Automatic Translation as a Model of the Human Translation Activity

ZOYA M. SHALYAP1NA

Institute of Oriental Studies, USSR Academy of Sciences, Moscow, USSR

The process of human translation is described in a formalised way as an organised sequence of operations, the main of which are: (1) perception and understanding of the original text; (2) interpretation and comprehension of the text from the point of view of the situation it describes; (3) translation proper ensuring the transition from the units of the source language to units of the target language; (4) construction of the target text by means of combining the translation equivalents of all separate units of the original into a connected text according to the rules of grammar of the target language; (5) evaluation of the adequacy of the resulting translation; (6) evaluation of its linguistic acceptability; (7) correction and modification of the translated text in case it is found to be inadequate or linguistically unacceptable. For each of these operations, kinds of formal procedures are indicated which can be regarded as their analogs in a system of automatic translation. It must be emphasised that one of the necessary conditions for any concrete AT system to constitute, in principle, an adequate model of the human translation process, is that the elaboration of even the initial version of the system (realising only part of the procedures mentioned) should be based on a sufficiently clear idea of the direction and methods of its subsequent transformation into a complete system, including all of the components in question.

One could define automatic translation (AT) as a kind of automatic text processing aimed at transforming a text in a certain natural language into a text in a different natural language with an equivalent semantic content. The same aim is pursued by human translation as well, so that insofar as the input text of an automatic translation system retains a natural language nature (after possible pre-editing procedures) and the output text of the same system can be considered to be an adequate translation of this input text, we may say that the functioning of such a system is a model simulating human translating activity.

Generally speaking, the structures of the functional model and its original may have absolutely nothing in common; it is sufficient if the application of the two systems in question to the same input objects (or data) yields results complying with the equivalence criterion used (a simple example of such a relation between two fundamentally different systems is given in [1, p. 28]).

In the case of automatic translation, however, it is important that the elementary operations used by the modelling system to carry out the required process (the operations of searching for certain elements of information, recording them, comparing the records available, modifying these records, etc.) do not include operations which could be said to have no functional analogies with the operations performed by human translators in their work. It seems natural to conclude that there must also exist fundamental similarities in the organisation of these operations in the AT systems, on the one hand, and in the process of translating natural language texts by human translators, on the other hand.

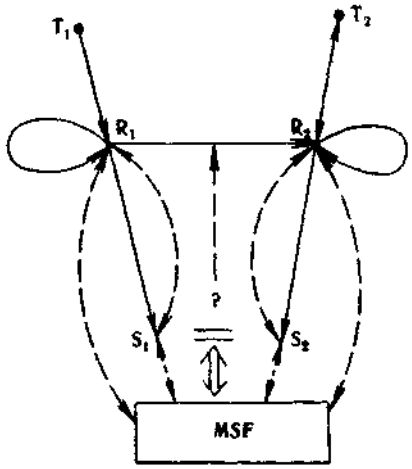
If we now take into account the fact that people as a rule try to organize their activity so as to use the means and instruments available to them in the most efficient and economic way, we may surmise that as far as the systems simulating this activity on the basis of functionally analogous means are concerned, the closer the internal structure of their functioning approximates the structure of the activity in question, the higher the adequacy and efficiency can be expected to be.

That is to say that an important direction of research in the field of automatic translation is that of trying to understand and formally describe human translating activity as an organised sequence of operations and procedures that could be simulated by some analogous operations in AT systems.

We are going to present here a version of such a description of the so-called 'written translation' process, as well as some conclusions relative to the possible composition of an AT system based on the model proposed. Our description may be called an integral formal model of the translation process.

The essence of our conception may be graphically represented by the following diagram*.

* An earlier (and simpler) version of fundamentally the same diagram as well as its general substantiation, is given in [2].



where T_1 stands for a text in the first (source) language; T_2 designates a text in the second (target) language constituting the translation of text T_1; R_1 and R_2 are the respective intermediate formalised representations of the two texts; S_1 and S_2 are the semantic representations of the texts in question; MSF is the model of the subject field with which text T_1 *is* concerned.

According to this diagram, the translation process includes the following principal operations.

1. Perception and understanding of the initial (source) text, resulting in the translator making out its meaning. In an AT system this operation will correspond to the automatic analysis of the text to be translated, designated in the diagram by the line $T_1 \rightarrow S_1$. From the formal point of view, such analysis consists of the transition from text T_1 as a sequence of graphic symbols and, possibly, blanks, to the semantic representation of the same text, aimed at providing the fullest possible explication of the semantic links and relations existing in the text (among other things, this representation is expected to contain data concerning the recurrence of various semantic elements within the meanings of different words, which forms the basis of the operations of redistributing these elements when using translating devices like synonymous transformations, logical deductions, etc.).

In some cases, however, it is possible to obtain an adequate translation of a text even if the understanding of this text by the translator is incomplete — for instance, if the translator comes across certain words in this text (special terms, realia, etc.) which he knows how to translate although their exact meaning remains obscure to him. With this fact in view one may conjecture that even in dealing with texts for which he can reach the deepest possible level of understanding (corresponding to level S in our diagram) the translator may, if the selection of the appropriate translation equivalents presents no particular problems here, restrict himself to a more superficial consideration of the text he translates, thus saving time and effort in performing the translation*.

In the diagram under consideration this situation corresponds to a reduced variation of the analysis process which does not result in a full semantic representation of the text to be translated. It stops at some intermediate level R (the segment $T_1 \rightarrow R_1$) and draws on the semantic information on the text (or, rather, on some of its fragments for which this information is accessible under the circumstances) only for the sake of more accurately representing the text in question at this intermediate level (the dotted line $S_1 \leftrightarrow R_1$).

2. Comprehension and interpretation of the text concerning the essence of the situation involved. The independent status of this operation becomes especially clear if we consider the problems of translating scientific and technical texts.

For the situational interpretation of the text to be translated the translator uses information about the subject field this text refers to, rather than about the language properties of various text units. Based on this information, he judges the adequacy of his conception of the text and chooses the optimum direction for further actions. To do this, the translator must either be a specialist in the given field himself (as required in many translation bureaus — see, for example, [4, p. 5]), or he must have the opportunity to consult such specialists, look up various encyclopaedic reference books, familiarise himself with the appropriate literature, etc.

In an automatic system the information relating to the subject field of the text may be drawn from some formal model of this field, represented, say, in the form of a thesaurus, a semantic network, or any other data base. Actually it means that an AT system must somehow interact with an artificial intelligence system designed for the same sphere of knowledge. Interpretation will be provided by checking whether there exists a correlation between this model and either the full semantic representation of the text in question (line $S_1 \rightarrow MSF$) or some intermediate representation of it (line $R_1 \rightarrow$MSF), depending on the structure of the model as well as on the level reached by the analysis process by the time the check is performed. Any contradictions revealed in the course of this procedure will point to the necessity of verifying and modifying the results obtained from the analysis. The simplest of such modifications is just eliminating the analyses having contradictory interpretations, and following alternative paths. In more complex cases transition to deeper levels of representation may be contemplated*.

3. Translation proper, providing the transition from the source language units to the target language units. In an AT system this is done by such operations as selecting the translation equivalents for the units included in the formal representation of the original text constructed in the course of the analysis process (in the simplest case — extracting the only equivalent associated with a certain unit from the data files available) and substituting these equivalents for the source text units, thus forming some sort of a translation text representation (line $R_1 \rightarrow R_2$).

What is important here is the way the translator solves the problem of what kind of units are most sui-

* It is this attempt to save the effort of fully understanding the text that is most likely to be one of the sources of such translation errors as 'literal' translation (see, for example. [3, p. 186—187]) if the translator underestimates the language difficulties involved in choosing the correct equivalents for the text units.

* Results of the situational interpretation of the text to be translated may also be used to check and improve the analysing rules (if, say, these rules are found to yield no results with a permissible interpretation when applied to a text known in advance to be correct); or to refine the model of the subject field itself (if the text contains some information overlooked in this model). These operations, however, reflect the process of language learning and the process of acquiring knowledge about the world, rather than the process of translation as such. Therefore they are not included in our translation model.

table as the basis of the transition in question: whether it is advisable, for instance, to use direct lexical correspondences or it would be wiser to give them up and to translate 'according to the meaning'. For an AT system this problem amounts to choosing one of the formal text representations envisaged by the analysis as the level at which the above operations of selecting translation equivalents and substituting them for the original units must occur. We can mention two main factors which seem to have a direct bearing on the problem under consideration.

On the one hand, the decisions taken by a translator will evidently depend on his general translation habits which, if we neglect various subjective aspects, are dictated primarily by the extent of the typological similarities and distinctions characterising the two languages in question (and reflected in texts belonging to the given subject field). In an automatic system the function of these habits is taken by the inventories of the grammatical and dictionary units provided with translation equivalents. In most cases such inventories will be supplied apriori, so that this factor is likely to play an active part only in the course of compiling or updating the linguistic information for a certain system; the functioning of the system may be influenced by this factor only in case it incorporates some learning devices.

On the other hand, it seems obvious that in deciding on the level of translation proper in a concrete case one cannot ignore such a factor as the degree of 'idiomaticity' of the text to be translated (or of its various fragments). If, indeed, for a given text fragment the syntactico-morphological means of the two languages can be matched accurately enough, one can translate 'word for word' or even 'morpheme for morpheme' (see [3, p. 179—181]). To put it in terms of automatic translation, the system can limit itself to the level of the morphological representation of the text — the level which reflects the results of the text's segmentation into separate word-forms as well as of the morphological analysis aimed at obtaining the lexico-morphological composition of each of these word-forms, but has no explicit indications as to the semantics of the word-forms themselves or the syntactico-semantic relations existing between them in this text

Now if such a limitation is found to be too strict in that it prevents the system from achieving the adequacy required from it, it is natural to consider correspondences of a more complex nature involving certain structural transformations, or even to resort to descriptive or interpretative translation. In an AT system it will be a question of performing the operations of translation proper at the level of various types of lexico-syntactical or semantico-syntactical structures (in which case separate translation equivalents are to be provided not only for lexical units and morphological characteristics, but also for the semantico-syntactic links and relations found to exist in the given text between the words it contains), at the level of semantic representation (which makes it possible to change the distribution of various meaning components between the units of the translation text as compared with the distribution of the same components between the units of the original text), and at the level of situational interpretation (providing, among other things, for inferences, conclusions and other operations of a logical nature), respectively.

On the whole, there seems to exist a tendency toward using (inasmuch as it does not contradict the requirement for translation adequacy) the most superficial levels possible — which, on the one hand, saves the translator time and effort (see point 1 above), and, on the other hand, helps him to reflect in his translation not only the meaning of the original text, but also some features of the expressive means used to convey this meaning (cf. the recommendations concerning the methods of translating phraseologisms given in [5, p. 92]). This tendency may be seen, for example, in the fact that even if translation within the limits of the level chosen by the translator (usually the lexico-syntactical level) presents certain difficulties (such as, say, absence of direct translation equivalents for some of the text units recognised at this level), he often rejects the methods of descriptive or interpretative translation in favour of a special translating device amounting to paraphrasing the text fragment under consideration in order to obtain a synonymous fragment differing from the original in that the adequate equivalents can be found for its units at a more 'superficial' level*. In an AT system similar possibilities can be provided by supplementing it with a subsystem of synonymous transformations of the source text, its functioning being limited to the representation level most often used for translation proper (cycle $R_1 \rightarrow R_1$).

4. Construction of the translation text by means of integrating the translation equivalents of all separate units of the original into a unified text according to the rules of grammar of the target language. If these rules do not accept some of the combinations of the equivalents obtained, the translator may resort to synonymous expressive means. The corresponding formal operation is the generation (synthesis) of the text required, understood as the transition from the intermediate representation resulting from the substitution of translation equivalents for the source language units, to a sequence of actual word-forms and punctuation marks in the target language (line $R_2 \rightarrow T_2$); the subsystem of synonymous transformations (cycle $R_2 \rightarrow R_2$) may be used if necessary.

The level of the intermediate representation R_2 will obviously depend on the level of the units substituted for translation equivalents (level R_1), although, generally speaking, these two levels need not coincide. It stands to reason that the farther this level is removed from the level of the text to be formed (level T_2), the more complicated the generation process will be. This process becomes particularly complex if the level in question is based on units specifying smaller components of text meaning than a separate word, so that one is faced with the task of choosing the correct lexemes. It is the solution of this very task that is probably one of the essential aspects of the so-called 'throes of wording' — the situation where a person knows what it is he wants to communicate to his reader or listener, but cannot find the appropriate expressive means. Therefore, within the framework of AT as well, in spite of the often inevitable inaccuracy of lexical translation equivalents, it seems justified to consider lexemes and units, compatible with

* This device is most widely used when translating from one's native tongue into a foreign language (see, for example, observations on the role of synonyms in the translation process given in [5, p. 190] or the description of the 'simplified' translation method in [6, p. 148]).

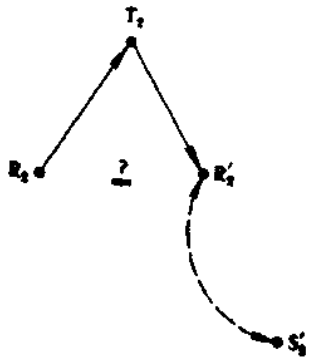
lexemes at the same level of text representation, the central units of translation. This decision seems all the more advisable as it makes possible the development of AI systems starting from the data already available in the usual ('human-oriented') dictionaries, grammars and various scientific papers (those concerned with research in the field of the theory and practice of translation, in the first place), known to be based in most cases on the notion of the word or some functionally analogous unit.

5. Evaluation of the adequacy of the translation obtained. The necessity of such evaluation as a separate translation procedure is motivated primarily by what we have already noted about the approximate nature of lexical translation equivalents (holding both for human translators and for automatic translation systems). In many cases it is actually far from safe to count on the absolute accuracy of such equivalents (moreover, as shown by the practice of translation, outside the range of terminology, absolute equivalence is almost an exception), and if several inaccurate equivalents co-occur inadvertently in a translation of one and the same text fragment, we run the risk of significantly changing the original meaning and giving rise to misinterpretations of the text we have translated. Serious misinterpretations may also result from 'literal translation correspondences', which a translator might use in the course of translation proper, fascinated by the linguistic peculiarities of the original (which he can easily notice and correct when re-reading the text obtained). In AT similar errors may be committed if the text representation chosen as the level of substituting translation equivalents for the source text units is not 'deep' enough. If, say, this representation does not provide for explicit description of elliptical constructions (filling in the units omitted) the Japanese sentence *Ichigatsu-wa Kyūshū-wa nanajūgo miri desu* will be translated as awkwardly as *In January Kyushu is 75 mm* (according to the context of this example [7, p. 150], it is actually the amount of atmospheric precipitation that is meant).

The formal procedure for evaluating the adequacy of the translation can be thought of as a combination of two main operations: first, the representation obtained for this text as a result of translation proper should be analysed with the object of explicitly stating its meaning (line $R_2 \rightarrow S_2$); second, the resulting semantic representation of the text to be evaluated should be checked against the similar representation of the original text ($S_1 = S_2$). Such checking may also involve interpretations of these texts with respect to their common subject field (in our diagram this possibility is shown by the double arrow drawn vertically between the sign of equality and the MSF), Inadequacy of the translation will be indicated by a disparity between S_1 and S_2 considered significant for the text type in question — that is, by the situation where the distance between S_1 and S_2 (in the mathematical sense of the word) exceeds a certain given threshold (the determination of this threshold being part of the work on the linguistic information for an AT system in the course of its development).

6. Evaluation of the language acceptability of the translation obtained. Besides stylistic considerations, this operation is dictated by two main factors. On the one hand, the translation obtained from the synthesis process may contain ambiguities due to accidental coincidence of some translation equivalents of unambiguous source language units in the target language—which may present difficulties in the interpretation of this text or even cause misinterpretations. On the other hand, dissimilarity of combination rules in the two languages (including rules of semantic combinability of their respective units), as well as semantic shifts brought about by the use of inaccurate equivalents, may give rise to certain combinations of words and constructions in the translated text that will be irregular from the point of view of some norms or standards of the target language, thus making this text difficult to understand. An experienced translator is usually able to detect defects of both kinds when re-reading the text he has written. In an automatic system the same could be done by means of checking representation R_2 formed in the course of translation proper and functioning as the basis for the synthesis process, against a formal representation of the same level, which the system can build by analysing the target text it has synthesised. It will be sufficient to use a reduced analysis procedure (see point 1 above) stopping at the R level (this is not to say, however, that information of deeper levels could not be drawn upon in deciding on certain concrete features of the representation being built — the same as with the reduced variation of the process of analysing the source text). The normativity (regularity) of the text will show itself here primarily as the very possibility of analysing this text by the rules available; if, in addition, the system contains data as to the frequency of occurrence of various grammar and dictionary rules of the target language in texts considered to be standard by the speakers of that language, the degree of the translation text normativity can also be assessed according to whether the rules involved in this case are more or less typical of the appropriate type of texts. As to undesirable ambiguities, their presence or absence in the translation text is revealed, respectively, in the presence or absence of contradictions observed between representation R_2 used in synthesising this text and the 'most natural' of its R-level representations obtainable from its analysis (in the case of ambiguities of this kind one would expect higher priority to be given by the analysis process to a representation inconsistent with R_2),

Generally speaking, this group of operations should be expressed in our diagram as:



In the diagram, however, we preferred, for the sake of clarity and simplicity, to represent these operations in a simplified form: $T_2 \rightarrow R_2 \leftrightarrow S_2$, corresponding to the specific situation where representations R_2 and R_2' match

(and so the translated text obtained can be considered linguistically acceptable).

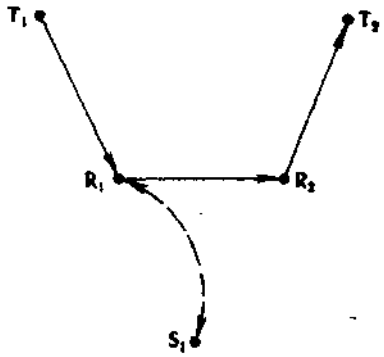
7. Modification of the translation text performed in case it is semantically inadequate or linguistically unacceptable. In connection with the human translation activity all operations of this type, classed as editing, are considered to be a single notion. In a formalised description, however, it seems more appropriate to distinguish between two groups of operations: operations of correcting the translation text if it is inadequate, and operations of editing proper — understood as refining a semantically adequate text from the point of view of the expressive means used (cf. [8, p. 20—22]).

Correction of an inadequate translation will amount to returning to the phase of translation proper (it is shown in the diagram by a vertical dotted line). Depending on the scale and nature of the discrepancies discovered between S_1 and S_2, one would expect here either a double check of the previously selected equivalents and, if found necessary, a replacement of some of these by different ones, considered to be more suitable in the given context; or, probably, a reconsideration of the entire procedure used in this concrete case to provide translation proper, up to abandoning the corresponding route of translation as a whole and repeating the operations of selecting translation equivalents at a 'deeper' level of text representation. In both cases the subject field data may be taken into account (lines $R_2 \leftrightarrow MSF$, $S_2 \leftrightarrow MSF$).

In general, the operations of editing the translation text may also involve switching to alternative translation equivalents of some source text units (followed by an obligatory check-up to make sure that the substitution of these alternative equivalents for the previous ones causes no semantic changes unjustified by the context). However, it seems that a better and more natural formal analog of this group of operations is furnished by the functioning of a system of synonymous transformations for the target language (cycle $R_2 \rightarrow R_2$). Ideally, such a system should have facilities for replacing separate words or word-combinations by their absolute or partial synonyms; modifying the grammatical constructions available; omitting units that are redundant according to the ellipsis rules of the language in question, as well as filling in the units that are not allowed to be omitted; breaking down cumbersome sentences into several simpler ones, etc. After the defects detected while evaluating linguistic acceptability of the translation have been removed, the functioning of the modification system should stop and the resultant text be output as the final version of the translation.

According to our description, it is only on condition that all of the operations enumerated above have been carried out that the process of translation can be regarded as completed and its results as wholly reliable. At the same time the description makes it clear that many of these operations (such as, say, interpretation of the text to be translated; evaluation of the adequacy and linguistic acceptability of the translation; its correction and editing) are actually necessary only in more or less complex situations where the simplest translation correspondences and the standard means of analysis and synthesis alone are inadequate to provide a good translation. It is also common knowledge that a human translator does not acquire the ability to cope with such difficult cases at once: not only special training in using various translation methods and devices, but also a certain amount of professional translating experience is usually required to acquire this ability. In other words, a person often begins work as a translator before he has mastered all the necessary techniques. During this period his translations need rather painstaking editorial correcting, the editor often duplicating his work — the analysis of the source texts included — due to the necessity of constantly checking the translations obtained against the original. It is only later that the translator becomes skilled enough for such duplication (obviously detrimental to the efficiency of the whole process) to become unnecessary.

Likewise, when developing an AT system, it would hardly be advisable to demand that it begin to function only after all the subsystems simulating the above components of the translation process have been fully worked out and incorporated in this system in their final form. A more reasonable approach would be to initially build up some simplified version of the system embodying just the 'nucleus' of the schematic proposed, so that the rest of the components envisaged could be gradually added to this nucleus as improvements in an already functioning system. It is clear that the minimum nucleus of an AT system corresponds to the fragment $T_1 \rightarrow R_1 \rightarrow R_2 \rightarrow T_2$ of our diagram. To the best of our knowledge, it is either to this minimum or, in some cases, to a somewhat extended fragment:



that all of the existing AT systems (whether already functioning or still under development) are confined; all the other operations are left to be carried out by a human inter- or post-editor.

As we have already mentioned, within the framework of the conception set forth in this paper, during the initial period of developing a concrete AT system, such a restriction is rather an advantage than a disadvantage, as it makes it possible to embark on practical employment of the system before it has been completed and, testing in action the solutions contemplated for various problems, decide on their efficiency and, hence, on the necessity to modify some of them prior to enlarging and improving the system as such. However, the fact that the process of translation as a whole covers a wider range of operations than those belonging to the 'nuclear' fragments of the general schematics, is of decisive importance in even constructing the very first modifications of the system, with all the simplifications that can be thought of. The point is that this imposes a general re-

quirement on any such modifications, however simplified, — the requirement of their compatibility with future subsystems, intended to provide for all the rest of the operations implied in the concept of the translation activity—such that the linking up of existing modifications in the system with the components added later will not involve fundamental internal rearrangements affecting its original structure and the organisation of its functioning. This means that even in working on the first version of an AT system its designers must have a sufficiently clear idea of the main directions and methods of its subsequent development. If we are to consider the system under construction a fundamentally adequate model of human translating, this seems to be a condition *sine qua non.*

REFERENCES

1. Gastev, Yu. A. *Homomorphisms and models. Logico-algebraic aspects of simulation.* Moscow: Nauka Publishers, 1975, 150 p. (in Russian).
2. Shalyapina, Z. M. On the problem of building a formal model of the translation process. In: *Translation theory and scientific grounds of training translators. Materials of the All-Union Scientific Conference. Part II.* Moscow: MGPIIYa, 1975, 165—172 (in Russian).
3. Barkhudarov, L. S. *Language and translation. Questions of general and special translation theory.* Moscow: International Relations Publishers, 1975, 240 p. (in Russian),
4. *Guide in methods for non-staff translators and scientific editors of the All-Union Centre for Translations.* Moscow: VTsP Press, 1976, 144 p. (in Russian).
5. Sobolev, L. N. *An aid in Russian-French translation.* Moscow: Foreign Literature Publishers, 1952, 403 p. (in Russian).
6. Revzin, I. I.; Rosenzveig, V. Yu. *Fundamentals of general and machine translation.* Moscow: Vysshaya Shkola Publishers, 1964, 243 p. (in Russian).
7. Golovnin, I. V. (ed.). *A course in Japanese. Part II.* Moscow: Vysshaya Shkola Publishers, 1973, 320 p. (in Russian).
8. Dokshtein, S. A.; Makarova, E. A.; Radominova, S. S. *A practical course in scientific and technical literature translation.* Moscow: Voenizdat, 1967, 494 p. (in Russian).