[From: *Overcoming the language barrier*, 3-6 May 1977, vol.1
(München: Verlag Dokumentation, 1977)]

OBJECTIVE EVALUATION CRITERIA FOR MACHINE TRANSLATION

A. J. Petit

Machine translation

The Canadian Government Translation Office

Abstract

If a translation of a complete text is considered a finished product, quality control principles can be applied. Defects likely to occur in a text need to be defined and the permissible tolerance limits established for each of these defects so that the text can be given a real acceptance test followed by approval or rejection of the product. The finished product is assessed irrespective of the production method used: manual or machine translation without revision or manual or machine translation with revision. The defects are defined and discussed. The tolerance limits are established and it is shown how this method enables an automatic translation system to be evaluated on the basis *of* design characteristics.

The technical translation as a tool

As part of its bilingual programme the Canadian Government classifies as tools all its technical manuals used for the operation, maintenance and repair of State equipment or equipment belonging to State-run companies. Once translated the technical text thus becomes a daily tool for all those who keep an industrialized country running: - operators, mechanics and workers. This tool must comply with very strict requirements if it is to remain of use.

Traditional methods of evaluating which consist in giving a translation an overall mark on the basis of the success or failure of the translator to produce a text are useful if an examinee has to be assessed or if levels of competence are to be established for administrative purposes. However, they are not suitable for checking the quality of a technical translation where a single error could have serious consequences and jeopardize the safety of persons and equipment.

The technical text (translation) considered as a tool is an industrial product which must be assessed on the basis of standard quality control methods. Once quality control has been carried out, one choice only is possible: the text is accepted or rejected. As with any industrial product, quality control can be carried out at various stages in production. For each production stage the acceptance criteria are determined which guarantee that the quality of the product is not irrevocably sub-standard and that it can be sent on to the next processing stage. It can be reasonably accepted that if this stage is passed, the product stands every chance of meeting the quality standard.

Checking the translation and checking the revision

Now and again a translator produces the finished product by himself (unrevised translation) but as a general rule the translation process is in two stages: translation and revision. An interim check allowing for rejection after the

translation stage enables any text to be discarded which would require an excessive amount of revision or which showed production faults which (in view of the means at the disposal of the reviser), make it difficult if not impossible to rectify by revision.

Quality control must therefore be based on a double set of criteria: one set concerning acceptance of the final text (i.e. revised) and another concerning the criteria for the acceptance of the text 'ready for revision'. Quality control during production offers a particular advantage where translation is computerized since it enables the quality requirements which can be imposed upon an automatic system to be objectively established, so that following normal revision, the text remains within the tolerance limits, 'normal revision' being taken to mean the rectification of occasional minor defects or of a limited number of major defects occurring regularly.

To reflect the real situation with any accuracy the quality control must take into account a complete unit of translation, or rather, of documentation. In practice a given piece of equipment is accompanied by a maintenance and instruction manual, for example, or it is described in a section or chapter of a larger manual. A unit of documentation corresponds by definition to a complete manual, a part of a manual or a particular chapter according to the complexity of the documentation as a whole. By way of illustration, typewriter maintenance instructions, a maintenance manual for a winch or the chapter on 'the undercarriage' of an aircraft maintenance manual represent units of documentation. The volume of the unit of documentation can vary from a few pages to 20 000 words but it is a single entity which represents a tool for the specialist or group of specialists: it is a working tool in its own right.

Inadequacy of conventional scales

Here, by way of illustration, is the conventional correction scale used at the Translation office:

1) Translator TR 1

Errors	Points
Grammar	max 3
Syntax	" 3
Usage	" 3
Misinterpretation	" 5
Gobbledegook	" 5
Omission of a passage	" 3 points per phrase constituent
General quality (subtract or add)	" 5
Spelling	" 2

2) New reviser TR 3

Errors	Points
Grammar	max 3
Syntax	" 3
Spelling and punctuation	" 3
Usage	" 7
Misinterpretation	" 7
Gobbledegook	" 7
Style	" 15 (to be added or subtracted as the case may be)

It should be added that this assessment is based on a fixed length of text.

Let us suppose that a piece of text of 300 words contains three cases of misinterpretation, one spelling error, but the style is quite acceptable. Translator TR 1 will lose 17 points (see scale) and the new reviser 24 points, but the latter may obtain a 'bonus' of 7 points say, for style. In both cases the final mark is more than 80/100, a result which is extremely satisfactory one would think.

On the basis of these data it is possible to extrapolate that 60 cases of misinterpretation occur in a unit of documentation 6 000 words long. The worker who follows the manual for the completion of a given task will on 60 occasions be given wrong instructions to follow. Let us return to industrial products: the number and gravity of defects tolerated are not proportional to the complexity or size of the equipment. On the contrary, the quality control on a large aircraft is more exacting and painstaking than that on a piece of office furniture.

Selection of objective criteria

In a technical text errors can have the following consequences: errors if considerable in number and of a primitive nature would make the specialist quickly realize that the text was useless and decide to consult the original or, if no original was available which he could understand, he would be led to continue his work without consulting it. On the other hand the error could be so inconspicuous as to cause the specialist to carry out a disastrous operation. Somewhere between these extremes are the variety of cases of incoherence which require a disproportionate effort to understand. Obviously the standards must be extremely rigorous. In any manufacturing process the tolerance aimed at from the start is zero but reality must not be lost sight of and it would seem reasonable to adopt real error rates as workable tolerances, such as can be found in good original-language technical documentation.

TRANSLATION TOLERANCES

The figures in the column headed 'Tolerance' (number of admissible defects) are based on a complete unit of documentation which may contain up to 20 000 words.

MAJOR DEFECTS

Defect	Consequence	Tolerance
Terminological error	The specialist must make an effort to find the item or operation referred to(wasted time, risk of confusion)	Not quite the right term: LESS THAN 10 wrong term: MAXIMUM OF 2 incoherence: LESS THAN 5
Nonsense and gobbledegook	The specialist does not understand and must: - carry out the work without the manual - have the offending passage retranslated(time wasted, error risk, expensive telephone or telex calls)	NONE
Misinterpretation or wrong meaning	Can cause the specialist to perform an operation leading to the total or partial destruction of equipment or risk to the lives of other members of the team or operators.	NONE
Error of syntax or punctuation	Can cause misinterpretation (see misinterpretation) If misinterpretation not possible (see minor defects)	
Illogicality (inconsistence)	Requires effort to be understood (time wasted,loss of credibility)	LESS THAN 10

Specialist = person for whom the manual is intended (worker? technician)

MINOR DEFECTS

Defect	Consequence	Tolerance
Wrong usage, anglicism, barbarism, spelling, wooden style, etc.	Can antagonize the reader. Frequent occurrence of these errors could provoke a deterioration in relations at work since the specialist could demand that 'his language be respected'.	LESS THAN 5 PER 1 000 WORDS

The tolerances given above apply to the revised text, in other words the finished product.

Normal conditions applying to quality guarantees

It is generally agreed that the level of revision necessary to obtain an acceptable finished product varies from one translator to the next. Revision may be quite extensive for a beginner and minimal, i.e. superfluous for an established translator. For any team of staff translators the average should be set with a view to striking a balance. A balance can be said to exist when a reviser can check and correct the work of three translators, a situation which can apply only in the following cases: 1) a good translator + and average one + a beginner; 2) three average translators, etc; once two beginners are in any trio the balance is upset and a drop in quality or production occurs, or both. Computerized translation is only economical when the output quality is at least the same as that produced by an average translator, which in practice means a translator with a number of years' experience in that field. In this instance the quality of the product remains unchanged and the work of four translators (three translators + one reviser) can be done by one (machine?) This level can thus be taken as threshold value of economy of operation with quality guarantee.

Unfortunately, the super translator remains indispensable, which means that the problem of the availability of human translators is still with us. In fact, the shortage of qualified technical translators is often the main obstacle to the expansion of any translation service and the factor which prompts the adoption of an automatic translation system. What use would an automatic translation system be which produced large quantities of translations only to be congested by the lack of revisers? If the quality of the finished product is not to be put at risk the standards applying to automatic translation systems must be raised so that a translation can safely be revised by a translator who does not have the experience and competence of a reviser as such, and this involves the elimination of any error where correction of that error requires specialist knowledge in that field.

Now on the basis of the requirements of the finished product, the intervention limits of the 'reviser/editor' become obvious, in other words we are talking about the degree of efficiency required of an automatic translation system.

Eliminating test

One of the first stages of translation consists in determining the exact function of each word in the sentence to be translated. Translators are required to sit an entrance examination on the basis of which candidates can be eliminated who do not have sufficient knowledge of the source language to recognize its structure (irrespective of meaning). In the human translation the problems caused by homographs become minimal, a few rare examples, often taken out of context, are the ones which linguists take a delight in quoting.

Any unresolved homograph obliges the reviser to retranslate the entire sentence. Just as a candidate who is unable to recognize the structure of the source language is eliminated at the admission test stage, any machine translation system

which does not have a syntactical analyser technically perfected enough to resolve problems of homographs is UNACCEPTABLE. Only such cases of natural ambiguity are tolerated as cannot be solved by the translator without requesting explanation from the author of the original text. This basic requirement is all the more reasonable as the current development of computerized linguistic research has enabled the problem of homographs to be completely eliminated, a problem which about a decade ago seemed insurmountable for the Georgetown generation systems.

Once a decision had been taken to get rid of systems unsuitable for translation purposes (and it must be admitted that these systems have been used successfully for rapid processing on the basis of 'quasi-translation' of enormous quantities of documents for information purposes and that they will be bequeathed to posterity along with the rest of the arsenal of the cold war) all that remains is to review the various criteria for acceptance, starting with the most important ones.

Defects leading to the distortion of the technical content (misinterpretation and wrong meaning or loss of it (nonsense and qobbledeqook)

A system which is satisfactory from the syntactical point of view must furthermore render the exact meaning of each of the following categories of words:

1) the 'grammatical' words: this category covers conjunctions, prepositions, adverbs, and in particular such polyvalent words as 'as, since, when, while,' which leaving aside homoqraphy, cannot be translated correctly without appropriate semantic features. Any error in translation of these words involves distortion or loss of meaning and requires considerable modifications to the sentence. Furthermore, a certain proportion of these errors could possibly go unnoticed by the average reviser. Consequently, any system which does not correctly translate the grammatical words is UNACCEPTABLE.

2) Non-technical words which take on a particular meaning according to the technical context. The main verbs in this category are 'to drive, to replace, to supply' etc. and nouns such as 'assembly, unit, area' etc. These words are the greatest source of misinterpretation and often have contradietary meanings. For example, 'to replace' can mean 'return to original location' or 'substitute' and only a specialist will be able to detect any misinterpretation. It is mainly because of the difficulty of translating these words that a machine translation system would have to be designed for a defined field of application, in other words a limited field, and that the 'experience of the translator' would have to be incorporated into the system in the form of a 'semantic thread' in order to obtain correct translation in each case. If a semantic thread is not included, the output is comparable with the work of a translator without any knowledge of the subject and those unable to understand the original text. The misinterpretations arising from these 'non-technical' words can be corrected only by a specialist translator and a certain number of errors could possibly go unrevised. Consequently, a system which does not correctly translate frequent non-technical words is UNACCEPTABLE.

3) Technical words and expressions. Wrong translation of a technical term could cause absurdities or misinterpretations. Only a specialist translator can reestablish the meaning but in the second case the error could possibly go unnoticed during revision. Simple technical words such as pin, valve, pad, etc. have a considerable range of quite precise meanings which cannot be identified without the aid of semantic rules similar to those used for non-technical words. Furthermore, technical expressions which may contain more than ten words are not within the grasp of the average 'reviser/editor'. Consequently a system which does not translate correctly technical words and expressions is UNACCEPTABLE.

Minor defects or flaws not affecting the technical content (wrong usage, anglicisms, non-agreement, articles, etc.)

These are the sort of faults which can be expected from a good automatic translation system and which the reviser/ editor would have to be ready to correct.

Only defects inherent in automatic translation have been examined. Obviously errors in coding or use of the system which do not fall within the limited indicated could adversely affect the quality of the translation.

Conclusion

If a technical translation is to come up to reasonable standards of quality, an automatic translation system must be used which is equipped with an accurate syntactical analyser and where the output can be revised by a translator without specialist knowledge of the subject. Thus correct translation of 'grammatical' words, frequent non-technical words and technical words and expressions is INDISPENSABLE.