[From: *International Social Science Journal* 19 (1967), 64-78.]

Linguistics and automatic translation

I. A. Melchuck

Twenty years ago the very expression 'automatic translation' (or machine translation) would have seemed meaningless to the linguist and, no doubt, to any reader.

Ten years ago, 'automatic translation' (AT) was discussed only by a small circle of specialists, while the scientific general public viewed it with curiosity, amazement and even perplexity. Today it has become a familiar concept; it is known and discussed everywhere; scientific works and popular articles are written about it; the main linguistic journals of the world publish papers on AT and the universities of at least ten countries include it in their philology courses.

However, familiar does not mean comprehensible. So far there is no single, universally adopted conception of AT and its place among other disciplines. Moreover, there is a very widespread, very narrow and, in our opinion, mistaken attitude, which consists in regarding AT purely as an applied discipline and primarily as a technical problem with a practical orientation (i.e., directed towards industrial and economic purposes). In other words, while AT is acknowledged to be of great value and no less scientific interest, it is conceived as a kind of linguistic engineering, somewhat analogous to the manufacture of a new type of electric razor or the technical development of a new product.

The purpose of this paper is to give a methodical and well-founded exposition of another conception of AT, without presenting an all-round view or going into all the fundamental problems and concrete results achieved in recent years. Without being too technical, with the minimum of references and concentrating solely on the ideological aspect, we shall try to demonstrate, on the basis of well-known facts and common sense, that in principle the solution of AT problems may be reduced to the construction of exhaustive operational models for language in general and a series of natural languages in particular, regarded as means of communication, and this seems to us to coincide with the central problem of synchronie linguistics. And so there appears to be no clear-cut natural boundary between AT in the broad sense and synchronic linguistics (both theoretical and descriptive linguistics).

In the early years of AT, linguists working in this field stated their problem as follows: 'We want to teach a universal electronic computer to translate texts (mainly technical texts) from one language into another. The human translator can do this by making use of the information he possesses on the two languages and their word equivalents (such data—the sum total of all essential data—are contained in full in grammars, dictionaries, etc., and are assimilated by the student during his study of the language). Translation may be regarded as a simple elementary operation. Accordingly, all we have to do is to present the data in our possession on a pair of languages, word equivalents and the translation process, in a form accessible to the machine, namely in the form of an algorithm adapted to a particular pair of languages and to a given machine. The task of the linguist working on AT may be summed up in the elaboration of algorithms of this type.'

Let us consider the principles implicit in this view:

1. There are in existence entirely acceptable theories of language and translation and in addition sufficiently complete and adequate descriptions of natural languages; it is not necessary to modify them in essence, but merely to adapt them to the practical problem of AT.
2. Translation is a fairly simple linguistic operation, perhaps even the simplest of all; consequently it is here that we should start with the mechanization of linguistic activity and in general of intellectual activity.
3. All necessary data concerning a language should be presented by the linguists in the form of special descriptions—algorithms capable of the process of analysis or synthesis of given linguistic objectives (as opposed to the 'static', 'classificatory' descriptions of traditional linguistics). For this reason linguists have concentrated on command routines and on details of text processing, as manifested in the search for 'operational morphology' and 'operational syntax'.
4. When constructing an algorithm, the linguist must have in mind a given computer or at least the performances of existing computers. For this reason linguistic descriptions were established in the light of low-capacity machine memories and/or command routines for specific machines (which explains the frequency of the terms 'cell', 'address', etc., in early linguistic papers on AT.

However, even in the first years of research on AT it was realized that the problem could not be tackled in this manner.

At present most researchers start from quite different premises, which may be summarized as follows.

In the first place, there is no 'ready-made' theory of language and no 'ready-made' descriptions of natural languages which merely have to be applied, or to some extent adapted to AT. Moreover, in linguistics little attention has been paid to how a man understands the meaning of a text and how he expresses the thought he needs. But this is of vital importance for AT. In addition, the degree of strict logic and accuracy in traditional linguistic work was below the level required by AT.

In the second place, translation is not a simple linguistic operation: it may be divided into two stages, comprehension (analysis) and expression (synthesis). In order to translate a text, we must first understand it (if only superficially), that is to say, we must extract from it a certain message which must be conserved in translation; and this message must then be expressed appropriately in the target language. All AT algorithms elaborated so far (several dozen!) break down into analysis and synthesis, which is not a matter of chance, but corresponds to the nature of things. For this reason, before there can be any question of elaborating high-quality algorithms for automatic translation proper, it is necessary to start by developing satisfactory algorithms for automatic analysis and synthesis. For this purpose a convenient notation system must be devised, in order to record the results of analysis and the initial data of synthesis, which makes it necessary to define the required depth of analysis (and therefore of the beginning of synthesis). In other words, the most urgent problem facing linguists working for AT, after choosing a satisfactory practical depth of description from the syntactic or semantic standpoint, is to elaborate formal rules for passing from text to description and vice versa, i.e., to construct dynamic models of comprehension and expression.

In these circumstances translation at syntactic level only is often impossible; in order to obtain high-fidelity AT it is necessary somehow to detect the message so as to secure 'meaningful' translation and to create models of comprehension (analysis) for the transition 'text → meaning' and models of expression (synthesis) for the reverse operation 'meaning → text'. It is quite clear that at present most research on AT concentrates on automatic analysis and (to a less extent) on synthesis. Translation as such remains at present on the fringe of scientific attention; in fact, automatic translation is left without any translation!

In the third place, it seemed reasonable to separate the description of language units proper, their meaningful (or syntactic) characteristics, and the equivalences between the two, from the description of the process of moving from units to characteristics, and vice versa.

We have in mind the breakdown of AT systems (and in general the automatic processing of a text) into 'grammar' and 'mechanism', or 'grammar' and 'universal algorithm', or 'table of linguistic constants' and 'algorithm proper', etc. This breakdown is in fact adopted and practised everywhere. More specifically, the first part, 'grammar', is concerned with the description of the actual language (units, characteristics, equivalences) and belongs entirely to the linguist's province. The second part, 'mechanism', models the processing which the human user does on the data in his head, and this part has to be studied not only by linguists, but also (perhaps even in the first place) by psychologists and mathematicians. In particular, the search for the optimum process for moving from objects of one type to objects of another type (e.g., from natural language sentences to their syntactic structures) may be interpreted as a strictly mathematical problem, analogous to problems of linear programming or discrete analysis.

At the present time, linguists working on AT rightly concentrate on the construction of natural language 'grammars' and their theory. Accordingly, in the most important, purely linguistic sector of AT there are no algorithms in the strict sense, although even quite recently these seemed to be the dominant feature of linguistic research on AT.

In the fourth place, while working on natural language description and/or language theory for AT purposes, the linguist should not restrict himself to one particular computer or even, more generally, accept the limits of existing computer performances. Such limitations prove artificial; they deform reality and, however paradoxical it may appear, make the achievement of computer linguistic descriptions more difficult.

The only demand made on the linguist is that his descriptions shall be formulated logically, strictly and exhaustively; the practical side, adaptation to a given computer, constitutes a separate problem, which the mathematicians will solve all the more easily if the basic description is not deformed by taking into consideration the number of 'bits' in a memory cell or the specific range of computer routine. Linguistic research on AT does not as a rule depend, at least as regards theory, on the actual computer; the logical machine is felt here as a necessary scientific abstraction; potential computer performance is an adequate prerequisite ensuring the strict logic (though not the truth!) of linguistic descriptions.

[*NB.* Collaboration between mathematicians and linguists has had interesting results of quite another kind; it develops habits of precision in the latter and teaches them to observe facts correctly. Moreover, it must not be forgotten that in linguistics a computer has an irreplaceable role as a powerful instrument of research—rather like a telescope in astronomy; this point will be discussed later.]

And so automatic or machine translation (at least in its linguistic aspect) subsists even without a machine[1].

Automatic translation with no translation, no machines and no algorithms?

What then should be the content of subsequent linguistic work on AT? In the author's view, there is only one possible answer: the elaboration of operational models of language, logical systems giving multiple-meaning equivalences between text and meaning in both directions. That is the problem which linguists working on AT are in fact now trying to solve and on their success will depend the practicability of AT in the strictest sense of the term.

It seems difficult to deny that the above problem is essentially one of linguistics. The study and description of the relationship 'text-meaning' in all its aspects (including the historical, social, psychological and other

1. This is one reason why we prefer the term 'automatic translation' to 'machine translation'.

aspects) has always been recognized as the central problem of the science of language.

It is not by chance that in recent years the main efforts of linguists (apart from AT) have been concentrated on the elaboration of operational models of language (cf. the most authoritative trend in modern linguistics—the theory of formal grammars, primarily associated with the name of N. Chomsky). The problem of AT thus coincides in principle with the problem of synchronie linguistics as such. This is confirmed, in particular, by the fact that most linguistic papers on AT published from 1963 to 1966, even in such highly specialized journals as *La Traduction Automatique,* 'Научно-техническая информация" (Scientific and technical information), etc., or in the proceedings of the Conference on Computational Linguistics (New York, May 1965) are essentially ordinary linguistic work. There is nothing surprising in this. On the one hand, the development of cybernetics has led quite naturally to the problem of teaching the human language to logical machines; this is the fundamental way of improving them and creating real 'thinking machines'. On the other hand, the autonomous, inherent development of linguistics has led to the conception of operational models of language (generative/recognition grammars). But the first problem can be solved only with operational linguistic models, and these require the use of computers if they are to be verified and improved on a sufficiently wide scale. We must point out that a computer is also necessary in 'pure' linguistics, as a research tool which not only saves human labour but can process a previously inaccessible volume of data; moreover (and this is the main point) it can undertake the experimental verification of linguistic models, which can probably not be done by any other means.

In this connexion we would mention the interesting work on automatic sentence synthesis based on generative grammar (V. Yngve, D. A. Dinneen, A. G. Satterthwait, N. N. Arsenteva).

Nevertheless, although the fundamental problems of AT and linguistics coincide, AT is still to some extent a separate sector, possibly for the following reasons.

At the present time we may detect two trends in AT—a theoretical one deriving from the conception set out above, and an empirical one aiming at constructing 'direct' (binary) algorithms as quickly as possible, even if they are very far from perfect. Most groups and most experts (in the U.S.S.R., United States, United Kingdom, France and elsewhere) follow the first trend; however, the same countries also have centres engaged in constructing and testing binary algorithms—without aspiring to solve the general problems of AT, but in the hope of perfecting these algorithms to the point of obtaining usable translations in practice. In so doing, they are accumulating valuable experimental material. Owing to the existence of this empirical 'wing', AT has not been entirely merged with linguistics.

The theoretical 'wing' of AT also differs from ordinary linguistics in that it specifically envisages language from the standpoint of translation, i.e., transformation of meaning. This attitude forces AT linguists to concentrate on equivalences of text and meaning (or at least, meaningful connexions—syntactic structure), to study and to describe first of all the 'text-meaning' relationship. Of course the argument asserting the leading role of this relationship is recognized, as we have already seen, outside the sphere of AT; but while it is recognized in theory, it is often forgotten in practice. In AT it is impossible to forget it. The 'translation approach' imposes a rigid frame of reference, which compels the researcher to move only in a useful direction, to try to model the linguistic behaviour of the speaker precisely in order to conserve the 'text ↔ meaning' exchange.

In the near future the 'translation approach' will doubtless become essential even in linguistics. Not without reason does an authority like R. Jakobson emphasize the fact that meaning cannot be defined otherwise than by translation ('meaning is what is translated', 'meaning is the invariant of synonymous transformations, and in particular of translation'), and that translation is one of the basic linguistic operations; N. Chomsky's transformational grammar also strengthens the 'translation approach' in linguistics (Chomsky's transformations are of a special type; the dynamic process of transforming, of obtaining something from something else, is characteristic of the 'translation approach' as opposed to the static, 'classificatory' approach of former linguistics).

However, at present the systematic use of the 'translation approach' *is* still a prerogative of AT.

Linguistics cannot be reduced to mere synchrony: a language must be studied as a phenomenon variable in time and space, in its normality and in its pathology, in its social aspect and in its individual aspect, from the standpoint of its aesthetic function, etc.

In the context of AT, however, all these aspects of language study have not yet found their natural place. That is why AT remains separate from linguistics as a whole, although drawing closer to and merging with synchronie linguistics.

Lastly, we cannot overlook the special status of AT research centres in all countries with regard to administrative organization.

Moreover, it is clear that there is no impassable natural boundary between linguistics and AT. There is only one science of language, with one main objective: to create an exhaustive operational model of the language in all its aspects. The first stage towards this model is to solve the problem of constructing a 'text ↔ meaning' system, i.e., a language model considered solely in its essential aspect, communication.

When this problem is solved, then and only then shall we see the birth of 'real' AT as an applied discipline, its subject-matter being to create economic, and practical translation systems, based on established linguistic theory.

Until then 'AT' remains a vague (and perhaps even puzzling) term designating a complex research network; summing up what has been said above, we may characterize it as follows.

AT is not an applied discipline aiming at 'practical' results; it quite simply has nothing to apply! On one side it appears to be an experimental sector, but on the other, and most important, side it is a theoretical discipline dealing with the construction of operational linguistic models and thereby represents an important trend in modern linguistic theory (similar in many respects to the theory of formal grammars).

We should emphasize that the very manner in which the problem of AT is presented encourages the researcher to explore a series of ideas of fundamental value to linguistics. In the first place, there is the functional and active character of the models, as opposed to purely descriptive, taxonomic and static models. We do not claim that the concept of generative and recognition grammars in linguistic theory is solely due to the influence of AT. This concept is deeply rooted in linguistics, and derives quite normally from the essence of language. Nevertheless, it has been necessary to clear the way and do battle in its defence and it has not yet conquered the main body of linguists; whereas to the AT researcher any other approach seems meaningless and the idea of an operational model perfectly natural and usual. It is not impossible that the influence of AT may in fact be helping to establish this concept in its proper place at the centre of modern linguistics.

The other concept relates to the exchange 'text ↔ meaning'. It is not fortuitous that AT should be making the first attempts to develop Chomsky's idea that generative/recognition grammars should be capable of generating sentences corresponding to a given meaning or of discovering the meaning of given sentences. Although this idea has long been mooted among linguists, it still has to win its rightful place, whereas in AT it derives naturally from the objective (see below).

AT thus appears as an effective catalyst helping to establish important concepts and approaches in linguistics. It is also a kind of experimental 'training ground' where linguistic theories and descriptions can be subjected to thorough detailed verification.

It would now no doubt be expedient to ask what AT has achieved in its fifteen years of existence.

On the empirical side, we can point to the development of binary algorithms for several pairs of languages (English-Russian, Russian-English, English-Japanese, French-Russian, etc.). These algorithms have been programmed and tested on computers in a whole series of experiments. We give below extracts from some machine translations:

1. French-Russian. U.S.S.R.,'Проблемы кибернетики". (Problems of cybernetics), 1962, vol. 8, pp. 286-7:

Nous venons de dire que la constante: arbitraire, si elle est convenablement choisie, entre linéairement dans l'intégrale générale de l'équation de Riccati. En fait, nous avions déjà obtenu ce résultat par une voie élémentaire, quand nous avons démontré	Мы только что сказали, что (чтобы) произвольная постоянная, если он (она) надлежащим образом выбирается, входит линейно в общий интеграл уравнения Riccati. Действительно, мы уже получили этот результат элементарным путем; ко-

que le rapport biharmonique de quatre solutions de l'équation précédente est une constante. On peut remarquer que les intégrales de cette équation ont non seulement leurs points critiques fixes, mais aussi leurs pôles fixes.

гда мы доказали, что (чтобы) отношение biharmonique 4 решений предыдущего уравнения есть постоянная. Мы можем заметить, что (чтобы) интегралы этого уравнения имеют, нет (не) только их фиксированные критические точки, но также их фиксированные полюсы.

2. Russian-English. United Kingdom, November 1964, National Physical Laboratory, personal communication of John McDaniel:

Использование свойство коммутации магнитного потока внутри разветвленного магнитопровода значительно упрощает конструкцию устройств и уменьшает число входящих в них элементов. Однако, несмотря на то, что сложная форма разветвленного магнитопровода позволяет легко получить многофункциональные устройства, заменяющие большое число простых элементов, на практике широкое распространение получили в основном устройства с двух- и трехдырочной формой магнитопровода. Такое положение объясняется значительными трудностями, возникающими при выборе оптимальной конфигурации разветвленного магнитопровода.

Allocation of properties of commutation of magnetic flow inside branched magnetic circuit considerably simplifies structure of devices and decreases number of elements entering into (in) them. However, despite that (then), that (that) multiple form of branched magnetic circuit allows lightly to receive mnogofunccionalic devices, replacing big number of simple elements, in practice wide propagation received in the main device about (from, with) two (other, few)— and trechdyrent by form of magnetic circuit. Such position is explained by considerable difficulties, springing up in (with) selection of optimum configuration of branches magnetic circuit.

3. Russian-English. United States, University of Georgetown, *Computers and Automation,* No. 5, 1963, p. 29 (the original Russian text is not given) :

If compared words do not coincide, as a result this operation receives which anyone a number, but not 0. In this case occurs switching on the following word of a dictionary, and so up to these being time, meanwhile upon a subtraction does not receive 0.... Now necessary to know which corresponds to it in by friend a tongue. Side-by-side with each word of converted tongue is indicated the number of the cell, containing the corresponding combination of this tongue, on which convert. A when subtraction gives as a 'result 0, switching occurs already not on the following word of a dictionary, but on this cell of the second tongue, a number which is side-by-side with the given word.

4. Russian-English. United States, Ramo Wooldridge Inc., *Fulcrum Tech niques to Languages Analysis,* RADC-TDR-TDR-63-168, March 1963, p. 59-60 (the original Russian text is not given):

Moon from immemorial times (periods) attracted the attention of the man. Still (yet) into ancient times (periods), philosophers expressed the correct idea, that moon the independent celestial body, the practically spherical form, can be similar to ground (earth). The new period in the study of our natural satellite began in (into) (NTS) the year, when (in which) Galileo directed on moon ones (its, my, our) first primitive telescope. He (it) discovered on moon of the plain and mountains. From this moment (momentum), it began the creation of new science is (are) the selenography, occupied (engaged in, concerned) by the study of formations (productions at surface of moon).

Of course the quality of these translations leaves much to be desired. However, since they were produced, algorithms have gradually been improved and the standard of quality has obviously risen (unfortunately the author has no recent texts available).

Far more important and more significant for the future are the theoretical achievements of AT, although it is incomparably more difficult to demonstrate them to the reader. We shall select and try to elucidate three groups of problems which, in our view, are of great interest in linguistics.

First of all, let us consider the problem of the representation of syntactic structure. It is certainly AT which has given impetus to contrasting studies such as the constituent method ('parentheses') or the dependency method ('arrows')—cf. the work of D. Hays [1, 2][1] and of E. V. Paducheva [3]. It was in connexion with AT that the advantages and disadvantages of these two methods were first seriously considered (e.g., A. Sestier [4]) and other compromise methods proposed. It is with a particular method of structure representation that we usually associate a given type of formal grammar. Corresponding to the constituent method we have the IC grammars,[2] and to the dependency method the so-called dependency grammars. The concept and theory of IC grammars were evolved long before AT; similarly, the formal study of abstract IC grammars continues outside the field of AT. However, although the idea of dependency grammar derives from traditional linguistics and was developed by L. Tesnière, the detailed study of such grammars is at present carried out in the context of AT. Moreover, it is in this field that formal grammars of a new type have appeared—the 'push-down store' type. The push-down store was originally proposed by Oettinger as a convenient means of analysing real sentences (cf. below). It is impossible to examine in detail the various methods of syntactic structure representation and the various types of grammars, and we shall merely point out that it seems to have been in the field of AT that it was felt necessary to distinguish systematically between the method of representing syntactic structure and the method (strategy) of detecting it in the course of analysis.

The second problem, to which we shall devote a little more time, is the detection of the syntactic (or semantic) structure. Hitherto linguistics has taken little interest in determining exactly how the user of a given language establishes relationships between textual units.[3] The individual does it so easily and spontaneously (this refers of course not to the conscious determination of relationships as, for example, in parsing, but to the uncons-

1. Figures in brackets refer to the bibliography at the end of this article.
2. Immediate-constituent grammars.
3. Interest has centred mainly on how the linguist can discover the syntactic laws of a language he does not know, i.e., modelling the behaviour of the researcher rather than of the language user.

cious use of such relationships in the process of understanding a text) that we have the impression that there is no special problem.

However, when the syntactic structure of a sentence has to be discovered by a machine (which is inevitable in AT), we are forced to inquire exactly how the information available to the machine is utilized.

It is natural to imagine the following process: for any class of words, all syntactic possibilities are recorded in a special table; the algorithm scans the phrase from left to right consulting the table, and gradually establishes the syntactic relationships. In Russian, for example, for the substantive in the genitive case (S gen.), we state that it may act as attribute to a preceding non-pronominal substantive S, as complement of a verb of the type избегать (avoid), as complement of a negative transitive verb, as adverbial complement of time (name of a month accompanied by an ordinal number), etc. On meeting S gen. the algorithm checks whether it is preceded by S; if it is, it relates that S to S gen. (or verifies some other more complex condition); otherwise it looks for a verb of the избегать type, etc. At each step the algorithm takes one decision; that is to say, among the multiple syntactic functions of a word, it must choose only one, after exhausting the information available at this stage.

In some cases, the algorithm makes mistakes, but in subsequent stages of the analysis it must elucidate and eliminate them. At step $n + 1$, the algorithm will have utilized all the information accumulated during the preceding n steps, the aim being at each moment to determine the structure as precisely as possible, before continuing the analysis. Such strategy may be called sequential or progressive or local (in so far as the sentence is not examined in its entirety, but only locally). Most syntactic analysis algorithms developed in the early 1960s were of this nature.

However, a situation is possible (at least in Russian) where an algorithm of this type is unable to make a correct analysis (or at least no one has yet managed to construct a local algorithm capable of handling this situation). Take, for instance, the sentence beginning as follows:

1. Для этого числа... (For this number...) (For this [the] numbers...). The local algorithm will establish the link between preposition and noun (as in для этого числа мы имеем... [for this number we have . . .]), but this may prove incorrect.

2. Для этого числа разбиваются на три группы. (For this [the] numbers are classified in three groups). In order to avoid the mistake we must introduce the restriction 'the preposition is connected with the noun following it:

S gen. sing./nominative-accusative, plur.

only if the clause contains no plural verb requiring a subject'. In these circumstances sentence 2 will be analysed correctly. However, there is the possibility of a sentence such as:

3. Для этого числа находятся все детерминанты. (For this number are obtained all determinants) where our restriction does not work. We must specify further ' . . . and if there is such a verb, it must be possible

to find another S plur. capable of being the subject'. But even that is not enough: in Russian the noun 'number(s)', the corresponding verb, and its potential S plur. subject may be separated from each other by any number of clauses containing both verbs and S plur. and also S plur. of the type 'number(s)', for example:

4. Для этого числа, как указывалось в главе 4, которая, если мы..., ..., находятся с помощью метода, который..., все детерминанты. (For this [the] number(s), as indicated in Chapter 4, which, if we...,..., are obtained by the method which . . . , all determinants).

It does not seem clear how we can describe such a situation (should it be possible) within the limits of a local algorithm. Of course, such complex sentences are practically never encountered in a real text, and it *is* possible to construct a local algorithm to handle only fairly common phrases. This algorithm will, however, be of only limited scientific value. It is incapable of a certain important human faculty, namely, the ability to understand phrases correctly (though perhaps only slowly), whatever their degree of complexity. A syntactic analysis algorithm must be able to analyse sentences as complex as are desired (including those not encountered in the text, but grammatically quite correct); just as a multiplication algorithm will supply the correct product of any numbers (including such large numbers that in fact no one has ever multiplied them). However paradoxical it may appear, an algorithm of this kind, because of its greater generality, may prove not only more powerful but even simpler than an algorithm designed solely for sentences encountered in practice.

It is this very tendency to make algorithms more powerful and at the same time simpler, to get rid of the many restrictions, application conditions, restrictions on application conditions, etc., which leads to another idea—global or general strategy. Such strategy consists in the following procedure: at first we apply to each word in the phrase all hypotheses regarding its possible syntactic relationships—that is, we relate the word hypothetically to all words to which it is in principle capable of being related. We thus obtain a set of hypothetical structures, which are systematically checked with the general rules of correct syntactic structure in the particular language. These rules act as filters which reject incorrect structures and let through only those structures which fulfil all requirements. For this reason it is termed the 'filter method'. Researchers must therefore establish and formulate all general laws governing sentence structure in the particular language. This yields very interesting results for 'normal' syntax. Moreover, the general algorithm is in principle very simple; it is expressed in a series of lists enumerating all conditions required for correct syntactic structure;[1] if all these lists are right, the general algorithm guarantees correct analysis of any sentence however complex, for all possible structures are examined. The filter method proposed in 1960

1. In practice, it is necessary to look for an optimum classification of the hypothetical structure, but this is another matter.

by Y. Lecerf and D. Hays was employed in syntactic analysis algorithms by O. S. Kulagina, S. Ja. Fitialov, G. S. Cejtin and L. N. Iordanskaja. The description of general conditions for correct syntactic structure in Russian has been published [5]; the results of experiments on mechanical syntactic analysis by the filter method [6] are very satisfactory. It has also been shown that the filter method can be applied effectively not only in analysis but also in synthesis; in the case of multiple-meaning synthesis (for a given meaning all possible expressions are constructed), it is convenient to start by generating all conceivable structures at every level—from meaning to 'depth' syntax, and from 'depth' syntax to 'surface' syntax, etc., up to the string of real word-forms—and then to eliminate unsatisfactory structures by means of various constraints, formulated with separate words and rules [7].

Lastly, apart from local sequential algorithms and filter algorithms, there are the so-called multivariant algorithms or multiple-path analysers (A. Oettinger, and then M. Sherry, S. Kuno, W. Plath); the algorithm scans the phrase from left to right, handling each word in turn, just like a local algorithm; however, in the case of a multiple-function word, it does not seek a single solution, but tests all functions, each in turn ('analysis branches into several paths'); after selecting one of the functions enumerated for a given word, the algorithm moves on ('takes the next path in the analysis'), and on encountering another multiple-function word, it again tries each of its functions in turn ('the next path divides into several branches'). Finally, after following a given path, the algorithm either arrives at the end of the sentence (i.e., it detects one of the correct structures) or it reaches an *impasse,* where no other syntactic function can be associated with the word examined (this means that the path followed was wrong, or at least that one of the solutions selected was not right). In both cases the algorithm backtracks to the last branch point where some functions remain untested, and in this way reviews all possible paths. In short, the algorithm provides all correct syntactic structures (syntactically polysémie sentences may have several) and only these. For the construction of algorithms of this type, there is a remarkably elegant and effective technique—'push-down store'—and a convenient form for describing syntactic possibilities—'prediction pool'—(information regarding the structures which should in principle follow the given word); the analysis takes the form of plausible hypotheses [cf. 8, 9, 10,11]. At present push-down store grammars are being studied in the abstract, as one of the possible procedures for modelling linguistic behaviour.

The problem of finding the optimal strategy for machine analysis corresponding most closely to real recognition strategy in human beings cannot yet be regarded as solved. However, a clear definition of this problem and the detailed elaboration of various methods of textual analysis indubitably provide a valuable contribution to linguistics.

The third sector in which AT has played a leading role is the famous problem of meaning or sense. Although in theory this problem has always been acknowledged to be of exceptional importance, in fact little attention has been paid to it. The key ideas of structural linguistics on the advisability of detecting elementary sense units, i.e., the description of discrete units of language content (L. Hjelmslev's content figures, Z. Harris's discourse analysis, etc.) were seriously developed in connexion with AT research and automatic information retrieval; we refer to the so-called semantic factor systems [cf. 12], which were developed most fully in the AT laboratory work of the first Pedagogic Institute of Foreign Languages in Moscow [13]. In exactly the same way, it was for AT purposes that fruitful research was started on the description of semantics with special dictionaries called thesauri (cf. the work of the Cambridge AT group, for instance [14]). Contrary to the mistaken view that AT takes account only of the external characteristics of a text and completely disregards the meaning, sooner or later researchers find that meaning occupies the forefront of their attention; moreover, any translation is before all else the transmission of meaning, i.e., a transformation retaining the sense. That is why AT arrives quite naturally at the problem of constructing notation, systems (i.e., elaborating means of noting the content of synonymous statements in an identical manner) and 'text ↔ meaning' algorithms. To our knowledge, the most complete system at present in process of development (the system of the Linguistics Research Center of the University of Texas [15]) covers semantic analysis (reduction of synonymous phrases to a single type) and semantic synthesis (construction of all acceptable expressions with a given meaning). In the U.S.S.R., research on semantic synthesis [7] has also been undertaken. In this connexion we feel it expedient to draw attention to two interesting linguistic results.

To describe lexical compatibility in any natural language we introduce 'semantic parameters'—very general elements of meaning, each of which has numerous expressions, the choice of the appropriate expression depending entirely on a given key word. For example, the semantic parameter *Magn* (high degree) gives: fierce, bloody (battle); thunderous, prolonged (applause); sworn (enemy); bitter, piercing (cold); driving, torrential (rain); great, grave (prejudice); thick (fog); (know) perfectly, inside out, thoroughly; etc. The parameter *Incep* (beginning): (the wind) rose; (rain, snow) began to fall; (the fire) caught; (the exhibition) was. opened; (the debate) began; (an abscess) forms; to start (a conversation); to strike up (an acquaintance); (revolt) breaks out; to begin to (speak, sing, etc.). The parameter *Opem* (do (what should be done with the given object)): to exert (pressure); to render (assistance); to bring (help); to make (a phone call, a translation); to have (confidence); to carry out (an inquiry); to sit on (a chair); to take (a step); to provoke (an explosion); to sing (a song); to commit (a crime); to indulge in (debauchery); etc. With two or three dozen general parameters (applicable to all words) and a very few specific parameters (significant for given semantic groups only) it is possible to describe most non-free associations[1] in any language; in a

1. Excluding 'true' idioms expressing a global meaning, such as: 'to twiddle one's' thumbs, 'stick-in-the-mud', etc.

dictionary, the meanings of all parameters of each word should be shown.[1] Parameters may be compared to cases: just as in Russian it is possible with six cases to represent the whole variety of case morphemes (for each root we indicate which ending should be affixed for a given case), with parameters we can represent a much larger variety of adjectives, verbs, etc., matching a given word (for each word we indicate the word to be associated with it in order to express the parameter in question, i.e., the desired meaning).

One of the practical uses of parameters is in translation (not only in AT but also in multilingual dictionaries for human use). For instance, in English, 'dramatic success' *(Magn* + success) gives in Russian *Magn* + (успех) большой, необычайный, шумный, сногсшибательный; and in French: *(un succès) énorme, considérable, extraordinaire, bœuf,* etc.

In terms of semantic parameters which (owing to their semantic character) are universal categories, we may formulate semantic equivalences which are also of universal value. With these equivalences (about thirty are known to date) we can describe in general form cases of synonymity such as: John helps Peter ↔ John is Peter's assistant; England gave them support ↔ They were supported by England; John likes to read in the afternoon ↔ John willingly reads in the afternoon; Set A contains element x ↔ Element x belongs to set A; etc. Semantic equivalence forms the basis of this system of synonymous periphrasis, which permits the construction of all acceptable expressions of a given meaning.

We have presented here as examples only some of the most outstanding theoretical results obtained in AT. For reasons of space we are obliged to stop there. We feel, however, that the examples quoted are sufficient to enable the reader to grasp the author's argument. This article will have achieved its purpose if it has given even an approximate idea of the aims of linguistic research in AT and of how such research *is* related to modern theoretical and descriptive linguistics.

The author wishes to express his profound gratitude for the helpful comments and advice of all those who were kind enough to read this paper, and in particular to F. P. Filin, L. N. Iordanskaja, L. L. Kasatkin, O. S. Kulagina, A. A. Reformatskij, E. M. Wolf and A. K. Zholkovski.

BIBLIOGRAPHY

1. HAYS, D. Grouping and dependency theories. In: *Proceedings of the National Symposium on Machine Translation. . . .* Englewood Cliffs, 1961, p. 258-66.
2. HAYS, D. *Dependency theory: a formalism and some observations.* Santa Monica, Calif., 1964, 39 p. (Memorandum RM-4087-PR.)
3. PADUCHEVA, E. V. [On ways of representing the syntactic structure of a clause.] Voprosy Jazykosnanija (Linguistic questions), no. 2, 1964, p. 99-113 (In Russian.)

1. A semantic parameter may be compared to a function (in the mathematical sense): $f(x) = y$, where f is a given parameter (e.g., *Magn), x* a key word (e.g., 'battle'), and y the value of the paramater (in this case fierce, bloody).

4. SESTIER, A. *A system of grammatical description of the natural languages using the concept of grammatical function.* 39 p.
5. IORDANSKAJA, L. N. [Properties of correct syntactic structure and an algorithm for detecting it (based on Russian).] *Problemy Kibernetiki* [Problems of cyber netics], vol. 11, 1964, p. 215-244. (In Russian.)
6. VAKULOVSKAJA, G. A.; KULAGINA O. S. An algorithm for textual analysis. *Problemy Kibernetiki* [Problems of cybernetics], vol. 18. (In press.)
7. ZHOLKOVSKI, A. K.; MELCHUK, I. A. [Possible method and instruments of semantic synthesis.] Naucno-techniceskaja Informacija [Scientific and tech nical information], no. 6, 1965, p. 23-8. (In Russian.) (A fuller version of this paper is to be published in *Problemy Kibernetiki,* vol. 19.
8. OETTINGER, A. Automatic syntactic analysis and the push-down store. In: *Proceedings of symposia in applied mathematics,* vol. 12: *Structure of language and its mathematical aspects.* Providence, 1961, p. 104-29.
9. KUNO, S.; OETTINGER, A. Multiple-path syntactic analyser. In: *Information processing,* 1962 *(Proceedings of International Federation for Information Processing Congress,* 1962. Amsterdam, 1963, p. 306-12.
10. KUNO, S. The multiple-path analyser for English. *Mathematical linguistics and automatic translation.* Cambridge, Mass., Harvard University, 1963. (Compu tation laboratory report NSF-11.)
11. PLATH, W. Multiple-path syntactic analysis of Russian. *Mathematical linguistics and automatic translation.* Cambridge, Mass., Harvard University, 1963. (Compu tation laboratory report NSF-12.)
12. PERRY, J. W.; KENT, A.; BERRY, M. M. *Machine literature searching.* New York, London, 1956, 102 p.
13. Masinnyi perevod y prikladnaja lingvistika [Automatic translation and applied linguistics]. Moscow, 1964, vol. 8, 252 p. (In Russian.) Work of the first Pedagogic Institute of Foreign Languages, Moscow.
14. MASTERMAN, M. *Semantic message detection for machine translation using an interlingua.* Teddington, 1961, 36 p. (National Physical Laboratory, paper 36).
15. *Report No*. 16. *Final report.* Austin, Tex., Linguistics Research Center of the University of Texas, 1963, 85 p.

Igor Melchuk is senior scientific officer at the Institute of Linguistics of the Soviet Academy of Sciences where he is concerned mainly with problems relating to automatic processing of linguistic data and general linguistics. Since 1954, he has conducted research in the field of automatic translation and his numerous publications include a work on mechanized syntax analysis (Moscow, 1964), as well as a number of articles on automatic translation, the Romance and Semitic languages, Hungarian, etc.