mentioned incidentally. In the latter category is for instance the actual nature of the speech disturbances. This is precisely the area that would particularly interest linguists, and where their skills in language analysis might produce important clinical and theoretical discoveries. But the book makes only a few superficial statements about the phonological derangements of aphasic or electrically stimulated patients, and there is no mention whatever that morphology or syntax either were or were not affected in the utterances of their patients.

Despite these critical remarks we must bear in mind that the topic of brain and speech mechanisms is of great complexity, and that at this time we cannot expect any book to cover the subject exhaustively or satisfactorily. Most of the essential spadework either on the cellular or on the intercell level has not yet been done. Speech can only be studied directly on human subjects, and so it is doubtful that the most pertinent data are likely to be fully assembled in the near future. Our only hope is that many more scientists like Penfield and Roberts make public their invaluable observations in connection with their privileged interference with human brain function. Thus we shall eventually compile individual research components that can be assembled into one gigantic experiment by collation. In this sense we owe Penfield and Roberts a great debt. Their book will be found stimulating in private perusal and useful as collateral reading for students in linguistics and psychology.

Materialy po mašinnomu perevodu, Sbornik I. [Edited by N. D. Andreev.] (Leningradskiĭ ordena Lenina gosudarstvennyĭ universitet imeni A. A. Ždanova.) Pp. 228. Leningrad: Izdatel'stvo Leningradskogo Universiteta, 1958.

Reviewed by Morris Halle, *Massachusetts Institute of Technology**

The book under review is a collection of papers on various aspects of machine translation (hereinafter MT) by workers of the Experimental Laboratory for MT recently established at Leningrad University. Although at the time of publication the group had been active for something less than two years, it had already succeeded in starting work on a wide range of problems. One cannot fail to be impressed with the great variety of languages which are being investigated (Arabic, Burmese, English, Hindi, Indonesian, Japanese, Norwegian, Russian, and Vietnamese), by the great industry of the workers, and by the purposeful direction of the project. Since the book is basically a progress report in a rapidly developing field, it seems to me that a critical examination of detailed proposals made in the book would be of little general interest, for in all probability new procedures have long since superseded many or most of those described. I have therefore chosen to focus the present review on a more general issue, that of the mutual relevance of linguistics and MT, which is raised both explicitly and implicitly by this book.

The suggestion that it is possible and desirable to program an electronic

* This work was supported in part by the U. S. Army (Signal Corps), the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U. S. Navy (Office of Naval Research); and in part by the National Science Foundation.

computer to perform translations from one language into another has captured the imagination of many workers, professionals as well as laymen. Newspapers have given favorable publicity to even the most modest achievement in this area, and organizations supporting scientific research have been very generous with their assistance.[1] As a result, a considerable number of energetic workers have been drawn into the field, and a great deal of activity has developed there in a comparatively short time. If the success achieved so far has been less than sensational—newspaper stories implying the contrary notwithstanding—this is not due to any lack of earnest effort or material support, but rather to the very great difficulty of the problem, which for a nontrivial solution requires ideas of the highest order of originality and sophistication.[2]

MT is fundamentally an engineering problem. Since important laws of natural science have often been discovered in the solution of engineering problems, linguists are well advised to follow developments in MT. Yet progress in MT does not necessarily imply progress in linguistics. From the fact that an engineering design works one cannot conclude that nature works in the same way. A functioning device cannot even be taken as conclusive evidence that its designer had a sound understanding of the nature of the phenomena he was manipulating, for lack of such understanding has not always been a hindrance to the construction of highly practical and successful devices.

In view of this I cannot share the opinion of Steblin-Kamenskij that 'the framing of absolutely clear and consistent rules, which are required by a translating machine, is at the same time also a critique of all concepts of traditional grammar' (3). The weakness of Steblin-Kamenskij's position is clearly brought out by the very examples he cites in support of it. I do not see that linguists can learn anything useful from the fact that 'in the course of working out an MT algorithm, it was found convenient to proceed from assumptions according to which the Russian word *porosja* ['suckling pig'] would be considered a reflexive verb with the meaning of a noun' (6). The form *porosja* can be interpreted as a reflexive verb only if the fact that a form ends in *-sja* is taken as a sufficient reason for assigning it to that category. A slightly more elaborate approach would no doubt also take into account the fact that Russian verbal forms cannot end in /ro/, and that there is, therefore, a very simple formal reason for rejecting the above analysis. Nevertheless, that crude analysis may well be ade-

[1] In a recent report prepared for the Office of Naval Research, Y. Bar-Hillel estimates that in the United States in 1958 about 150 workers were engaged in MT research with a total budget of $1,500,000. He believes that the MT effort in the USSR is of the same magnitude.

[2] In view of this it seems unlikely that an MT scheme significantly superior to word-by-word translation will be put into operation in the near future. For the present, the most realistic solution of the practical problem of disseminating information available only in a foreign language is to teach the language to the final consumers—the scientists, engineers, and government officials who must read the material in question. Most of these need to know the language only well enough to scan writings in a restricted field—e.g. nuclear physics, automation, or genetics—so as to decide whether the material warrants further study. From my experience with language courses for graduate students preparing for their foreign-language reading examination, I am sure that in most languages the needed facility can be acquired in six to twelve weeks.

quate for purposes of MT even if it is patently inadequate for purposes of a linguistic description.

The need for a clear distinction between solutions that work for MT and solutions that are satisfactory from a linguistic point of view is further illustrated by the problem of mechanizing morphological analysis, which incidentally is the one on which most effort has been expended by the Leningrad group. We shall consider three different solutions of this problem, all of which are noncontradictory and complete in the sense that they yield the correct analysis of every form of the language.

The first solution makes use of a simple list of all forms in the language. Since a lexical item can have only a finite number of inflected forms—e.g. a Russian noun such as *syn* 'son' has at most ten orthographically different flectional forms—it is possible to list these, together with all relevant grammatical information. The MT procedure for morphological analysis is then nothing more than a simple look-up procedure that can be programmed in a straightforward manner on most existing computers. For example, given the word form *syna*, the computer would be instructed to print out the information that the input form is either the gen. sg. or the acc. sg. of the noun *syn*; given the word-form *mesti*, the computer would be instructed to print out the information that the form is either the infinitive of the verb *mesti* 'to sweep' or the gen., dat., or loc. sg. or the nom. or acc. pl. of the noun *mest'* 'vengeance'.

Although conceptually the simplest solution of the problem, this brute-force approach is not the one favored by most MT workers, for in order to store in the computer a list containing all forms of a language it would be necessary to utilize tape memory units, which are relatively impractical.[3] It is therefore imperative to reduce the size of the list. If the language makes extensive use of the process of suffixation it is natural to propose—as is done by the Leningrad group[4]—that each word form be analyzed into a suffix and a stem. The list can then be reduced to contain only stems and suffixes. In Russian, as in most suffixing languages, the process of suffixation is moreover accompanied by predictable morphophonemic changes. If the mechanical analysis routine is limited exclusively to splitting off suffixes, it is necessary to postulate numerous pseudo-stems as well as pseudo-suffixes. For instance, the cited form *mesti*, if taken as an infinitive, is the result of the application of general morphological rules of Russian to the sequence {met + t,í};[5] the shift from {t} to {s} before the infinitive suffix is an automatic consequence of the operation of these rules. An analysis routine in which all grammatical operations other than suffixation are

[3] Since it is probable that computers with the requisite memory capacity will soon become available, some MT workers (among them V. Yngve of MIT) have decided to by-pass morphological analysis. This is evidently sensible from an MT viewpoint, for in the face of the many conceptually unsolved MT problems, the search for a more elegant solution of a solved problem may well be a superfluous luxury.

[4] The solution for Russian contained in the joint paper by L. N. Zasorin, N. B. Karačan, S. N. Medvedeva, and G. S. Cejtin (136–90) resembles in its conception the solutions proposed by some American workers; see A. G. Oettinger *A study for the design of an automatic dictionary* (Harvard diss. 1954).

[5] R. Jakobson, The Russian conjugation, *Word, 4*, 155–169 (1948).

neglected must list the pseudo-stem *me-* and the pseudo-suffix *-sti*, in addition to the true stem {met} and true suffix {t,i}.

A solution like the one just discussed, which is limited to splitting off suffixes from stems, cannot work in languages in which morphological information is conveyed by such devices as reduplication, vowel apophony (ablaut), and consonant mutations, or by discontinuous morphemes.[6] The method of listing all word forms is of course possible in these cases too, but again not practical at present. In the analysis of such languages, therefore, a procedure rather different from the two outlined above is needed. This procedure, which elsewhere we have called 'analysis by synthesis',[7] can best be explained with the help of an illustrative example.

Assume that for some task it is necessary to distinguish integers that are seventh powers from all other integers. It is impossible here to have recourse to a list, since there is an infinity of such integers. But it is evidently possible to determine in a finite number of steps, by simple multiplication, whether any given integer is a seventh power: one can begin by calculating the products $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$; $3 \times 3 \times 3$ ..., etc., until the calculated product equals or exceeds the integer in question. This procedure is by no means wholly impractical, for it can be materially shortened by various shortcuts. For instance, depending on whether the given integer is odd or even, only odd or only even roots need to be investigated; if the given integer has more than 14 digits, no roots below 100 need to be tried; and so on. But even in the absence of short cuts the suggested procedure will always work, since any integer can be tested by a finite number of multiplications.

In applying this method to morphological analysis one would store in the computer memory all true stems and the morphological rules for operating on stems, i.e. the flectional morphology of the language. The stored information is all that is needed to generate the list of all forms of the language. To analyze a particular form appearing in a text, we simply find the form in the generated list. The procedure whereby a particular list form was generated is then the desired analysis. If there is more than one way of generating a particular form—e.g. Russian *mesti*—there will be more than one occurrence of that form in the generated list. Such multiple occurrence indicates that a form is morphologically ambiguous.

The proposed solution is completely general, in the sense that it is applicable to any describable morphological process. It also happens to be economical with regard to its utilization of the computer memory. Its obvious shortcoming is its very cumbersome operation, since in order to analyze even a single word it is necessary to generate all forms of a language. But this is not unavoidable;

[6] The efforts of the Leningrad group to handle languages with such morphological processes have been limited in the main to an attempt, of peripheral importance in these languages, at splitting off suffixes and prefixes; see the discussion of MT routines for Arabic (112–26) and for Indonesian (88–97). This may well be due to the group's very understandable desire to avoid difficult problems at a very early stage of their work.

[7] See M. Halle and K. N. Stevens, Analysis by synthesis; W. Wathen-Dunn and L. E. Wood (edd.), *Proceedings of the seminar on speech compression and processing* (Air Force Cambridge Research Center Technical Report 59-198; Bedford, Mass., Dec. 1959).

it is not difficult to think of various shortcuts which would make the suggested procedure comparable in speed to other approaches.[8]

It is evident that the first method, that of listing all possible forms of a language, has no serious linguistic interest, since it is inconceivable that man performs morphological analysis without taking any advantage whatever of the regularities of the language. For essentially similar reasons it is doubtful whether the second method, that of splitting off affixes, is of major linguistic interest. This, too, fails to exploit fully the structural regularities of the forms to be analyzed. The second method, moreover, throws no light on how morphological analysis might be performed in a language with morphological processes other than affixation. In addition, both methods postulate an analysis that is basically different from the synthesis. Only the third method, that of analysis by synthesis, does not suffer from these shortcomings. It is therefore of considerably greater linguistic interest than the others. This does not mean, however, that the third method should necessarily be chosen for MT. A host of nonlinguistic factors, such as the state of computer technology and the need for speed, must be conconsidered before deciding which of the three methods is best for MT.

The method of analysis by means of synthesis may also be applicable to syntactic analysis.[9] To demonstrate this in the requisite detail is a difficult task which needs considerable further work. Such a demonstration would be of great importance for linguistics, since it would provide the basis for a change in our view of the relationship between speech production and speech perception. At present, it is usual to consider perception and production as two basically different processes, each with its own rules. Some linguists have accordingly spoken of two distinct grammars, one for the speaker and one for the listener. If it can be shown that syntactic analysis, like morphological analysis, can be performed by a method that involves in an essential manner the generation of the nessage by the listener, there is no need to postulate the existence of two separate grammars. In addition to its obvious interest for linguistics, the demonstration that analysis by means of synthesis may be applicable to language on all levels promises to yield important insights for psychologists and physiologists interested in perception. And since listing of sentence types is not a realistic machine procedure for

[8] G. H. Matthews and I have written a computer program for the morphological analysis of Russian by the method just described and hope to put it into operation in the near future. The shortcut we have adopted makes use of a preliminary analysis into pseudo-stem and ending. The pseudo-stem is used to select from the list of true stems stored in the computer memory those relatively few stems that cannot be excluded by some simple rules. As a result only a very small number of forms must be generated by the synthesis program and checked against the form to be analysed. It was found useful to write a simple computer program for converting the input forms, which are given in conventional orthgraphy, into a morphophonemic transcription, and to perform the rest of the analysis and synthesis in morphophonemic terms.

[9] Judging from the work reported in the book, the Leningrad group has yet to come to grips seriously with the mechanization of syntactic analysis. In a paper on the analysis of simple English sentences, B. M. Lejkina outlines a few rudimentary rules—e.g. '1. The first noun not preceded by a preposition is the subject of the sentence; 2. The first verb following the subject is the predicate verb' (216)—which evidently represent the very first attempts and have by now no doubt been superseded by much more adequate and sophisticated routines.

syntactic analysis, even if the memory capacity of computers should be multiplied manifold, the method of analysis by synthesis may prove also of practical value in MT.[10]

Kindersprachforschung mit Hilfe des Kindes: Einige Erscheinungen der kindlichen Spracherwerbung erläutert im Lichte des vom Kinde gezeigten Interesses für Sprachliches. Von W. Kaper. Pp. xxiii, 244. Groningen: J. B. Wolters, 1959.

Reviewed by W. F. Leopold, *Northwestern University*

The book appeared earlier in 1959 as an Amsterdam dissertation under the present subtitle as the only title. The author or the publisher obviously realized that such a long, unwieldy title would not do and reissued the book under the new main title, with a new cover and title page, omitting the reference to the dissertation character of the publication from the title page and replacing the Dutch preface by a brief note in German. The new title is handier but not skillfully chosen. The book deals with what I have called 'linguistic consciousness'. My choice for a title would have been 'Sprachliche Bewußtheit des Kindes'. For this specialized topic of the language learning process, this study is surely the most thorough, most scholarly treatment in existence.

Willem Kaper is a teacher of German in Rotterdam, who has now been called to the University of Amsterdam, where he will be in charge of the training of teachers of German. He kept a careful record of the learning of Dutch by his two sons 1;7–8;2. The present study deals with a small portion of his material, extracted for the purposes of the sharply restricted topic. Kaper finds it impossible to exclude psychological considerations from the treatment, but keeps in mind conscientiously that he is not a psychologist himself. His approach is definitely linguistic, as learned in Anton Reichling's lectures on child linguistics at the University of Amsterdam (incidentally the only linguistic university course on child language of which I have ever heard).

In 37 chapters Kaper deals with a great variety of problems, all somehow related to the central theme of linguistic consciousness. The restricted topic makes it possible to describe the situations in which the utterances were made in great detail. All Dutch words and sentences are also given in German translation. Lexical items play a leading part, but the investigation is not limited to them. In addition to using his own collections made in personal observation, Kaper constantly coordinates his findings with the records of previous investigators in detailed reports and discussions. He makes good use of the Western international literature, with special attention to the child language records of linguists. He is fully cognizant of Jakobson's important theoretical work. Stern's *Kindersprache* is also amply used as a source for parallels, but Kaper is often critical of Stern's interpretations. Dutch studies are more fully utilized than in other books and articles.

[10] An English translation of the volume has recently been published by the U. S. Joint Publications Research Services under the title *Soviet developments in information processing and machine translation* (JPRS–2150–N; New York, 1 Feb. 1960).