluation of human translation and MT will be facilitated by the development and definition of types of translation appropriate to each method.

Analysis and formalization of text-form-related types of translation will lead to categories of derived texts suitable for interlingual communication. The end user's knowledge of these categories will permit him to specify more clearly the type of translation suited for his needs and thus increase his satisfaction with the product.

The diversity of products of translation will increase with development of MT systems. The user must be given a choice of viable alternatives which he can only exercise if he has full knowledge of all the translation types available together with the various time and cost factors.

BIBLIOGRAPHY

(1) Aslib, Proceedings, Machine Aids for Translators, Vol. 33, (1981) 7.

(2) Bacon, J., Drazil, M., Report on the Seminar on Translation Specification, Newsletter of the Translators' Guild, 6 (1981) 1, pp. 7-14.

(3) Bathgate, R.H., Studies of Translation Models, in: The Incorporated Linguist. 19 (1980) 4, pp. 11-12 + 20 (1981), pp. 11-16.

(4) Barbizet, J., Pergnier, M., Seleskovitch, D., Actes du Colloque International et Multidisciplinaire sur la Compréhension du Langage, Didier, Paris, 1981.

(5) Gouadec, D., Paramètres de l'évaluation des traductions, in: META, 26 (1981) 2, pp.99-117.

(6) Simpkin, R., What the Customer Wants, in Newsletter of the Translators' Guild, 5 (1980), 4, pp. 12-17.

(7) Van Slype, G. Critical Study of Methods for Evaluating the Quality of Machine Translation, Final Report, Brussels/Luxembourg, CEC, 1979.