SOME METHODS OF MECHANIZED TRANSLATION

by
R.H. RICHENS and A.D. BOOTH

Lisez, lisez; jetez vos grammaires
au feu.
Schlegel.

The following outline summarizes some suggestions about ways in which translation may be mechanized that were worked out by the authors in 1948-1952. The treatment is not exhaustive, and many modifications in detail would certainly be necessary should the ideas here mooted be put into active operation. It is considered, however, that it would be serviceable to describe the techniques here envisaged in their present rudimentary form, rather than await their further refining, since they may prove of interest to others working along the same lines.

The paper is divided into three parts. The first deals with general principles, the second with specimen translations, while the third considers in detail the working of some specific mechanical translation schedules.

I. General Principles of Mechanized Translation

A language is a series of symbols representing ideas. The simplest conceivable written language would have one symbol per idea, together with appropriate rules, possibly involving extra symbols for syntactical relations. Such a language does not exist. The nearest approach is probably Chinese or the numerical notation in general use. In both these cases the symbols represent ideas directly. Usually, written language symbolizes, not ideas, but other symbols, namely sounds, spoken words, that represent ideas. Since, however, there are many logical deficiencies in the way in which sounds are used to express ideas, and in the ways in which writing is used to represent sounds,the double process of redaction complicates considerably the business of translation.

Taken in its most general sense, translation is the substitution of one language for another to express the same set of ideas. It should proceed by a one-one substitution of symbols for each of the ideas expressed, together with the additional changes required by a possible change in the syntactic rules.

In the foregoing statement, it has been implied that a written text is a sequence of symbols, each representing an idea or syntactical relation. This assumption obviously needs qualification. In the first place, the elementary unit of a written text is the letter, using this term in a wide sense to include in general the units assembled by a compositor. It would thus include letters as ordinarily understood; it would also include syllabics, for example the Japanese hiragana script, and it would include too the Chinese ideographs. These three categories are not, however, sharply demarcated. Thus the u of use in English is a syllabic, i.e., y + u; in the Slav languages, the yodolized vowels are also syllabics. Conversely, the Japanese hiragana syllabics include symbols for the pure vowels and for n. The Chinese ideographs normally represent ideas, occasionally, however, especially when transliterating from other languages, they act as syllabics. Some languages, notably those using Arabic script, omit short vowels. Mechanized translation, in this case, must be based entirely on the residual consonantal skeleton.

Practically all modern languages are punctuated. In the case of Chinese and Japanese no further demarcation is made, and the sequence of letters between punctuation marks, ideographs in Chinese, and a mixture of ideographs and syllabics in Japanese, constitute what may be termed a word block. In most languages, spacing is used to subdivide letter sequences

into separate words. It must not be supposed, however, that the word is necessarily the semantic unit. It is well known for instance that languages such as German and Finnish use long words compounding several ideas, where English would use several separate words.

In order to translate, it is necessary to discover the semantic units of the text, which will in some cases correspond to words or word blocks, but less often than might be supposed.

Thus in English dog does not represent just the notion of the animal; it also includes the accessory notion of being in the singular, a limitation not imposed by the corresponding Chinese ideograph chuan. Very frequently words and word blocks consist of a sequence of semantic units, usually divided by grammarians into such categories as roots and affixes. For purposes of mechanized translation, however, every semantic unit, that is, any component to which a distinct significance attaches, should be given an equivalent status.

Translation, then, in the first place involves recognition of semantic units, but before this operation can be done mechanically it is necessary to render the text into a functional form, that is, into some medium that lends itself to mechanical operations. This stage is the one around which the economic feasibility of mechanized translation hinges. Subsequent operations, though perhaps more interesting theoretically, demand little in the way of man-hours. Rendering the script into functional form is a different matter, for the simplest method, teleprinting or some comparable process involving use of a typewriter keyboard, is time-consuming in man-hours, and, in the case of the uncommoner languages, notably those in nonRoman scripts, and worst of all, those employing Chinese ideographs, the initial typing is so arduous that the translation may not be worth the trouble. In the case of handwriting, it seems unlikely that a human intermediary can be dispensed with. The process of recognition, involving a complicated set of psychological and topological operations, is performed very efficiently by human operators, and there is little prospect of machinery being developed capable of dealing with irregularly formed handwriting.

Printed matter, however, constitutes a different problem. The regularity of type face presents opportunities for comparison with standard letters using photoelectric cells, and while practical difficulties, especially with the more complicated scripts, may prove serious, there appears no theoretical reason why printed matter in any language should not be rendered functional without human intermediaries, apart of course from the necessity of setting up suitable scanning machinery. It is true, of course, that scanning machinery is incapable of allowing for smudged or malformed letters. Usually, however, any errors caused thereby will be easily spotted in the final translation. It is possible too that different letters of nearly the same shape, such as occur in Hebrew, may not be distinguishable by machinery. In this case it would be possible to treat pairs of words differing only slightly in appearance as single words with two meanings.

The aim of the functionalizing operation is to render the original text into a manipulative form, the most important being mechanically or electrically sensed punched holes, light signals photoelectrically sensed, or electromagnetically sensed magnetic signals. When rendered into a functional medium, each letter of the original text becomes represented by a distinctive modification of the functional medium, for example, a specific pattern of punched holes or magnetic signals. It is necessary however to make use of the notion of cotypes. Roman B, for instance, is differently shaped from italic B, and they may be recognized photoelectrically after comparison with different standards. However, both variants must have the same equivalent in the functional medium. Similarly, Arabic letters differ according to their relative position in the word, whether initial, medial or final. All variants however of the same letter should have the same functional equivalent, excepting possibly the t of the feminine suffix at in Arabic, where the special form of the letter has a

semantic significance.

Diacritical marks are more complicated. In Spanish, with few exceptions, the acute accent is merely a guide to the tonic accent, and can be ignored. Thus e and é in Spanish can be regarded as cotypes with the same functional equivalent. In German, however, a and ä must be distinguished, as must o, ó, ő and ö in Hungarian. Conjoint letters, as in the Devanagari and other Indian scripts should be represented in the functional medium by their components, though they would probably be recognized photoelectrically as unit characters; and in those languages using alphabets derived from the Brahmi script, including Burmese and the Dravidian languages, where some vowels precede the consonants after which they are sounded, they should be transposed into their natural order in the functional medium.

The next stage in translation is the mechanized equivalent of consulting a dictionary. The basic premise on which this chapter is based is that syntax is of quite minor importance in understanding a language. Language teaching is naturally based, in general, on the assumption that the learner will wish to express himself correctly in the language concerned. Syntax must be stressed under these circumstances. If, however, a language is being studied merely in order to translate from, the matter is different. It is true that in poetry, and in prose works depending for their effect on obscure syntactical relations, a thorough knowledge of syntax may he essential. Generally, however, in straightforward descriptive writing, the mere sequence of words, without any knowledge of syntax at all, is sufficiently revealing. This applies in particular to scientific writing, where literary obscurities are not encouraged. It has to be borne in mind too that many syntactical rules are more or less universal. Thus an uninflected subject usually precedes an uninflected object, and, even when it does not, common sense will usually detect the deviation. It is a matter of semantic indifference where the verb comes. It is important, however, to know whether a double negative is meant to be an affirmative as in English, or a negative as in Russian. It is not maintained that dictionary searching will infallibly be sufficient. It is maintained, however, that it is adequate in s high a proportion of cases, that special methods for dealing with syntactical relations between separate words may be superfluous. It is hoped that the examples given below will serve to justify this point of view.

A mechanical dictionary consists of each of the invariant semantic units of a language, whole roots or affixes, together with its equivalent or equivalents in English. The dictionary must include all the invariant words or part of words (stems) of the language to which a definite meaning or meanings attach. Thus, for Latin, am, love will appear, but none of the derived forms of amo will be given, as they are formed regularly by suffixing on to am. In the case of rego, I rule, however, three entries are required, namely, reg, rex, and rect, three different stems being involved. In other languages, notably Greek, where assimilation may result in a considerable number of invariant stems for a single verb, the same verb must be entered a corresponding number of times. The general rule is that, whenever the derivatives of a word are not formed by simple affixation, then the derivatives' stems must each appear as a separate entry. Such cases are unimportant in many languages. But they bulk large in Celtic languages, where initial mutation, as in Welsh, may necessitate that such a word as ci, dog, shall appear in four places, namely ci, gi, nghi, and chi, while eclipsis and aspiration in Irish require that for bad, boat, the stem shall appear also under bhad and mbad. Similarly the alterations in initial and final letter required by the rules of sandhi in Sanskrit necessitate that the words so affected must be replicated under each of their variants. Internal changes must be regarded as equivalent to forming a new stem, thus German Bruder, brother, and its derivative Brüder, must be entered separately. Stems in which doubled letters appear as in Dutch heb, hebb, have, or Greek words with initial r, which may be doubled in compounds, should also be treated separately.

It may be possible, however, to reduce the number of entries in the mechanical dictionaries of languages with mutable stems by arranging for certain letters or combinations of letters to have the same equivalent in the functional medium. Thus, in Welsh, it may prove possible to have a single equivalent for p, ph, or mh, which are variants of the same radical letter. Similarly the number of entries in Irish or Gaelic could be reduced by treating h as nonexistent. Such a procedure would occasionally result in two different words having the same form in the functional medium. In this case, the resultant form could simply be treated as a single word with two meanings.

Affixes are treated as separate words, for example Rumanian -lor (genitive plural definite), and Estonian -ksime (first person plural conditional active).

The process of utilizing a mechanical dictionary consists then in decomposing the functional text into its semantic units and then matching each of these with the terms in the dictionary. The process of decomposition is facilitated by punctuation and the word structure of the language. A semantic unit, as understood here, will not usually include more than one word, though there may be several semantic units in a single word. This rule does not apply of course to metaphorical or idiomatic constructions, but it is not intended to translate these other than literally. Should a word correspond exactly with a term in the mechanical dictionary, the English translation is obtained forthwith. Should the word have no exact equivalent in the dictionary, it is necessary, starting at the beginning of the word, to find the longest segment that does so correspond, treat this as the first semantic unit, then begin again where this ends, and so on till the end of the word, e.g. the Spanish word comprarlo could be decomposed as follows:-

compr	buy
ar	(infinitive)
lo	the/it

It happens in many languages, however, that the meaning of an inflection differs according to the part of speech to which it is attached or even with the particular declension or conjugation. Thus -i in Latin may be:-

genitive singular	2nd declension
nominative plural	2nd declension
dative singular	3rd declension
ablative singular	3rd declension
1st person singular perfect indicative active	2nd and 3rd conjugation
present infinitive passive	3rd conjugation
2nd person singular imperative active	4th conjugation

In addition to these meanings i is also a stem of the verb eo, I go. It is necessary in such cases to devise means whereby the translation of an inflection is determined by the preceding stem. This is possible with magnetic drums by including after the entries of stems an indication as to what part of speech they are, and if necessary the declension or conjugation. For inflections, a range of meanings is given, one for each category of stem to which they may be attached. It must then be arranged so that the particular translation made is that corresponding to the grammatical category of the prefix. It is also possible to make a similar arrangement using punched cards.* The cards carrying the stems in the mechanical dictionary should be punched for the grammatical category, declension or conjugation. Then, when the words of the original text are being decomposed, the secondary pack, into which the inflexions are punched, together with the punching for serial order, should also be punched for the grammatical category of the stem. It can be arranged so

* The way in which punched cards can be used as the functional medium and compared with a dictionary of punched cards is described below.

that the dictionary cards for the inflexions carry the translation for each grammatical category in a different panel of the card, so that, when the secondary pack is collated with these cards, the punching for grammatical category in the secondary pack can be made to determine which translation shall be punched into the latter cards.

Even when due allowance is made for inflections with various meanings, it may still happen that the rigid application of the method of decomposition described above will result in misdivision. Thus, in Latin, given the stems desid, be idle, and desider, desire, the verbal form desideremus will be decomposed into:-

desider emus desire (1st person plural present subjunctive active)

This would be a possible reading, but the correct decomposition may be:-

desid eremus be idle (1st person plural imperfect subjunctive active)

The word desidere would prove untranslatable since the decomposition desider e attaches a suffix e to a conjugation in which it does not occur. The solution to this type of difficulty is to enter any such awkward words in the mechanical dictionary in their undecomposed form, together with their true meaning or meanings.

The process described above of decomposing words into their semantic units and matching these with the entries in the mechanical dictionary must be undertaken for each word or word block of the functional text. In order to perform this operation expeditiously, two possibilities present themselves. The first is to go through the entire dictionary at high speed for each word. This process is possible using magnetic drums rotating at high speed. The second possibility is to rearrange the functional text into alphabetical order. This is possible with punched cards. If, for instance, each word of the original text is punched onto a card, together with a serial number, the cards after the entire text has been punched, can be sorted into alphabetical order. Comparison with a dictionary pack can then be effected in a collator; the English equivalent can be punched into the cards carrying the original text; and then the latter can be sorted back into their original order using the punched holes denoting serial order. In the case of compound or inflected words, after the longest segment of each word has been coped with, the residue can be punched into a secondary pack together with the serial number, and the secondary pack, after being sorted alphabetically, can be dealt with as before, and if necessary the process can be repeated. All the additional packs so obtained are then merged with the original pack and sorted into the original order as described above.

Mechanical translation admits of various degrees of precision. Increased precision normally involves the listing of more alternatives and a rather cumbersome translation. It is obviously possible to prepare mechanical dictionaries variously graded, and with supplements of technical terms for particular subjects.

The translation of any particular semantic unit may be a single expression or a row of alternatives. These can be conveniently indicated by diagonal strokes, e.g., in Spanish,

llam call/flame

Sometimes, as above, the semantic unit is translated by a semantic term of corresponding meaning. In other instances the unit to be translated may be an inflexion with a grammatical or syntactic significance. This can be conveniently represented by the use of brackets,

e.g. Lithuanian -uose (locative plural).

In highly inflected languages such as classical Greek, a standard set of abbreviations for grammatical categories would probably be useful, and these abbreviations might be employed, with advantage, for all other languages.

Thus Greek -thesontai (3d person plural future indicative passive)
or (plural feminine accusative)

could be represented in some such form as IIImgic/mfa.* Grammatical elements such as gender, which may be syntactically significant in indicating, say, which adjective is to be attached to which noun, are omitted from translation. In prose writing, ambiguous translations arising from the omission of gender will be very rare, and common sense will normally be adequate to make any decision.

In many languages, words occur that have no semantic significance, such as the various sorts of euphonic particles. French t in a-t-il is an example. This can be translated as

t = (vacuous) or v

There are also a large number of near-vacuous terms, such as the definite article in English and other languages, and the numerous adversative particles. These terms are frequently quite devoid of meaning, though not invariably. Some latitude may be allowed in their treatment. An exacting reader can employ a dictionary in which French le comes out as the whereas for others the translation will be (vacuous) or v.

It will inevitably happen that any extended passage will contain terms not in the dictionary, either because they are proper names, words from another language, misspellings or faulty printings. In all these cases the translating machine must modestly reply (untranslatable) or u.

It has been stated above that translation admits of various degrees of precision. The most precise mechanical translation is not necessarily the most suitable, since common sense may be quite sufficient to resolve any ambiguities in a laxer treatment. There may be some advantage, for instance, in having a single category for oblique cases, possibly excluding the accusative, in such languages as Latin or Sanskrit. It may also prove redundant to specify the singular number, the third person of a verb, or the present tense.

Much play has been made by critics of word by word translation of the point that separate words that form a single semantic unit, in particular words such as ne...pas or ne...que in French, or verbs with separable prefixes in German, cannot be dealt with thus.

There are two possible solutions to this problem. In the first place it is possible, using magnetic drums, to incorporate standing instructions so that when the first member of a double semantic unit appears, the translation is deferred until the second member is encountered. It would also be possible to use a parallel device with punched cards. This would require that the card carrying each word of the text should bear not only a serial number for the word, but also for the clause in which they are contained. This could be arranged by automatically changing the clause serial number at each punctuation mark. Then, assuming, for example, that the words stellt...dar occur in a German text, it can be arranged that, when collated with the dictionary cards, not only is the translation of each separate word punched in the original cards, but also a number indicating possible membership of a double semantic unit, say a 1 for a potential first member and a 2 for a potential second member. Cards with potential double semantic units could then be separated from the rest, subdivided into potential first or second members, sorted into serial clause order, and then matched with each other in a collator. It should then be possible to punch into the cards carrying the first potential member, the punching of any second potential member with the same clause serial number. Thus, a card would now be obtained punched for stell dar. Then, if a supplementary dictionary is prepared for double semantic units, this can be collated with the double-punched cards and the translation of the double unit can be punched into them. The potential first member cards would then carry not only the translation of the potential first member taken in isolation, but also its translation as a double semantic unit. This method would clearly break down in some cases, as when a clause is subdivided by commas, but it would probably work for the majority of cases.

---

* Abbreviations for grammatical units are listed in section II.

Alternatively, in view of the comparatively restricted semantic range of such double units, it would be possible to use a much simpler method, applicable to punched cards as well, of simply listing the range of meanings of each member of the double unit taken alone and then the meanings when either participates as a component in a double unit. The meanings of the double unit may appear under one or both of the components according to convenience, there being an obvious advantage in distributing the load, so to speak, as evenly as possible. Thus in French, the negatives could be treated as follows:-

ne	not
pas	not/step
que	that/which?/but only (following not)

The compounds of the German verb stellen involving separable prefixes could be treated as follows:-

stell	place/restore (after hither)/represent (after there)
an	at/on
aus	out
be	(transitive participle)
dar	there
ein	a/one/in (when prefix)
fest	firm
her	hither
heraus	outside
vor	before

In the case of German verbs with inseparable prefixes there is, of course, no problem. If the prefix retains its independent meaning, it can be treated as a separate semantic unit. If the prefix + stem constitutes the semantic unit, then the compound is entered in the dictionary. Similarly, as shown above, when a separable prefix retains its independent meaning fairly obviously as in ausstellen or einstellen, there is no problem, and the stem and prefix are dealt with separately. In other cases, however, as herstellen, to restore, and darstellen, to represent, the meanings are not immediately inferable from the components, and are given separately under stell.

A similar example of double semantic units occurs in Greek and Latin where the meaning of a preposition may vary with the case following it. It would be possible again, using rotating magnetic drums, to defer the translation until the case of the following noun has been ascertained. It may be simpler, however, merely to list the alternatives under the preposition.

The result of the foregoing operations will be a sequence of words and grammatical directives with a vague approximation to a stereotyped form of pidgin English. It would doubtless be possible to refine this further mechanically so as to approximate closer to standard English, but it is doubtful how far this would be worthwhile. Any mechanical translation must be rewritten to bring it into line with the highly idiomatic structure of standard English, and this rewriting can be done as well from a word for word translation as from a smoother redaction. Provided preconceived ideas of word order are set aside, there should be no special difficulty in anyone understanding a word for word translation of a normal prose text. It is well to bear in mind too that many languages dispense with such refinements as grammatical number, the article, and frequently in Japanese, even with a grammatical subject at the same time without loss in intelligibility. Syntactical intricacies, though of stylistic importance, contribute far less to the understanding of a text than the mere succession of significant words, which, it is felt, should be the prime concern of a translating machine.

One minor point requires mention. A mechanical translator, like the

Sorcerer's Apprentice, is unable to desist. It will continue to translate even when not required, as for example, when it encounters proper names. The context will almost certainly prevent misunderstanding, but the reader must be prepared for Tours to come out as turn/tower *(plural)* and for Mr. Kondo to appear as Mr. near wistaria.

II. Some specimen translations

Samples are given in this section of the sorts of translation to be expected from the application of the above-mentioned procedures to specific languages. The sentences chosen have been selected at random from the biological literature in these languages, merely avoiding sentences with proper names or numerical data. For convenience of type-setting, the samples are taken only from languages normally written in Roman script, but sentences from two oriental languages, Arabic and Japanese, have been transliterated to illustrate further points.

To illustrate the relative unimportance of syntax, many near-vacuous words, in particular the articles, have been treated as vacuous, and no indication is given of grammatical forms indicating the singular number, third person, or present tense. The resultant translation thus combines an English vocabulary with a grammar or lack of grammar more characteristic of Chinese. It is probable that a somewhat less extreme treatment of near-vacuous terms would be more useful in practice.

The grammatical abbreviations used are as follows:-

a	accusative
d	dative
f	future
g	genitive
l	locative
m	multiple, plural or dual
n	nominative
o	oblique
p	past
q	passive
r	partitive
s	subjunctive
v	vacuous
z	unspecific

The asterisk symbol (*) denotes the locus of decomposition.

Albanian

Hardhi*ja hyn në çlodh*je në vjesht*ë nga shkak*u i temperatur*ave të ulët*a.

vine z enter in rest z in autumn/harvest z from/whence reason z v temperature op v low z.

The vine becomes dormant in autumn because of the drop in temperature.

Danish

Sam*arbejde*t mellem de land*økonomisk*e Forening*er og Dansk*e Landbo*forening*ers Frø*forsyning er fortsat efter samme Retningslinie*r som i foregaaend*e Aar.

together work z between m country economic z union m and Danish z rural-dweller union mg seed/frog supply is continue p after same line m which/as in/you previous year.

Cooperative work between the Rural Economy Association and the Seed Supply Service of the Danish Rural Unions continued along the same lines as in the previous year.

Dutch

De ziekte treed*t dus zeer hevig op en heeft in vele gevall*en een totale mis*oogst ten ge*volg*e.

v disease come z thus very rapid up and has in many case z a/one total amiss crop then p follow z.
The disease thus appears very rapidly, and a total crop failure has then followed in many cases.

Finnish
Muut neljä ulkoma*ista kantaa ovat osoittautu*neet viljelys*arvoltaa*n kovin epävarmo*iksi.
other m four foreign country (out of) standpoint r/standard r/bear are show oneself pm cultivation value g/a very insecure (become).
The other four standard varieties from abroad have proved very unreliable as regards cultivation value.

French
Il n'est pas étonn*ant de constat*er que les hormone*s de croissance ag*issent sur certain*es espèce*s, alors qu'elles sont in*opér*antes sur d'autre*s, si l'on song*e à la grand*e spécificité de ces substance*s.
v not is not/step astonish v of establish v that/which? v hormone m of growth act m on certain m species m, then that/which? v not operate m on of other m if v one dream/consider z to v great v specificity of those substance m.
It is not surprising to learn that growth hormones may act on certain species while having no effect on others, when one remembers the narrow specificity of these substances.

German
Wenn in ein*em gröss*eren Gebiet zwei Form*en neben*einander leb*en, ohne sich zu vermisch*en, so gehör*en sie verschieden*en Form*en*kreis*en an.
if in a/one d large (more) area two form m beside one another live z without self to/too mix z, so belong/hear p z z different m form m circle m at.
If, in a largish area, two forms coexist without intermixture, they will belong to different form-cycles.

Hungarian
Az apró*bogyó*jú fajták, úgy termés*mennyiség*ben, mint száraz*anyaghozam*ban, felülmúl*ják a nagy*bogyó*jú fajták*at.
v small berry v variety m so crop/fruit quantity in as dry matter yield in surpass m v great berry v variety m a.
The varieties with small berries surpass those with large berries both in fruit yield and dry matter content.

Indonesian
Pe*semai*an*^{2} kadang^{2} demikian hebat di*rusak*kan*nja, schingga harus di*semai sekali lagi.
(causative) sow v m sometimes thus enormous v damage vv, till/so that ought v sow once more/also.
The sowings are sometimes so seriously damaged that it is necessary to sow a second time.

Italian
E' stato prov*ato che i cereal*i d'invern*o cresc*iuti in serra mostr*ano poc*a resistenza al freddo, mentre gli stessi cresc*iuti in campo apert*o, sono molt*o più resistent*i.
is been/status prove p that/which? v cereal m of winter z grow pm in mountain/crowd/greenhouse show m little v resistance to cold while v same m/ /is ps grown pm in field open v are much v more resistant m.
It has been proved that winter cereals grown under glass show little resistance to cold, while those grown in the open are much more resistant.

Latin

Possibil*e est, at non expert*um, omn*es speci*es ejusdem generis ab eadem speci*e ort*um trax*isse.

possible z is however not prove/lacking z all m species appearance same g genus/son-in-law z from same z species/appearance o arise z draw p.

It is possible, though not proved, that all species of the same genus have been derived from the same species.

Latvian

Tomēr var*am jau noteikt*i run*āt par mūsu up*ju ozol*mež*u tipa plaš*aku izplat*ību agrāk*os laik*os, kā arī par to, ka šo mež*u iznīksan*ai, vismaz pa daļ*ai, pamat*ā bijuši klimatisk*i iemesl*i.

however is able we already fixed z speak v concerning our river z oak forest z type z extensive (more) z spread z earlier lm time lm as also concerning this z that this z forest z dying out d at least through part d basis l been climatic z reason m.

We can conclude however that the oak forests of our rivers were more widely distributed in former times, and that the dying out of the forests has been due, at least in part, to climatic causes.

Norwegian

Avling*ene av høst*hvete var mere variabl*e fra år til år enn avling*ene av vår*hvete.

growth m of autumn wheat was/wary more variable m from year to year than growth m of spring wheat.

Winter wheat crops varied more from year to year than spring wheat crops.

Polish

Kierunk*i wygię*ć strzał odpowiada*ją kierunk*om panując*ych wiatr*ów i należy sądzi*ć że szablast*ość jest powodowa*na przez wiatr*y.

direction m bend v shot/trunk answer m direction dm dominant om wind gm and it behoves judge v that sword-shaped (abstract noun) is cause pq through wind m.

The direction in which the trunks are bent corresponds with the directions of the prevailing winds, and it may be concluded that the sword-like form is caused by the winds.

Portuguese

A existência de um número variável de semente*s dentro do fruto indic*a que os vari*os óvulo*s desta planta têm idêntic*a possibil*idade de se desenvolv*er.

the/to existence of a/one number variable of seed m within of fruit show z that/which? v various m ovule m of this plant has identical v possible (abstract noun) of self develop v.

The existence of a variable number of seeds within the fruit shows that each of the ovules of this plant has an equal potentiality for development.

Rumanian

Cromozom*il orzur*ilor cultivat*e sunt de un calibru mai mar*e decât cei ai orzur*ilor sălbatic*e.

chromosome m barley g m cultivated z are of a/one diameter more great z than those v barley gp wild z.

The chromosomes of cultivated barleys are of greater diameter than those of wild barleys.

Spanish

El estudio de la distribución de las temperatura*s mínim*as annual*es, como es obvi*o en tod*o trabajo, supedit*a su justeza a la densidad de las estacion*es y al record de observacion*es de cada una de ellas.

v study of v distribution of v temperature m minimum m annual m as is obvious v in all v work, reduce v v justification to v density of v station

m and to record of observation m of each a/one of v.
The study of the distribution of the minimum annual temperatures, depends, as is obvious for all work, on the density of the stations and on the record of observations of each one of them.

Swedish
Om jord*en varit tjäl*ad länge och djupt, har ingen skadegörelse av klöver*röt*an erhåll*its.
round/if earth v been freeze p long and deep, has no injury of clover rot v get pq.
If the earth has been frozen for a long time and to a considerable depth, no injury due to clover rot has resulted.

Turkish
Bütün asidite bakım*ından daima zengin ol*mayan şarap*larımız için malik asid bölünmesi keyfiyet*i arzu*ya şayan değildir.
entire acidity view v (from) always rich is/become (not) wine m our because malic acid decomposition condition a/v desire d suitable is not.
Since our wines are not always satisfactory from the point of view of total acidity, the decomposition of malic acid is not desirable.

Arabic
The following sample has been transliterated without indicating the short vowels, which are not given in the original. No attempt has been made to differentiate among the several different Arabic consonants corresponding roughly to such English letters as t, h and s, nor is hamza indicated apart from its bearer. Round brackets are used to indicate combinations of English letters conventionally used to represent the sounds of single Arabic letters. All the distinctions implicit in the Arabic text will, of course, be differentiated by the mechanical translating techniques used.
wy*hd(th) ahyan*a tjzty ll*sb(gh)y*at al*hdd*t, w*(dh)lk a(th)na al*anksam al*mnwy al*a(kh)yr, ntyj*t l*hwad(th) mytwzy*t.
and occur time m z division d chromosome m v limited z and that period v division v sperm v last result z d occurrence m mitotic z.
Fragmentation of the limited chromosomes takes place occasionally as a result of mitotic disturbances during the last spermatogenial division.

Japanese
Japanese texts are written in a mixture of three alphabets, namely Chinese ideographs, the hiragana syllabics, and the katakana syllabics. In the sample appended, the Chinese ideographs are put into capitals and the hiragana syllabics into lower-case letters. Katakana syllabics did not occur. Most of the Chinese ideographs have at least two alternative phonetic readings in Japanese. For purposes of mechanical translation, it is irrelevant which is correct, and for uniformity the on reading is used throughout. All phonetic irregularities, such as the use of the syllabic tsu for representing doubled consonants, have been ignored as semantically irrelevant. The sentence given is a word block with no punctuation to assist in decomposition.
ko-no-*TAI-*KAN-*SEI-*no-*SA-I-*ha-*SHI-BAI-TAI-*ga-*NI-BAI-TAI-*ni-*HI-*shi-*SHIN-TO-ATSU-*gu-*KO-*ku-*na-tsu-te-*wi-ru-*ko-to-*mo-*DA-ki-*na-*GEN-IN-*to-*KO-i-*ra-re-ru.
this endure cold sex/disposition g difference (as for) tetraploid n diploid (at) sort/compare do/also osmotic pressure n high (adverb) becoming is fact v large v reason with/when consider q.
It is thought that the resistance to low temperature of the tetraploids when compared with the diploids may be largely due to the higher osmotic pressure of the former.

III. Illustrative schedules of mechanical translation

It has been shown, in sections I and II of this paper how a dictionary translation can be made from any "foreign" language into the base language, and also that such a translation may be a useful one. In this section some of the actual machinery whereby such a process can be realized will be discussed and estimates will be given to show the expected speed of operation of these devices.

In the first place it can be stated that two classes of machine are already in existence which can perform the required operations. These, are firstly the standard punched card equipment currently used for business purposes, and. secondly almost any of the available all-purpose automatic digital computers. An unfortunate limitation of the latter class of device lies in the fact that none has, at present, the storage capacity to make it worthwhile as a translator. On the other hand, as will appear later, the former class of machine has adequate storage capacity, but is limited in speed.

Before proceeding to a detailed examination of the above alternatives, it is perhaps worth while setting down in a general way what would be the ideal performance of a translating device of the dictionary type. It would appear that such a device should have a keyboard input and a separate output typer and should be of such a speed that as soon as a particular word has been typed in, the output printer would commence the writing of its translation. This assumes, of course, that the input word does not form a part of a possible multiple semantic unit, in which case the machine would either await the conclusion of the phrase, or alternatively output some such indication as (incomplete information) to assure the user that the hiatus was not due to internal malfunction of the machine.

Since the procedure developed in previous sections is tentative, it is not proposed to exhibit a complete scheme of operation for its mechanization. Instead, however, the limited problem of single stem and ending translation will be examined, and it will be at once apparent that any more complex word or phrase unit can be handled by a straightforward application of the same principles.

As a first example, the use of punched card machinery will be considered. It will be assumed that the following equipment is available:

1. Hand card punch.
2. Sorter.
3. Collator.
4. Reproducer (gang punch).
5. Printer.

In fact, it is not necessary to have all of the above equipment. For example either the sorter or the collator may be omitted. Likewise a tabulator may replace the collator. The effect of such changes is mainly on speed, and the optimum speed occurs when the full range is available.

The operation of translation is then as follows:

1. Have available: (a) Stem dictionary pack
   (b) Ending dictionary pack
   both sorted into alphabetical order
2. Punch text to be translated on to blank cards together with a number giving word position in original message.
3. Sort text pack into alphabetical order.
4. Using collator, mesh sorted text pack with stem dictionary pack (a) arranging connection with reproducer so that translation field of dictionary pack, word order number from text card, and residual ending of text word are punched on blank card. Collator feeds are arranged so that text and dictionary cards go to separate hoppers and so do not need subsequent separation and resorting.
5. Mesh partial translation cards obtained in (4) with alphabetic-ending pack (b) exactly as in (4). This will now produce a pack containing:

   Stem translation.

Ending translation.
Original word-order mark.

6. Sort pack obtained in (5) into numerical order.

7. Reproduce card contents (excluding order number) on printer.

It is to be observed that with the multiple word units in for example, German, several repetitions of operations (4) and (5) may have to be made. A technical point of collator design also requires mention: It is necessary in the above scheme to arrange the residual ending punching resulting from (4) so that it occurs starting in a fixed position. This can be done by means of suitable pyramid selection circuits.

When a collator is not available, the following scheme will produce the same results as those given above, but at the expense of rather longer operating time (vide infra):

1 and 2. as above.

3. Sort text pack and stem dictionary pack (a) into alphabetical order.

4. Pass pack resulting from (3) through tabulator, arranging control change so that output punch is stimulated only when a text punching occurs, in which event preceding card field (i.e., translation) is punched or output, together with order punching and suitably shifted ending from text card.

5. Repeat operations (3) and (4), this time using ending dictionary pack (b) and partial translation pack resulting from (4).

6 and 7. as in previous scheme.

8. Pass meshed text and dictionary (a) cards from (4) through sorter operating on message order column. This separates the two packs, and the (a) cards having no order punching arc recovered from the reject box in original alphabetical order.

9. Same as (8), but on (b) cards from (5).

The defect of this method lies in the fact that time is wasted in the complete dictionary sorts in (3) and (5) and also in the separation operations (8) and (9).

To exhibit more clearly the relative merits of the above processes, and at the same time to give some idea of the general efficiency, a concrete example will now be considered.

Let stem dictionary (a) contain 5,000 cards.
... ending ... (b) ... 5,000 ...
... message consist of 1000 words of maximum length 10 letters.
Operating speeds are taken to be:

1) Sorting 450 cards/min.
2) Collating 240 cards/min.
3) Tabulating and/or punching 100 cards/min.
4) Text punching 100 words/min.

Then the times of operation are:

	Scheme I	Scheme II
1)	--	--
2)	10 mins.	10 mins.
3)	22 mins.	132 mins.
4)	25 mins.	60 mins.
5)	25 mins.	192 mins.
6)	22 mins.	22 mins.
7)	10 mins.	10 mins.
8)	--	14 mins.
9)	--	14 mins.
TOTAL	1 hr. 54 mins.	7 hr. 84 mins.

It is thus seen that the latter scheme is much inferior. This is wholly on account of the redundant sorting operations which have to be performed on the dictionary packs, and a material improvement results if a shorter dictionary of, say, 1000 stem and 1000 ending words is acceptable.

Viewed in competition with a skilled human translator, even the first figure of approximately 2 hours is not impressive, since the human

translator could probably make a much better translation in only slightly more than one half the time. On the other hand it must be remembered that the above process requires no skilled linguist and has the advantage that the machinery will deal equally well with any number of languages; this is particularly valuable since so many multilingual humans appear to have a marked antipathy to the sciences and an almost complete absence of scientific vocabulary in both their acquired and their parent tongues.

Next consider the use of an automatic digital computer for the purpose of translation. Here a new feature is at once evident, namely, the possession of a high speed storage device or memory. In available machines this has a capacity of 1000-4000 "words," each of about 30-40 binary digits. In many cases this storage is of a transient type and would require filling from an auxiliary medium (e.g. punched cards) at the start of each translation from a different language. It is also true to say that, at present, all forms of storage (used in existing computers at least) except one, are not absolutely reliable over periods of more than about 30 minutes.

The exception is the so-called magnetic drum which consists of a rotating drum coated with a magnetic material on which binary data can be stored as elementary magnetized regions of either N-S or S-N polarity. These regions may be erased and reused at will and are read by means of the electric currents which they induce in a pick-up device usually known as a head. Volumetrically, such storage is exceedingly efficient; in a typical computer, for example, 6000 binary digits can be stored upon one square inch of surface. This compares very favorably with the 1,000 binary digits that can be stored on the 23 square inch surface of a standard punched card.

The most obvious way of using a computing machine for translation is to code each letter of a proposed dictionary word in 5 binary digit form and to store the translation in that memory location having the same digital value as the aggregate value of the dictionary word. Translation would then consist in coding the message word (a teletyper does this automatically) and then extracting the translation in the same or next lower, digitally valued, storage position, the first case giving exact translation and the second the stem translation with a remainder. The stem translation would be at once printed at the output station and the undecodable remainder shifted to occupy the extreme left hand position of the word storage register. When this had been done, the number remaining would be used to refer to the ending dictionary whose entry would again be typed at the output. In the event of the word forming part of a multiple semantic unit the whole would be sent to an auxiliary storage position to await receipt of its remaining part or parts, an indication being meanwhile typed.

Despite its simplicity, this scheme is quite impracticable with any existing storage device because the naive process of placing each dictionary word in a storage position equivalent to its own binary numerical version is grossly redundant. For example, 5000 words should require a storage capacity of only 5000 positions, whereas if the maximum word length is 10 letters, in straight binary code any number of up 2^{50} might represent a word (although in fact only 5000 or approximately 2^{12} would do so) so that an available storage capacity of about 10^{15} words would be required!

To overcome this difficulty, a slightly more sophisticated approach has to be adopted. Dictionary words are stored in alphabetical order and in consecutive memory locations. Possibly these might be broken up into initial letter groups such that each group starts in a storage location indicated by the binary coded form of its initial; this would increase the speed of operation by a considerable factor without greatly adding to the storage requirements. The coded word to be translated is sent to an arithmetic register and is there compared, by subtraction, with the dictionary words. In this way, by means of the normal conditional control transfer order, the dictionary stem corresponding to the word can be

obtained and its translation printed. Next the remainder of the message word left after subtraction of the stem, is shifted to occupy the extreme left hand register position, and the comparison process is repeated, this time with the words of the ending dictionary. This process would lead to a further typed output giving the required grammatical information, after which the machine would be ready to receive the next message word for translation. Parts of multiple semantic units would be stored for future treatment exactly as in the more elementary scheme outlined above.

To complete the discussion, it is only necessary to make some remarks about times of operation of general purpose computing machines. The relevant feature so far as the second machine treatment is concerned, is the time of a single subtraction, and this can be taken conservatively as .001 second. Thus, even assuming a complete dictionary search over 10,000 words to be necessary the total translation time would be a maximum of 10 seconds and a mean of 5 seconds. To this must be added the typing time which, for 40 letters, would be about 6 seconds. It would thus be uneconomical to reduce the translation time by a factor of more than about 5 and this could be attained easily by means of the initial letter group storage technique outlined above.

For comparison with the punched-card examples previously considered the total time of translation of a 1000 word message, with a 5000 + 5000 word dictionary, would now be about 2¼ hours so no gain results from the use of an electronic machine with existing teletype output facilities. On the other hand, at least one computer is equipped with punched card printer output and with this, the translation and printing time of the same message could certainly be reduced to less than 30 minutes, and very probably to 15 minutes.