MECHANICAL TRANSLATION

DEVOTED TO THE TRANSLATION OF LANGUAGES WITH THE AID Of MACHINES

VOLUME THREE, NUMBER THREE DECEMBER, NINETEEN FIFTY SIX

News

RUSSIA

The Russian journal Voprosy Yazykoznania (Problems in Linguistics) has started a series of articles devoted to mechanical translation, the first three of which appeared in the September - October issue of 1956. They are abstracted in the bibliography section of this issue of MT.

The editors of MT take pleasure in welcoming this outstanding linguistics journal to the field of mechanical translation.

GEORGETOWN UNIVERSITY

The Eighth Annual Round Table Meeting on Linguistics and Language Studies, April 12 and 13, 1957, at the Institute of Languages and Linguistics of Georgetown University, Washington, D.C., had as its general topic: Aspects of Research in Machine Translation. About a dozen papers were presented on mechanical translation and related topics. The proceedings, including discussion are being published by Georgetown University.

The mechanical translation project at Georgetown has issued a number of mimeographed seminar work papers which are available on request.

ENGLAND

The Cambridge Language Research Unit announces receipt of a grant from the National Science Foundation, Washington, D.C., U.S.A., to finance its work for one year. Margaret Masterman has been appointed principal investigator. The unit will concentrate on investigating logico-mathematical methods for the analysis of languages for mechanical translation. The unit is now publishing copyright workpapers, a limited number of which will be available for distribution.

ITALY

We hear from Silvio Ceccato: "During the last months I have perfected

a) my studies on mechanical translation, and
b) my working hypotheses for the (anatomical and physiological) individuation of the organs of the mental and perceptive activity.

As to (a) I think that by now only the analysis of the particularities of the single languages in question is lacking for an industrial utilization of the principles. Also the steps in order to formulate the operations into a program for a calculating machine are clearly sketched."

UNIVERSITY OF WASHINGTON

Erwin Reifler reports that an initial project from June 1956 to March 1957 analyzed one hundred eleven Russian texts from thirty-one fields of science and supplied approximately 14,000 Russian-English entries consisting of Russian "semantic units" belonging to the technical and general language vocabulary occurring in these texts, additional lexical units selected from high frequency lists, and their target equivalents. An expanded project, to be completed in October, 1957, has as its object the preparation of a new list of lexical units consisting of all of the paradigmatic forms of the technical terms from one of the thirty-one fields and of the general language vocabulary found in the Russian scientific texts together with their target equivalents. This will increase the number of entries to about 160,000.

The first translation machine will not have the logical equipment necessary for the automatic resolution of grammatical and non-grammatical problems. Consequently Russian word order will be retained in its output which will, moreover, still be cluttered up with "strings" of target alternatives. There are, however, many opportunities for output improvements by changes in the original lexicography which take advantage of the very large storage capacity of the memory device under construction.

PASADENA, CALIFORNIA

Peter Toma, who recently gave a demonstration of an initial computer program for mechanical translation at the California Institute of Technology, has sent the following brief summary of his procedure. A detailed description is being prepared.

"The translation is carried through sentence by sentence. First, about one paragraph of the text is block transferred from magnetic tape to a special location of the high speed working memory. A short program finds the end of the first sentence and transfers it to a particular location where the translation actually takes place. After a check of the word endings and a computing process (arithmetical operations which actually do context analysis) the computer knows approximately where to find in its large, systematically arranged but relatively slow access memory all the material pertinent to the sentence. This material is grouped around "base words" which are mostly nouns and verbs. In the large memory these nouns and verbs are listed as "heads" of semantic units. These units contain usually two or three words and are selected according to frequency occurrence in the field to be translated. About one third of the large memory contains a dictionary which remains unchanged in every field while two thirds has to be changed from magnetic tapes according to the subject field under translation.

"Whenever a "pertinent dictionary" has been transferred to the high speed working memory the translation of the sentence actually begins. The computer again looks up all the words from the beginning of the sentence, combines them into the longest possible logical sequence and, by making a constant search in the pertinent dictionary, finds largest semantic units with their English equivalent for that particular sentence. For example: A, B, C, D, E, ... K are words in a sentence. In the particular dictionary for this sentence (which has already been block transferred) there is a combination of AB and DE and the meaning of the word A and C. As the computer starts out, it combines A and B, it checks the dictionary, finds the equivalent and writes it out. Then it checks the next word which is C, combines it with D, finds nothing in the dictionary, takes CDE and finds nothing, then CDEF and again finds nothing. After the fifth combination (that is the longest unit which will be considered) the computer automatically goes back and treats C separately, finds an equivalent, writes it out and then looks up DE. If this is not available it continues the same procedure again and after the fifth unsuccessful combination goes back to D. Even if it does not find D, the computer types out the word in its original untranslated form and continues with the combination EF, etc.

"My demonstration program has been an experimental presentation of some phases of the above mentioned procedure. I have been limited because I could not use the Datafile storage unit, and the capacity of my working memory has been only 4,000 words. However, considering the frequency occurrence of word pairs or even three word units, especially in scientific texts, I believe that soon we shall have large enough memories to store all those units."

Some Psychological Methods for Evaluating the Quality of Translations †

George A. Miller and J. G. Beebe-Center, Harvard University, Cambridge, Massachusetts

The excellence of a translation should be measured by the extent to which it preserves the exact meaning of the original. But so long as we have no accepted definition of meaning, much less of exact meaning, it is difficult to use such a measure. As a practical alternative, therefore, we must search for more modest, yet better defined, procedures. The present article attempts to survey some of the possible methods: One can ask the opinion of several competent judges. Or, given a translation of granted excellence, one can compare test translations with this criterion by a variety of statistical indices. Or a person who has read only the translation may be required to answer questions based on the original. The characteristic advantages and disadvantages of each method are illustrated by examples.

ONE HEARS it said that MT is currently rather crude, but that workers in the field are striving to improve and refine their translations. A brief encounter with the unedited output of an automatic dictionary is sufficient evidence of the tremendous range of quality between the simplest mechanical 'translation' and the product of a skilled, human translator. The question is whether this intuitive judgment of the quality of a translation can be made more precise by any psychological techniques of scale construction.

A scale of the quality of translations should be reliable, valid, objective and easy to use. In addition to these general desiderata for all scaling procedures, there are certain special features that this particular scale should have. For example, it should be applicable to any translation, whether produced by a machine or by a human translator. This feature would enable us to compare the output of a particular machine to the output of a human who had had a known number of years of study in the foreign language. Furthermore, the scale should be applicable to translations from or into any language whatsoever, and so should not take advantage of any characteristics peculiar to a given language, say English — Whether or not a single scale can apply to all languages and still make linguistic sense is a debatable question. And, preferably, the scale should be unidimensional, so that different translations could be compared with respect to a single 'figure of merit'. Finally, we would like to have one or more cutoff points indicated along the scale; "completely unusable," "useful for scanning as to subject matter", "useful after post-editing", "immediately readable, " and "suitable for publication" are some criteria that we might hope to locate along the scale.

All these features would be desirable, but it is not obvious at present that they can be achieved.

Subjective Scaling

Perhaps the most direct approach is to give both the original passage and the translation to be tested to a person who understands both languages and to ask him to assign a number between 0 and 100 to the translation, where 0 means that it is equivalent to no translation at all and 100 means the best imaginable translation. This method fails the criterion of objectivity, of course, and cannot be applied when a polyglot is not available to judge, but we expected to be able to map out the general territory in this way and to use subjective ratings

† Preparation of this paper was supported under Contract AF 33 (038) — 14343 between the U.S. Air Force and Harvard University and appears as Report Number AFCRC —TN—56 — 61, ASTIA Document Number AD 98823. Reproduction for any purpose of the U.S. Government is permitted.

We would like to acknowledge the assistance of Peter Aldin, Martha Taylor, Soon Duk Koh and Elizabeth Friedman.

as a criterion against which to test various other scaling techniques.

In a short exploratory study, however, we obtained somewhat confusing results. We found much disagreement among different raters. Perhaps we should have used foreign language teachers as our judges, for they probably have skill in grading that ordinary, bilingual persons do not seem to have, but we did not anticipate that the ratings would be so difficult.

For the purposes of this study, we selected four summaries of articles from the journal Acustica, two in German and two in French. The journal also gave an English translation, so we had the work of a theoretically competent translator to use for comparison. (The published translations were not the best possible, but they represent the sort of thing that is available in the current scientific literature.) Then we prepared mechanical translations, simulating by hand the possible operation of an automatic dictionary. Each word of the original text was written on a card. These cards were then alphabetized, and on the reverse side we listed the possible English equivalents in approximately the order of their frequency of occurrence, as well as we could judge it on intuitive grounds. From this pack we then constructed six different translations: (1) the first English alternative was chosen from each card; (2) an editor selected the best of the first two alternatives from each card, making his selection in complete ignorance of the other alternatives or the original passage; (3) an editor selected the best one from all the alternatives on each card, still in complete ignorance of the original passage; (4) an editor rewrote the English passage from a knowledge of only the first alternative on each card; (5) an editor rewrote the English passage from a knowledge of only the first two alternatives on each card; and (6) an editor rewrote the English passage from a knowledge of all the alternatives on each card, but without seeing the original passage. In all cases, these editors were monolingual Americans with no linguistic training. The first three procedures did not lead to grammatical English, of course, so we obtained a fairly wide range of quality by these procedures. These six translations, together with the translation taken from the journal and the original passage, were presented to judges who rated them on a scale from 0 to 100.

As a sample of the sort of materials produced, consider a single sentence taken from a French passage:

Original. Il résulte de ceci qu'une atmosphère stratifiée doit toujours réfléchir et donc produire des échos.

(1) He result of this which a atmosphere stratified must always to think and therefore to produce of the echoes.

(2) It results from this which a atmosphere stratified must always to reflect and therefore to produce of the echoes.

(3) It results from this that a atmosphere stratified must always reflect and therefore produce echoes.

(4) The result of this is that in a stratified atmosphere, one must always think of the echoes that are produced.

(5) It results from this that a stratified atmosphere must always reflect and therefore produce echoes.

(6) It results from this that a stratified atmosphere always reflects and therefore always produces echoes.

Published translation. It follows from this that a stratified atmosphere should reflect sound and produce echoes under all circumstances.

A similar sample taken from one of the German passages is the following:

Original. Bei beliebiger Impulsform ergibt sich das Faltungsprodukt aus Membran- und Impulsform.

(1) By any form of the impulse yields -self the products of the folding out membrane- and form of the impulse.

(2) By any form of the impulse yields the products of the folding out membrane- and form of an impulse.

(3) By any form of the impulse yields the products of the folding out membrane- and form of an impulse.

(4) Any form of the impulse is yielded by the interaction of the bending out of the membrane and the form of the impulse.

(5) The impulse in any form yields the products of the folding-out membrane and the form of an impulse.

(6) Any form of the impulse yields the products of the membrane-folding.

Published translation. With a given impulse form one obtains a resultant effect of the shapes of the impulse and of the disk.

Table I

Mean Ratings of Quality of Seven Translations

Method of Translation	French I	French. II	French Mean	German I	German II	German Mean
(1)	21.9	28.2	25.1	27.1	22.2	24.7
(2)	35.5	30.1	32.8	21.6	37.0	29.3
(3)	47.3	27.7	37.5	13.3	29.0	21.2
(4)	38.2	70.1	54.2	45.6	31.8	38.7
(5)	90.5	80.4	85.5	24.0	34.0	29.0
(6)	75.9	54.3	65.1	45.5	77.5	61.5
Published Translation	89.5	80.1	84.8	77.0	75.5	76.3

When the seven translations were given to subjects to judge, of course, no information was supplied as to the method of translation. It is interesting to note that supplying several alternative English equivalents seems to be more useful in translating from French than from German, but this judgment is based upon only these four samples of about 75 words each.

Eleven judges were used for the French passages and ten for the German. The judges were able to speak the language from which the translations came, but had no linguistic training; they were instructed to compare each translation with the original and to take time enough to be sure of their judgments. The means of their ratings are summarized in Table I.

There was so much disagreement among the judges (which was reflected in their bitter comments about the difficulty of their task) that even the means reveal only very general trends. These trends are clearer if we pool the data further, as in Table II.

From Table II we see that far more success is possible with French than with German, and that selective editing helps a little but not so much as complete rewriting. These conclusions are intuitively correct, and it would be disappointing indeed if they failed to appear. The error variance is so large, however, that these conclusions are barely significant.

We were slightly surprised that rewriting made as much difference as it did, since the people who rewrote had essentially the same information about the original passage as was contained in the selectively edited translations. The superiority of the rewritten translations indicated that the judges relied rather heavily upon the grammaticalness of the translation in reaching their decisions. In order to check this notion, we asked another group of subjects to act as judges, giving them the same instructions as before except that they were not shown the original French or German passages. Their ratings correlated closely with the original ratings, especially for the translations from German. It seems, therefore, that people will not regard favorably an ungrammatical translation even though they are able to understand it correctly.

Table II

Mean Ratings for Three MT Procedures for French and German

Method	French	German
No editing (1)	25.1	24.7
Selective editing (2-3)	35.2	25.3
Rewriting (4-6)	68.3	43.1
Means	53.4	38.6

We can conclude that a simple word-for-word substitution, method (1), is not satisfactory, but that an automatic dictionary combined with rewriting is a fairly satisfactory solution for translating from French into English. The problems with German are more difficult and seem to require that the machine recognize syntactic features. These conclusions, however, are of less immediate importance to us than the conclusions we can draw about this method of estimating the quality of translations: (a) The method is subjective; (b) Raters dislike the task; (c) There is considerable error variance, so that many judges are needed in order to obtain reliable means; (d) The literary skill of the rewriter is an important factor in the ratings; (e) An attempt should be made to obtain more experienced judges — either language teachers or professional translators.

Word Scores

Another way to approach the problem is to consider what a grader does when he evaluates a pupil's translation. Introspective reports indicate that he looks for two kinds of errors: (1) errors in vocabulary and (2) errors in construction. It is difficult to make these introspections more precise, for vocabulary and syntax are complexly intertwined. Nevertheless, it seems worthwhile to try.

The fact that a grader can recognize errors at all implies that he must have some personal standard against which he compares the student's work. In its most rigid form, this might consist of his own written translation; more often it is probably a rather vague set of translations that would be about equally acceptable. In order to imitate his procedures, therefore, we should have one or more explicit translations, written out in advance, that we will use as criteria. The task is then to obtain some objective measure of the relation between the test translation and the criteria.

Given a test and a criterion translation, the simplest thing to try first is to ask if they use the same words. That is to say, a score can be given by taking the number of words in the test translation which are duplicates of words in the criterion translation and then expressing this number as a fraction of the total number of words in the criterion translation. This method ignores the order in which the words are written. As an illustration:

Original: La maison se trouve à droite.

Criterion: The house is on the right.

Test: The house leans to the right.

From the criterion translation an alphabetical check list of words is prepared and the words in the test translation are checked against it:

house	1 √	
is	1	
on	1	Score = 4/6 = 0.67
right	1 √	
the	2 √√	

A number of exploratory experiments have been conducted with this method, using translations produced by students attempting to pass their language examinations in French or German and by competent translators. These studies have explored various possibilities, but none of them has been followed up with large amounts of data. Disregarding levels of significance, the studies can be summarized as follows:

(1) Five subjects with a good knowledge of both languages translated a sentence from German into English. These translations, all assumed subjectively to be 'good', were evaluated against a criterion translation. The scores ranged from 0. 73 to 0. 86. With students whose knowledge of German ranged from low to high, scores ranged from 0.19 to 0.70. For three persons with little knowledge of German, the mean score was 0.31. Four persons with a relatively good knowledge of German had a mean score of 0.65.

(2) One passage was translated from French into English by a simple word-for-word substitution, taking the first English equivalent that occurred in a French-English dictionary. The score for this translation was 0.40.

(3) One person who knew no Turkish but was familiar with the general subject matter translated a short, technical passage from Turkish into English. No dictionary was used. The score for a language as little related to English as this was 0.20. The fact that the score was not zero is due to the occurrence of common words in the two languages.

(4) In order to study the variability of the score, eleven French sentences were translated with a mean score of 0.65. The standard deviation was found to be 0.12.

(5) Seven translations of two German sentences were made by students. These were scored and the scores were compared with scores given by a grader on a longer passage containing these same sentences and also with scores on an 'objective test' of German language ability and achievement. The three measures of the students' ability were in close agreement.

(6) Since the use of a particular criterion translation may seem rather arbitrary, the check lists from six different criterion translations were combined and used to score the students' translations. With one criterion translation, there was a ceiling of about 0.86 and a mean of 0.50. When six criterion translations were combined, the ceiling rose to about 0.95 and the mean increased to 0.58. No significant changes in the rank order of the test translations resulted from this broader definition of the scoring criterion.

(7) When successive pairs of words, instead of individual words, were used to construct the check list, the scores were lower but were linearly related to the scores for individual words. With sequences of three successive words used to construct the check list, scores were very low and discrimination appeared to be lost.

(8) A word-for-word substitution of Korean equivalents for English words was made with ten sentences totalling 171 words in length. The Korean words, in the English order, were given to three Korean students at Harvard. They were asked to rewrite the sentences in Korean, ignoring as best they could their knowledge of English. Their rewritten sentences were then scored against a criterion prepared by an experienced translator. The three scores averaged 0.49. However, if differences in inflection are ignored and the word is considered correct if the root is identical, the average was 0.75. It is very likely, however, that the subjects' familiarity with English was a considerable aid to them.

(9) These same sentences were then translated again, this time using some simple rules for pre-editing the English. (a) Articles were omitted; (b) Idioms were underlined; (c) When 'of' occurred in a possessive phrase, the order of the words was inverted; and (d) When 'to' occurred in an infinitive construction, it was indicated. With this pre-editing, the word-for-word translation was repeated. The two sets of sentences, translated with and without pre-editing, were given to two groups of 31 students each in the Kyung-Bock High School, Seoul, Korea, and they were asked to rewrite them into intelligible Korean sentences. Their sentences were then scored against the criterion translation. The average score without pre-editing was 0.125; with pre-editing, 0.218. These scores are probably too low; the students were being given instruction during the summer vacation because of their poor school records.

These studies support some general comments. For human translators, a simple measure of correspondence of vocabulary correlates rather well with a subjective evaluation of the quality of the translation; a student who has achieved a given level of competence in vocabulary has probably achieved a corresponding level of competence in grammar, so the vocabulary measure will be correlated with any other measure of quality. For MT, however, the correspondence is not so close. It is possible to imagine a mechanical translation that is completely unintelligible yet contains most of the correct words. That is to say, the vocabulary measure is necessary but not sufficient. Nevertheless, we have been pleasantly surprised that so mechanical and simple a procedure gives us any discrimination at all.

Word-Order Scores

In order to supplement the simple vocabulary score, we would like to have some indicant of the syntactical adequacy of the translation. Before bringing to bear the more sophisticated concepts of modern linguistics, we decided to try the simplest possible comparison with a criterion translation. The simplest method we could think of was to compare the order of the words which were common to the test and the criterion translations. For example:

Criterion: The young boy walked fast.

Test: The fast boy had walked.

From the criterion translation a check list is again prepared, but this time the ordinal position of each word is indicated:

	Position in Criterion	Position in Test
boy	3	3 √
fast	5	2
the	1	1 √
walked	4	5 √
young	2	

The word score is 4/5 = 0.80, when scored as before. If we consider the four shared words, we find that the three checked words correspond as to order. Thus the word-order score can be stated as 3/4 = 0.75.

Thirteen people, whose knowledge of French varied from low to high, were given four 300-word French passages to translate. These translations were scored by the word-order method and also by a more subjective technique, with a grader scoring errors in words and in phrases. Furthermore, each person took two forms of an objective examination in French language achievement.

The word-order scores ranged from 0.20 to 0.72. The error scores given by the grader ranged from 1.6 to 24.4. The objective examination scores ranged from 252 to 750 (where 250 is chance performance). Thus all three measures discriminated among the translators. The average correlation between word-order scores and error scores was about 0.70, and between the word-order scores and the objective examination scores was about 0.60.

The reliability of the word-order score is reasonably good and could probably be improved by lengthening the passages. The correlation with error scores and objective examinations provides evidence for some degree of validity, at least for human translators. This technique is useful to discriminate against very poor translations, but the present evidence indicates that it may not discriminate accurately in the range that might be labelled 'good' to 'excellent'.

A slightly more sophisticated and less mechanical way to get at the syntactic aspects has been used by Koh in the Korean studies. A scoring key is constructed in advance by noting which words modify other words in the original English passage. If the rewritten Korean translation contains this same relation, one point is given. When the rewritten translations produced by the Korean high school students were scored by such a key, they obtained an average score of 8.5% on the passages without pre-editing and 23. 3% with pre-editing. The method is rather arbitrary, inasmuch as the experimenter must select in advance those syntactic relations for which credit will be given, and it is less mechanical than the word-order score, since it requires some intelligent judgment both in constructing the key and in doing the scoring. Nevertheless, it is a technique that deserves further exploration.

These methods involving a statistical comparison of the test translation with a criterion translation are certainly effective at the lower end of the scale. Whether the statistical net can be woven fine enough to catch the subtle shades of meaning that differentiate between 'acceptable' and 'good' or 'excellent', however, is still an open question.

Measures of Transmitted Information

One goal, although an unrealistic one, that we might hope to attain in translation is reversibility. That is to say, we could recover the original passage exactly by translating back again. We do not usually aspire to this goal, because it is not necessary to recover exactly the original passage. Various alternative wordings may be adequate for purposes of communication; so we hope merely to land somewhere inside this set of acceptable alternatives. When we translate we hope that something will remain invariant under translation. This something might be called the meaning or it might be called the information. Since techniques for estimating amounts of information have been developed, this line of thought leads to the suggestion that we should attempt to compare different translations to see how much information they have in common.

The method we have explored is one developed by Claude Shannon for estimating the redundancy of printed texts. Subjects guess repeatedly at successive letters, advancing to letter n + 1 after they have correctly guessed letter n. Shannon has shown how to estimate the amount of information, in bits per letter, from the frequency distribution of correct responses on the first, second, third, etc., guess. In fact, Miller and Friedman[2] have found that it is not necessary to obtain repeated guesses, since the amount of information per letter can be estimated rather closely from the percentage of times the first guess is correct. The relation is H = 5Q, where H is the number of bits per letter, and Q is the probability of being wrong on the first guess.

---

1. Shannon, C.E., "Prediction and Entropy of Printed English", Bell Syst. Tech. J. 1951, 30, 50-64.

2. Miller, G.A., and Friedman, E.A., "The Reconstruction of Mutilated English Texts", Information and Control, 1957 (in press).

The strategy we have used involves an approximation to the information formula,

$$T = H(x) - H_y(x),$$

where T is the amount of information common to x and y; H (x) is the amount of information in x; and $H_y(x)$ is the amount of information in x when y is known. Now suppose that x and y are two alternative translations of the same passage. We can estimate H(x) by asking a subject to guess successive letters according to Shannon's technique. Then we can take another subject and show him translation y; with y available to him, he now proceeds to guess successive letters in x, and so gives us an estimate of $H_y(x)$. Assuming the two subjects to have identical guessing habits, the difference between these two measures should give us an estimate of the amount of information common to the two translations. If one translation is a criterion translation, the value of T should be high when the test translation contains essentially the same information, and low when it contains relatively little of the same information as the criterion.

In a preliminary study we found that T averaged 0.8 bits per letter for two 'good' translations of a given sentence and 0.05 bits per letter for one 'good' and one 'poor* translation. Although these results indicate that the method may be feasible, it is laborious and time-consuming; we have not explored a wide variety of conditions in this way and will probably not do so unless it becomes of some further theoretical interest. It does have the slight advantage that the measure is given in bits per letter, which may be more meaningful to computer designers than some more arbitrary scale.

Reading Comprehension Tests

A possible criticism of the methods discussed so far is that they are too much concerned with the small details of a translation and too little concerned with the general purpose of making translations in the first place. The purpose, of course, is communication. The translation should be judged successful if this purpose is achieved.

In ordinary situations outside the psychologist's laboratory, we have a simple check on whether we have communicated successfully. We ask questions. For example, after a series of communicative acts that he calls 'lectures', a teacher will evaluate his success by a procedure that he calls an 'examination'. If the recipients of a message can answer correctly questions which they could not answer before they received the message, we conclude that the communication was successful.

One way to apply this technique is in the form of commands that must be carried out by some gross, bodily behavior. A more convenient way is to ask questions that can be answered verbally. For example, in order to evaluate the readability of a particular passage, psychologists give the reader a few minutes to study it and then ask him a series of questions ranging from very simple to very difficult. Once a set of passages has been standardized for readability on a large sample of readers, it can be used to measure the reading skill of other individuals. Such a set of passages with related questions is called a 'reading comprehension test'. It should be relatively straightforward to apply this same technique to measure the comprehensibility of a translation.

The translation to be tested would be presented to a person along with a list of questions that he must answer about the meaning of the passage. These questions should be simple enough that an intelligent person equipped with a good translation could answer them all, yet difficult enough that a person with no translation could not answer any of them. We have hesitated to adopt this approach because the phrasing of the questions requires much skill and the test should be standardized on relatively large groups of subjects.

For example, the subject might be presented with the following word-for-word translation of a German passage:

> The theory the passage of sound through plates is — for even waves and bounded bundle — in such form given that the relation with it the free waves of the plate in appearance steps. Cremer's conception the total number of passages as 'coincidences' the falling in wave with it free waves of the plate, certain exceptions hereof and the influence a final cross section of the wave are discusses. The conclusions are experimental with it ultra-sound on aluminum plate proven.

Then he would be confronted by questions like the following:

1. What does the form of the theory reveal?
2. What was done with the conclusions?
3. What kind of incident sound was studied analytically?

4. What kind of incident sound was studied experimentally?

5. Was Cremer's theory accepted without qualification?

6. What did Cremer think was coinciding?

Although these questions have not been tested in any way, it is hoped that they will be difficult to answer until you have read the following alternative translation:

> The theory of transmission of sound — plane waves and laterally bounded beams — through plates is given in a form which reveals the connection with the free waves in plates. Cremer's interpretation of total transmission as 'coincidence' of the incident wave with a free wave in the plate, certain exceptions from that representation, and the influence of the finite cross section of the beam are discussed. The conclusions have been examined experimentally on aluminum plates by ultrasonic waves.

This example should make clear the difficulties involved in formulating good questions. On the one hand, they should not be so specific as to require a particular word in answer, for this reduces to a vocabulary test. On the other hand they should not be so general that it is difficult to decide whether the answer is right or wrong. No doubt special passages would have to be constructed for the purpose; we have not yet undertaken this formidable task.

Syntactic Analyses

All of the scaling procedures discussed above are linguistically naive. We have been much impressed by the elegance of certain theories of grammar. For example, Z. Harris' constituent analysis should certainly yield some kind of measure of agreement between the true analysis and the constituents of the translation to be tested. However, these ideas have been difficult to apply because the translations produced by some of the simpler mechanical procedures are so bad that it is impossible to say what the constituents are. Such analysis is easier if the translation is grammatical.

Ideas concerning the degree of grammaticalness of a passage are suggested in the work of A. N. Chomsky. For example, if words are classified into syntactic categories, we might ask how often ungrammatical sequences of categories occur. As a variable we could examine the degree of precision of the syntactic classification. A very grammatical translation would have only permissible sequences even with the most refined analysis of categories, whereas an ungrammatical translation might not have only permissible sequences until the categories were reduced to something as crude as Noun, Verb, Adjective, and X, where X represents everything else. This is a forbidding task to undertake, however, and does not get at the question of whether the translation, grammatical or not, carries the same meaning as the original. Indeed, much syntactic analysis carefully avoids any contamination with semantics.

We have assumed, therefore, that such analyses are much more important for workers trying to develop translating machines than for those who would like to evaluate the finished product.

Our studies have not explored the closely related problem of measuring the "translatability" of the original passages. We have observed, of course, that with respect to English, French is more translatable than German. But there are many other differences. The literature in any given language is not uniformly translatable, and some schemes for MT may succeed with one author and fail with another. For example, a passage which is well written in the original language will usually be more translatable than a poorly written passage. Or, again, a passage written by a person who knows no English will usually be harder to translate into English than something written in the same language by a person whose first language was English. Only a large sample of different materials in the source language can inform us on this question, and it is impractical to generate such a sample by manual simulation. Thus there are important aspects of the evaluation problem that cannot be studied satisfactorily until the machines are running.

The Linguistic Basis of a Mechanical Thesaurus †

M. A. K. Halliday, Cambridge Language Research Unit, Cambridge, England

The grammar and lexis of a language exhibit a high degree of internal determination, affecting all utterances whether or not these are translated from another language. This may be exploited in a mechanical translation program in order to cope with the lack of translation equivalence between categories of different languages, by the ordering of elements into systems within which determination operates and the working out by descriptive linguistic methods of the criteria governing the choice among the elements ranged as terms in one system. Lexical items so ordered form a thesaurus, and the thesaurus series is the lexical analogue of the grammatical paradigm.

A FUNDAMENTAL problem of mechanical translation, arising at the levels of both grammar and lexis, is that of the carry-over of elements ranged as terms in particular systems; i.e., systems established non-comparatively, as valid for the synchronic and syntopic description of what is regarded for the purpose as 'one' language. The translation process presupposes an analysis, generally unformulated in the case of human translation, of the source and target languages; and it is a commonplace that a one-to-one translation equivalence of categories - including not only terms within systems but even the systems themselves - does not by itself result in anything which on contextual criteria could be called translation. One might, for example, be tempted to give the same name 'aspect' to two systems set up in the description respectively of Chinese and English, on the grounds that both systems are the grammatical reflection of contextually specified categories of a non-absolute time-scale in which components of a situation are ordered in relation to one another; not only would the terms in the systems (e.g. Chinese and English 'perfective') not be translationally identifiable: not even the systems as a whole (unless a neutral term was introduced to universalize them) could be assigned translation equivalence.

---

† This is one of a series of four papers presented by the Cambridge Language Research Unit to the October 1956 Conference on Me-, chanical Translation (for abstracts see MT, Vol. II, No. 2, pp. 36-37).

Syntax

Where translation is handled as a function between two given languages, this problem can be met by a comparative description of the kind that has come to be known as 'transfer grammar', in which the two languages are described in mutually (or unilaterally) approximating comparative terms. For mechanical translation this is obviously unsatisfactory, since each language would have to be analyzed in a different way for every new language on the other end of the translation axis. On the other hand the search for categories with universal translation validity, or even with validity over a given limited group of languages, whether it is undertaken from within or from outside language, could occupy many years; and while the statistical survey required for the intralinguistic approach would be, for the linguist, perhaps the most pleasing form of electronic activity, the pursuit of mechanical translation cannot await its results!

In practice, therefore, we compromise, and make a descriptive analysis of each language which is at the same time both autonomous and geared to the needs of translation. We then face the question: what is the optimum point at which the source language and the target language should impinge on one another? Let us suppose we possess two documents: one, consisting of a descriptive analysis of each of the two languages, the other, a body of texts in the two languages, the one text a translation of the other. In the first document we find that in Language 1 there is a system A with terms n, o, p, and in Language 2 a system B with terms

q, r, s, t. The second document reveals a translation overlap between these systems such that we can make a synthesis as follows: Language 1, system A_1, terms n_1, o_1, p; Language 2, system A_2, terms n_2, o_2, q, r, where the use of the same letter indicates probability greater than a certain arbitrary figure that translation equivalence exists. Meanwhile document one has specified what are the determining features (contextual, grammatical etc.) of the two systems, and the proportional overlap between the two sets of determining features represents the minimum probability of translation equivalence. The actual probability of translation equivalence is always greater than the determining features show, because although (a) if a contextual feature X determines both n_1 and n_2, there is predictable equivalence since by definition if X is present for one text, it is present for its translation, yet (b) if n_1 is determined by a grammatical feature Y of Language 1 and n_2 by a grammatical feature Z of Language 2, there is no predictable equivalence though equivalence will arise whenever Y is found to be the translation equivalent of Z.

Since translation, although a mutual relation, is a unilateral process, what we are interested in is the choice of forms in the target language, let us say Language 2. Document one (which is presumed for this purpose to be ideal, though it must be stressed that at present there is no language which does not still require to be swept by many maids with many (preferably electronic) mops before such an ideal description is obtained) has given us the determining features of all forms in Language 2, and document two has shown us what forms of Language 2 can be predicted with what probability to be the translation equivalents of what forms of Language 1. (However ideal document two, there can never be certainty of equivalence throughout; the reason will be clear from document one, which shows that it is not the case that all languages are determined by the same features differently distributed, but that features which are determining for one language are nondetermining for another.) The final output of the translation process is thus a result of three processes, in two of which the two languages impinge upon one another. First we have translation equivalence, second, equivalence of determining features, third, operation of particular determining features in the target language. This is not necessarily a temporal order of procedure, but it may be illustrated in this way: suppose a Chinese sentence beginning ta zai nali zhu-le xie shihou giu . . . Translation equivalence might give a positive probability of Chinese non-final perfective = English simple past perfective: zhu-le = lived. (This identification is chosen for the sake of example, and is based merely on probability.) Equivalence of determining features overrules this by showing that some feature such as "past time reference relative to absolute past time" determines English past in past perfective: zhu-le = had lived. A particular determining feature of English, however, connected with the non-terminal nature of the time reference (which is irrelevant in Chinese) demands the imperfective: so we get "When he had been living there for some time. ."

Now the 'ideal' translation may be thought of as the 'contextual' one: it is that in which the form in Language 2 operates with identical effect in the identical context of situation as the form in Language 1. Theoretically, the one thing which it is not necessary to have to arrive at such a translation is the original: the first of the three processes above can be left out. But in translation in practice, one always has the original (the text in the source language), and what one does not have is the complete set of its determining features. The human translator may implicitly abstract these from the text, but this may not be wholly possible in any given instance, since the text may not contain indications of them all; and in any case the computer cannot do this until we have the complete ideal linguistic description. In mechanical translation the second of the three processes becomes the least important because it can be least well done; and the computer must concentrate on the first and the third: that is, the translation equivalence between source and target language, and the particular determining features of the latter. The less use made of comparative systematization, the more use must be made of the particular systematization of the target language. In translation as in any other linguistic composition a great deal is determined internally, by the structure of the target language; if the source language is going to yield only, or mainly, translation equivalence (as it must unless, as said above, we are to have a different description for each language in each pair in which it occurs) maximum determination must be extracted from within the target language.

For this we require a systematic description of the target language, which will be the same

whatever the source language, since it is accounting for features that are quite independent of the latter. It is quite clear what this means for the grammar: a formal grammatical analysis which covers the description of the relations between grammar and context to the extent of those contextual features which can be abstracted from the language text (not those which are dependent on situational features not themselves derivable from the text). In the example given above, we have to get both the past in past (had lived) and the imperfective (been living) from English context-grammar alone (if you try to get them through the source language text the procedure will be immensely complicated and will depend on transfer grammar, thus losing generality, since each source language will then have to have a different treatment for every target language, i.e. the Chinese of Chinese-English will be different from the Chinese of Chinese-Russian, without in any way simplifying the treatment of the target language): to get the English tense-aspect complex out of the English is relatively simple, whereas to get it out of the Chinese is absurdly complicated. There will be in other words a mechanical grammar of target English to account for the internally determined features of the language. One has only to think of source texts in Italian, Russian, Chinese and Malay to realize how much of the grammar of the English output would be left undetermined by the highest common factor of their grammatical translation equivalences.

Lexis

The problem has been discussed so far in terms of grammar, but it arises in the same way with the lexis. The first stage is likewise one of translation equivalence, the second stage is the use of the determining features of the target language. The question is: how can the lexis be systematized so as to permit the use of 'particular' (non-comparative) determining features, and especially, is it possible to operate the second stage to such an effect that the first stage can be almost restricted to a one-to-one translation equivalence (in other words, that the number of translation homonyms can be kept to a minimum, to a number that will be as small as, or smaller than, the number of historically recognized homographic (or, with a spoken input, homophonic) words in the language), which would clearly be of great advantage to the computer?

What is required is a systematic arrangement of the lexis which will group together those words among which some set of 'particular' determining features can be found to operate. Any arrangement based on orthography or phonology is obviously useless, since orthography plays no, and phonology very little, part in determining the choice of a given word at a given time. A grammatical arrangement by word classes adds nothing if, as is proposed, grammatical features are to be carried over separately as non-exponential systems, since classification is also in the main irrelevant to word determination, and where it is not, the grammar will do all that is required. (This merely amounts to saying that we cannot use grammar to determine the lexis because grammar will only determine the grammatical features of the lexis.) The form of grammatical systematization suggested above gives the clue: what is needed is a lexical arrangement with contextual reference. The lexis will be ordered in series of contextually related words, each series forming a contextually determined system, with the proviso that by context we mean (a) collocation, that is specifically word context, the statistically measured tendencies for certain words to occur in company with certain others, and (b) those non-collocational features of the context which can be abstracted from the language text.

The lexis gives us two points of advantage over the grammar, in reality two aspects of the same advantage, which arise from the fact that lexis reflects context more directly than does grammar. In the first place, one-to-one translation equivalence has a higher probability of resulting in translation in lexis than in grammar — there are whole regions of the lexis, especially in technical vocabulary, where it works with near certainty; and in the second place, where there is no 'term' (word) equivalence there is usually at least 'system' (series) equivalence. So we exploit the first advantage by giving one-to-one equivalence at the first stage, and the second advantage by the 'series' form of arrangement.

Thesaurus

The type of dictionary in which words are arranged in contextually determined series is the thesaurus. Each word is a term in one, or more than one, such series, and the translation equivalents provided by the first stage of the dictionary program function as "key-

words" leading in to the second, the thesaurus, stage. Each word will pass through the thesaurus, which will either leave it unchanged or replace it by another word in the series.

Each thesaurus entry, that is one series with its "key-word(s)", thus forms a closed system among whose terms a choice is to be made. We are already in the target language as a result of the translation equivalence of the first stage, and a pre-thesaurus output would be an interlingual form of the target language including some elements which were not words — since some key-words are in fact non-verbal symbols introduced to deal with the 'partial operator' sections of the lexis, to which we shall return later.

By the time the thesaurus stage of the dictionary program is reached we have one word in the target language (more than one word in the case of homonyms, and a symbol in the case of partial operators). We may also have a general context indicator from the source language of the type that most mechanical translation programs have envisaged, giving a clue to the generalized class of discourse in which we are operating. How much is still left to be provided from the resources of the target language itself can be gauged from a few specimens of non-technical railway terminology given below. Only four languages have been used, English, French, Italian and Chinese; and three of these are in close cultural contact; and yet there is so much overlap that we have a sort of unbroken "context-continuum" ranging (in English) from "railway station" to "coach". It is admittedly something of a tour de force, in that the words used are not the only possible ones in each case, and adequate translation would result, at least in some instances, from the use of other words. But if we consider each language in turn as a source language, each one is a possible non-translation form, and a one-to-one word equivalence would clearly not result in translation between any pair of languages, let alone among the whole four. Moreover, the sentences used were not chosen as containing words especially liable to overlap, but merely because the present writer happens to be interested in railways and in the linguistics of railway terminology.

Each sentence is given in English, because it is the language of this paper, together with a brief indication of situational or linguistic context where necessary. The underlined words, and the words in the French, Italian and Chinese lists, are contextual translations of each other: that is, words which a speaker of each language would be likely to use in an utterance having the same 'meaning' (i.e. the same place in the same sequence of linguistic and non-linguistic activity) in the same situation. They are considered as operating in a spoken text, where much of the context is situational; but in a written text, which we envisage for mechanical translation at present, the absence of "situation" is compensated by a fuller linguistic context, which is what the computer can handle. It should be stressed that, although only one word is given in each case, this is not regarded as the only possible word but merely as one which would not be felt to be out of place (this is in fact implicit in the criterion of 'the same meaning', since if it were felt to be out of place it would alter the context-sequence).

Finally, the English is British English; I do not know the American terms, but I suspect that even between British and American English there would be no one-to-one translation equivalence!

As with grammar, the systematization of the features determining the choice among terms in a lexical series requires a vast amount of statistical work, the result of which will in fact be the simplest statement of the lexical redundancy of the language. This redundancy is reflected in the fact that the terms in the commutation system operating at any given point in a context sequence are very restricted. (Two terms in a system are said to commute if one can be replaced by the other in identical context with change of meaning. If no such replacement is possible, or if replacement is not accompanied by change of meaning, they do not commute.) The restrictions can be systematized along a number of different dimensions, which will vary for different languages. The sort of dimensions that suggest themselves may be exemplified from the sentences below.

(i) Chinese huochezhan, chezhan and zhan in (2), (3) and (4) do not commute; they might commute elsewhere (e.g. huochezhan and chezhan, to a bus driver) but here they are contextually determined along a dimension which we may call 'specification', ranging from the most general term zhan to the most specific huochezhan. In mentalist terms, the speaker or writer leaves out what is rendered unnecessary by virtue of its being either

"given" in the context (linguistic or situational) or irrelevant. The computer does not know what is irrelevant — in any case irrelevance is the least translatable of linguistic phenomena — but it does know what is given, and would select zhan here if certain words are present in the context (railway terms such as huoche, and the ting (stops) of (5)), chezhan if there is some reference to a specific form of travel, and huochezhan otherwise.

(ii) English track, line, railway: the choice in (12), (14) and (16) is not a matter of specification but of classification. Like the three Chinese words, they may denote one and the same physical object; but their connotations are as it were respectively 'ential', functional

NON-TECHNICAL RAILWAY TERMINOLOGY

Situational or Linguistic Context	English	French	Italian	Chinese
1. Here's the railway station (pointing it out on a map),	railway station	gare	stazione ferroviale	huochezhan
2. How do I get to the station? (inquiry in the street).	station	gare	stazione	huochezhan
3. Station, please! (to taxi driver)	station	gare	stazione	chezhan
4. There's one at the station (on the way to the station, to companion who inquires e. g. about a post office)	station	gare	stazione	zhan
5. How many stations does it stop at? (on the Underground)	station	station	stazione	zhan
6. It's two stops further on.	stop	arrêt	fermata	zhan
7. It doesn't stop at the halts (i.e. only at the staffed stations)	halt	halte	fermata	xiauzhan
8. Travel in this coach for the country platforms.	platform	point d'arrêt	fermata	yetai
9. They' re mending the platform.	platform	quai	marcia-piede	yetai
10. He's waiting on the platform .	platform	quai	marcia-piede	zhantai
11. The train's at Platform 1.	platform	quai	binario	zhantai
12. I dropped my cigarettes on the track (while waiting at station)	track	voie	binario	guidau
13. Don't walk across the line.	line	voie	binario	tiegui
14. The trains on this line are always late.	line	ligne	linea	lu
15. There's a bridge across the line.	line	ligne	linea	tielu
16. He works on the railway.	railway	chemin de fer	ferrovia	tielu
17. I'd rather go by rail .	rail	chemin de fer	ferrovia	huoche
18. Let's go and watch the trains .	train	train	treno	huoche
19. Get on to the train! (standing on platform)	train	train	treno	che
20. There's no light in this coach .	coach	voiture	vettura	che

and institutional. A purely locational context could give 'track', a proper name 'railway'; 'line' overlaps with both (cf. (13) and (15)) and might be limited to functional contexts such as 'main line'.

The word as a term in a thesaurus series is grammatically neutral: it is neutral, that is, as to all grammatical systems, both categories of the word (e.g. number) and word class itself. Since we cannot carry over the classes and other categories of the source language as one-to-one equivalences (e.g. Chinese verb ≠ English verb, Chinese plural ≠ English plural, even if both languages are described with categories named 'verb' and 'plural'), these are dealt with in the grammatical part of the program and only after having reached the target language do they re-enter the range of features determining word choice. The attempt to handle such categories lexically leads to impossible complexity, since every word category in each source language would have to be directly reflected in the thesaurus.

All mechanical translation programs have carried over some word categories non-lexically, word-inflections obviously lending themselves to such treatment. If in the thesaurus program the word is to be shorn of all grammatical features, including word class, the whole of the grammar must be handled autonomously, and the method proposed for this is the lattice program originated and developed by Margaret Masterman and A.F. Parker-Rhodes. The lattice program, which is a mathematical generalization of a comparative grammar (i.e. a non-linguistic abstraction from the description of a finite number of languages) avoids the necessity of the comparative (source-target) identification of word (and other grammatical) categories. The word class of the target language is determined by the L(attice) P(osition) I(ndicator), derived from the grammar of the source language; class is thus not a function of the word as a term in the thesaurus series, nor does the choice of word class depend on comparative word class equivalences.

The autonomy thus acquired by the lexis of the target language allows the thesaurus stage of the dictionary to be the same for one target language whatever the source language, and at the same time permits the maximum use of the redundancy within the target language by allowing different treatment for different sections of the lexis. This would be impossible if word classes were based on translation equivalence, since the thesaurus series could not form closed systems within which determination can operate. If for example one identified particularly (i.e. non-comparatively) a word class 'conjunction' in the target language, the redundancy of the conjunction system can only be fully exploited if it is determined (as it is by the LPI) that the choice word must be a term in this system. If we attempted to carry over to Chinese word classes from, say, English, where we could not identify any grouping (let alone class) of words which would have valid translation equivalence with Chinese 'conjunction', we should forfeit the redundancy of the Chinese system since the words among which we should have to choose could not be ordered as terms in any lexical series.

The thesaurus admits any suitable grouping of words among which determination can be shown to operate; the grouping may be purely lexical or partly grammatical (i.e. operating in the grammatical system of the target language). It might be that a word class as such, because of the redundancy within it, was amenable to such monosystemic treatment. This is clearly not the case with the 'non-operator" (purely lexical) sections of the lexis, such as verbs and nouns in English, but may work with some partial operators. (Pure operators, i.e. words not entering into lexical systems, which are few in any language (since their work is usually done by elements less than words) — Chinese de is an example — will not be handled by the thesaurus, but by the lattice program.) The nouns in the above sentences enter into lexical series, but no determination system can be based on their membership in the word class of 'noun'; prepositions, on the other hand, which are few in number — and of which, like all partial operators, we cannot invent new ones — can in the first instance be treated as a single lexical grouping.

It is simply because partial operators (which in English would include — in traditional 'parts of speech' terms — some adjectives (e.g. demonstratives and interrogatives), some adverbs (those that qualify adjectives), verbal operators, pronouns, conjunctions and prepositions) are in the first instance grammatically restricted that they have a higher degree of overall redundancy than non-operators. Knowing that a noun must occur at a certain point merely gives us a choice among several thousand words, whereas the occurrence of a verbal operator is itself highly restrictive.

An idea of how the thesaurus principle might be applied in a particular instance may be given with respect to prepositions in English. In dealing with the English prepositions we can begin by considering the whole class as a lexical series. We can then distinguish between the 'determined' and the 'commutable'. Most prepositions are determined in some occurrences and commutable in others. The 'determined' prepositions are simply those which cannot commute, and they are of two types: the pre-determined — those determined by what precedes (e.g. 'on' in "the result depends on the temperature at . . ", which cannot be replaced, or 'to' in " .. in marked contrast to the development of . .", which could be replaced by 'with' but without change of meaning), and the post-determined — those determined by what follows (e.g. 'on' in "on the other hand", or 'to' in "to a large extent"). In the system of each type we may recognize one neutral term, pre-determined 'of' and post-determined 'to'.

Determined prepositions will be dealt with not as separate words but as grammatical forms of the word by which they are determined. The combination of pre-determining word plus preposition will constitute a separate entry, a transitized form of the determining non-operator (verb, noun or adjective, including adverb formed from adjective), of which the occurrence is determined by the LPI. The features determining the occurrence of these forms are grammatical features of the determining word; they are connected in varying ways with the presence or absence of a following noun (group): 'depends / depends on A', 'a contrast / a contrast with A', 'liable to A'; but 'wake up / wake A (up)'. Which form of the word (with or without preposition) corresponds to which lattice position will be indicated if necessary in the same way as other word class information; in the absence of such indication the transitized form of words which have one is used before a noun. If a verb is not assigned a marked transitized form, it is assumed not to have one, and will be left unaltered in a lattice position that would require a transitized form if there was one; but if a noun or adjective without transitized form occurs in the corresponding lattice position the neutral term 'of' is to be supplied. Thus 'depend', 'contrast (noun)' have the transitized forms 'depend on', 'contrast to'; 'display', 'production', 'hopeful' have no transitized forms, and will thus give 'display of (power)', 'production of (machinery)', 'hopeful of (success)'.

Post-determined prepositions are always treated as part of a larger group which is entered as a whole. These are forms like 'at least', 'on the whole', 'to a large extent', and are single words for thesaurus purposes. The exception is the neutral term 'to' before a verb (the 'infinitive' form). This is treated as a grammatical form of the following word (the verb) and will be used only when required by the LPI, e.g. in a two-verb or adjective-verb complex where the first element has no predetermined (or other) preposition: 'desires to go' but 'insists on going' — all other prepositions require the -ing form of verbs —, 'useless to go' but 'useless for (commutable) experiment'.

Determined prepositions in the English version of the Italian pilot paragraph are:

Pre-determined:	of 1 - 6
Post-determined:	at least; on the other hand; in fact; for some time past; above all; to mechanize.

Commutable prepositions operate in closed commutation systems of varying extent (e.g. 'plants with/without axillary buds' (two terms only), 'walked across/round/past/through/towards etc. the field'), and each one may enter into a number of different systems. Those which are lexical variants of a preceding verb are treated as separate lexical items, like the pre-determined prepositions (e.g. 'stand up', 'stand down', and favorites like 'put up with'). The remainder must be translated, and among these also use is made of contextual determination.

The overlap in this class (i.e. among words in source languages which can be translated into words of this class in English) is of course considerable, as one example will show:

Sentences:	English	Italian	Cantonese
He went to London	to	a	
He lives in London	in	a	hai
He came from London	from		hai

We can however set up systems limited by the context in such a way that the terms in different systems do not commute with one another. For example, concrete and abstract: to / in / from commute with each other but not with in spite of / for / without. Within the concrete we have motion and rest: to / from commute with each other but not with at / on / under; and time and place: before / after / until

commute with each other (in some contexts before / until do not commute but are grammatically determined) but not with under / at.

Commutable prepositions of this type will go through the usual thesaurus program in which they form series on their own (whereas determined prepositions and the 'lexical variant' type of commutable prepositions do not); the context will specify in which system we are operating. If the source language has words to which English prepositions are given as translation equivalents, these will as usual be one-to-one (with limited homonymy where necessary: Cantonese hai would have to give 'be at (English verb or preposition according to LPI); from (preposition only)', since on grounds of probability the motion context equivalent of 'at' will be motion towards, not away from). Each key-word will in the usual way lead into a series the choice within which will be determined by the context category.

Commutable prepositions in the Italian pilot paragraph are:

Lexical variants:	none	
Free commutables:	with	(It. a, abstract 'with (/without)'
	for	1 - 4 (It. per, abstract)
	in	(It. in, abstract)

This paragraph is typical in that the freely commutable prepositions are a minority of the total prepositions in the English output.

Thus the thesaurus method,which uses the contextual determination within a language, is applicable to partial operators through the handling of redundancy at the level at which it occurs: where the use of a preposition depends on grammatical or lexical features (considering English forms like 'put up with' to be lexical, not contextual, variants) it will be handled
accordingly, and not as a term in a lexical preposition series. The method is far from having been worked out in full; the principle on which it rests, that of "make the language do the work", can only be fully applied after the linguists have done the work on the language.

Bibliography

Noam Chomsky 96
Syntactic Structures
Mouton and Co. , (1957), The Hague, 116 pages.

Three models for linguistic structure are discussed from the point of view of their respective advantages and disadvantages. Rejecting, because of their restricted usefulness, the communication theory model and the phrase structure model based on immediate constituent analysis, the author convincingly argues for the superior qualities of a third, transformation model. Phrase structure analysis, according to this theory, applied in the description of what the author terms kernel sentences, and the formulae for transforming these into more complex syntactical sequences, once they have all been elaborated, will in the simplest possible manner generate all the grammatical sentences in the language whose structure they state in terms of the kernel sentences. The grammar of a given historical language is thus regarded as a deductive system, one whose complexity must necessarily far exceed that of the formal language whose grammar is mathematical logic.

B. Ulvestad

J.E. Holmstrom 97
Bibliography of Interlingual Scientific and Technical Dictionaries
Published by UNESCO, third edition, (1953), 178 pages.

This bibliography lists 1629 dictionaries under 236 subject heads and 75 languages. The subject heads are classified in the Universal Decimal Classification. Many of the entries contain a brief description of the dictionary.

V.H. Yngve

Eugen Wüster 98
Bibliography of Monolingual Scientific and Technical Glossaries
Published by UNESCO (1955), 219 pages.

This bibliography lists 1581 monolingual scientific and technical glossaries arranged by subject according to the Universal Decimal Classification.

V.H. Yngve

A.N. Nesmjanov 99
Science and Industry
Vestnik Akademii Nauk SSSR, Vol. XXVI, No. 2, pp. 1-19,(1956).

This is a general report by the president of the Academy of Sciences of the USSR on the application of science to problems in industry and economy in the USSR. On pp. 16-17, the experiments on the BESM are reported.

V. H. Yngve

P.S. Kuznetsov, A.A. Lyapunov, 100
and A. A. Reformatskiy
Fundamental Problems of Machine Translation
Voprosy Yazykoznania, Vol. 5, No. 5, pp. 107-111 (Sept. -Oct. 1956)(in Russian).

This is the introductory article of a series to be published in this journal. The method of operation of computing machines is discussed. Certain prejudices against mechanical translation are dispelled. The linguistic problems associated with mechanical translation are discussed, particularly with reference to the grammatical and syntactic differences between languages.

V.H. Yngve

O.S. Kulagina and I. A. Mel'chuk 101
Machine Translation from French to Russian
Voprosy Yazykoznania, Vol. 5, No. 5, pp. 111-121 (Sept. - Oct. 1956) (in Russian)

The present article discusses the principles of formulating a dictionary and a "grammar" (a system of special rules) for mechanical translation of a scientific text in the field of mathematics from French to Russian, and describes the translation process itself.
A translation requires the following data: (1) a stem dictionary; (2) an idiom dictionary; (3) preposition translation tables; (4) rules for distinguishing homographs; (5) affix tables (in French and Russian); (6) groups of analyzing rules; (7) groups of synthesizing rules. These data are stored in the "memory" of the machine as needed.

Authors

L.I. Zhirkov 102
Limits of Applicability of Machine Translation
Voprosy Yazykoznania, Vol.5, No. 5, pp. 121-124, (Sept. - Oct. 1956) (in Russian).

The article states that there is a limit between what a machine can do effectively and what should be left to a thinking person. A broader role for the post-editor would be advisable, while the translation process itself can be left to the machine.
It also relates that in 1938 an inventor, P. P. Smirnov-Troyanskiy, claimed to have worked out a mechanical translation method which was not related to the concept of electronic computers. His invention was met with skepticism and an experimental model was never built. He has since died.

V.H. Yngve

A. Koutsoudas and R. Korfhage 103
Mechanical Translation and the Problem of Multiple Meaning
Mechanical Translation, Vol.3. No. 2, pp. 46-51,

A report of research done at the University of Michigan in the analysis of language structure for mechanical translation. The rules and vocabulary developed for translating a Russian text into English are presented. A simulated mechanical translation made with these rules and vocabulary by a volunteer having no knowledge of Russian is included with the original Russian text and a standard translation.

J. R. Applegate

L. Brandwood 104
Mechanical Translation of French
Mechanical Translation, Vol. 3, No. 2, pp. 52-61.

A description of the procedure developed at the computational laboratory of Birkbeck College. The research was directed toward the development of a computer program. Because of the limited storage capacity available, the problems discussed are primarily morphological rather than syntactic in nature.

J. R. Applegate

105
Reports from the International Conference on Machine Translation
Mechanical Translation, Vol. 3, No. 2, pp. 33-45

The reports presented by the several groups represented at the conference serve to indicate the goal toward which research in each group is directed. The project at the University of Washington, directed by Erwin A. Reifler, reported on methods that have been developed for the recognition of grammatical categories of words and groups of words in both Russian and English. Because there is close cooperation between linguists and engineers, the linguistic re.search has been influenced by problems of economical storage etc.
Research at the University of Michigan has been directed toward the solution of the problem of multiple meaning. From the analysis of a Russian text some progress has been made toward the formulation of algorithms which make possible the selection of one meaning from a list of several meanings given for each word in a sequence.

At the Institute of Languages and Linguistics, Georgetown University, a group directed by Leon A. Dostert has based its research on the postulate that translation is a problem of correlating two systems of signs each of which may be incomplete with multivalent symbols. They have been concerned with the development of signs that will make the two original systems correlate exactly.

The Language Research Unit at Cambridge, England has also been concerned with the problem of multiple meanings. Much of their work has been centered about the development of a mechanical thesaurus and the development of a code that permits the selection of the correct meaning equivalent after a consideration of contextual clues.

A similar approach is being considered by the International Telemeter Corporation. Here, however, the use of contextual clues for determining meaning is bolstered by the use of statistical information about the probability of given meanings in specific contexts. The basic assumption is that Western languages are 50% redundant.

At the University of California, experiments have been made with Russian scientific literature. As a result, it has been decided that the chief difficulty is the syntactic and semantic ambiguity of individual words. Recent research has been concerned with the use of contextual analysis to eliminate this ambiguity.

The Academy of Sciences, Moscow, Russia , reports that it has been found that maximum separation of the dictionary and syntactic parts of the mechanical translation program is desirable . The two main parts of the translation program at the syntactic level are analysis of the input sentence and synthesis of the corresponding output sentence.

At M.I.T., work has been focussed on the structural analysis of German sentences for the purpose of developing a routine for analyzing these sentences and synthesizing the corresponding English sentences. From the beginning the research has been directed toward sentence-for-sentence rather than word-for-word translations .

J. R. Applegate

MEETING PAPERS

At the National Convention of the Inst. of Elec. Comm. Eng. of Japan in April 1957, a paper "Syntax Analysis for Mechanical Translation into Japanese" was presented by Prof. Tsuneo Tamachi of Kyushu University.

At a meeting of the Seattle, Washington, Section of the Institute of Radio Engineers, May 9, 1957, a discussion of automatic language translation was held. The speakers were E. Reifler, L.R. Micklesen, W.R. Hill and R.E. Wall, Jr., all of the University of Washington.

The American Oriental Society, at its 167th Annual Meeting, April 24 - 26 at Princeton had a session on mechanical translation. J. Bigelow of the Institute of Advanced Study at Princeton and Erwin Reifler presented papers.

At the 12th national meeting of the Association for Computing Machinery at the University of Houston, Texas, June 19-21, 1957, a paper "Conclusions on Language Translation" was presented by A.F.R. Brown of Georgetown University.