L. BRANDWOOD

The Translation of a Foreign Language by Machine

Mr. L. Brandwood is a Nuffield Fellow in Machine Translation at the Computational Laboratory of Birkbeck College. This paper is intended as a sequel to that of Dr. Booth. It describes the present state of research, and explains in greater detail how the basic grammatical problems facing the would-be translator from French can be overcome.

In the near future an attempt to translate from a foreign language by machine will be made at the computational laboratory of Birkbeck College, which will differ from previous experiments in that the sentences or passages of French to be translated will not have been specially chosen or "doctored" in any way beforehand: on the contrary, they will be constructed by French scholars, who will be invited to do their best to fault the machine, What follows is an account of the method, or program, which makes such an experiment possible.

First of all, however, I think a brief explanation of the general procedure in translating by electronic computer will not be out of place. To a layman like myself the information that such a computer works in terms of electronic impulses conveys very little, but as I suspect that the majority of the readers are in the same boat, I shall no doubt be forgiven for using terms which though strictly not correct have at least the advantage of being more readily intelligible.

We may say then that the computer works, like any other calculating machine, with digits. It cannot recognise letters; at any rate not as they stand. In order that it may recognise them the alphabet has first to be coded, that it to say, each letter represented by a particular number. Since a word is merely a collection of letters and we have a number for each letter, we can now put any words we choose into the computer to be stored on a magnetic drum in terms of impulses. These constitute the machine's "dictionary", which may be looked upon as exactly like an ordinary French-English dictionary, but with the numbers representing the French words arranged in order of either ascending or descending magnitude. This is important for the rapid identification of the words in the text being translated, which takes place as follows. The machine identifies an incoming word by comparing the number which represents it with the numbers contained in its dictionary. The comparison takes the form of a simple subtraction. Starting with a number at the middle of the dictionary it subtracts from this the number of the word being translated: if the result is zero, the word has been identified and the English equivalent, which is marked alongside the French can be output; if the result is not zero, then the number sought lies in one or other of the two halves of the thus divided dictionary according to whether the result is positive or negative. So, taking the number at the middle of the relevant half, the machine carries out the same comparison as before, and if — as is more than likely — this too proves unsuccessful, it continues halving the numbers that remain until the one required is located.

This ingenious process of elimination, which was devised by Dr. A. D. Booth, ensures the speedy identification of words even in a very large dictionary. In fact the size of the latter can be greatly increased without appreciably lengthening the look-up process. To be precise, the number of comparisons required is doubled as the number of words in the dictionary is squared. This means that since on an average seven comparisons are required to identify any word in a dictionary of one hundred, only fourteen will be necessary to cope with a dictionary of ten thousand or twenty-eight to cope with one of a hundred million.

The capabilities of the look-up method might seem rather wasted when it is revealed that the maximum storage capacity of the Birkbeck computer is at present only 250 dictionary items (i.e. 250

foreign words each with one English equivalent), though this will shortly be increased to 2000. It must be borne in mind, however, that this is not the limit of all computers, even of those existing, and storage capacities will continue to grow larger. But that lies in the future, and for the time being all our attempts at translation have to be scaled down and done, so to speak, in miniature.

The machine's capacity allows us 250 dictionary items. We cannot use all of these for the actual dictionary, however, because we wish to include a few elementary syntactical instructions to enable the machine to produce something a little better than a word-for-word translation. We did not know beforehand in what proportion the dictionary space should he divided between vocabulary and grammar, but reckoned that with a vocabulary of about a hundred words there would be enough space left for the instructions in mind and a little over for subsequent modifications. What was wanted therefore, was a vocabulary which, though very small, would yet be capable of producing good French sentences in a reasonable variety; in short the proportion of the various parts of speech which it contained had to be correctly balanced. This was achieved in a very simple fashion by selecting a passage from a book and extracting the first hundred different words: to these were added any missing forms of the personal, possessive and relative pronouns encountered. The result was a vocabulary consisting of 25 verbs, 2 adverbs, 23 nouns, 10 adjectives, 3 numerals, 8 prepositions, 6 conjunctions, all forms of the unstressed personal pronoun (except *en* and *y*), all forms of the simple possessive pronoun, the relatives *qui, que, dont,* the definite and indefinite articles and the negative *ne . . pas.* Volunteers will be presented with this vocabulary and asked to construct sentences or short passages for the machine to translate. They will be permitted to use the singular and plural forms of all nouns and adjectives, the masculine and feminine of the adjectives, and all parts of the verb except the 1st person imperative and the subjunctive mood. The former is excluded because it requires a rather periphrastic translation, which combined with its comparative rarity makes it hardly worth bothering about, the latter because many of its forms are identical with those of the present and imperfect indicative, and could only be distinguished from these by taking into account the syntactical context. This, however, is out of the question with the present storage capacity. Nevertheless, disregarding the compound tenses, which are accounted for by the separate consideration of *avoir* or *être* and the perfect participle of each verb, there still remain about 30 different forms to each regular verb, the actual figures for the four conjugations being: *donner* 33, *finir* 29, *vendre* 32, *recevoir* 31. For an irregular verb, unless it is defective, the number is usually a little higher. In short the 25 verbs (13 regular, 12 irregular) produce well over 800 different forms, all of which the machine must be able to translate. Obviously such a large number will not fit into our machine's dictionary as they stand. How then are they to be incorporated? The solution thought of as long ago as 1947 by Booth and Richens, is that a word should be divided into a stem and an ending.

Here we must pause a while in order to define our terms. *A stem is not necessarily a stem in the sense that grammarians know it.* Generally speaking, it is the longest part of a word common to all forms (inflections) of that word. The stem of *donner,* for example, is *donn-;* what remains of each inflection of *donner* after this has been subtracted constitutes an ending. This, however, is not a universal rule, since there are instances where a single stem would account for all forms of the verb, but more than one is chosen for the sake of utility (e. g. the stem *acqu-* would cover all forms of *acquerir,* but in addition *acquier-* has to be used in order to comply with the general system covering all verbs). Utility might in fact be said to be the only principle of grammar in mechanical translation.

Verbs are divided, then, into stems and endings, and the following results cover each of the four regular conjugations:

	Pres.	Fut.	Condi.,	Imperf.	Past Hist.	Perf. Part.	Inf.	Pres.Part.	Imp.
1. donn-									
1st	e	rai	rais	ais	ai				
2nd	es	ras	rais	ais	as	é		ant	e
3rd	e	ra	rait	ait	a	ée	er	(ante)	e
1st	ons	rons	rions	ions	ames	és		(ants)	
2nd	ez	rez	riez	iez	ates	ées		(antes)	ez
3rd	ent	ront	raient	aient	erent				ent
2. fin-									
1st	is	irai	irais	issais	is	i		issant	is
2nd	is	iras	irais	issais	is	ie			isse
						is		(issante)	issez
3rd	it	ira	irait	issait	it	is	ir	(issantes)	
1st	issons	irons	irions	issions	imes			(issants)	
2nd	issez	irez	iriez	issiez	ites				
3rd	issent	iront	iraient	issaient	irent				isssent
3. rec-									
1st	ois	evrai	evrais	evais	us	u		evant	
2nd	ois	evras	evrais	evais	us				ois
3rd	oit	evra	evrait	evait	ut	ue	evoir	(evante)	oive
1st	evons	evrons	evrions	evions	umes	us		(evants)	
2nd	evez	evrez	evriez	eviez	utes	ues		(evants)	evez
3rd	oivent	evront	evraient	evaient	urent			(evantes)	oivent
4. vend-									
1st	s	rai	rais	ais	is	u		ant	
2nd	s	ras	rais	ais	is				s
3rd	—	ra	rait	ait	it	ue		(ante)	
1st	ons	rons	rions	ions	imes	us			
2nd	ez	rez	riez	iez	ites	ues		(ants)	ez
3rd	ent	ront	raient	aient	irent			(antes)	ent

If now the French verb stems and their English translations are entered into the machine's dictionary together with an instruction that each word fed in for translation is to be identified with the stem in the dictionary which it most closely resembles, we will be supplied with the basic meaning of every verb in the passage to be translated. The tense, however, is still lacking and for this we have to refer to the endings. These too need to be included in the dictionary, though at the same time they must be kept apart and distinct from the stems; otherwise confusions will arise (e. g. between the singular present indicative of *avoir* and the endings denoting the singular past historic of the 1st conjugation). Nor should they be included as they appear above for two reasons: Firstly the use of different inflections to indicate difference in person and number is on the whole a linguistic redundancy, so that where two identical endings occur we need in fact make only one entry in the dictionary: secondly, in as much as the endings are used merely to distinguish tense, there is no necessity to enter the whole ending but only sufficient to enable this distinction to be made (e. g. the *-rai* part of the ending *-evrai* in *recevoir* is enough to indicate that the tense is future). These distinctive endings are 46 in number, i. e.

Pres.	Fut.	Codit.	Imperf.	Past Hist.	Perf. Part.	Inf.	Pres. Part.	Imp.
e	F*ut.*	rais	ais	ai	é	r	ant	
s	rai	rait	ait	as	ée	re		
z	ras	rions	ions	a	és			
t	ra	riez	iez	mes	ées			
ois	rons	raient	aient	tes	i			
oit	rez			rent	ie			
	ront			is	(is)			
stem				it	ies			
				us	u			
(is)				ut	ue			
					(us)			
(it)					ues			

One or two explanatory remarks are necessary here. The "stem" means that there is no ending (e. g. *il vend).* The feminine and plural forms of the present participle are omitted because they belong strictly speaking to its adjectival function. The forms in parentheses, though characteristic of the

tense under which they are marked, are *also* found in another tense of which it is preferable to re-gard them as primarily characteristic. This is because the limited capacity of the machine prevents the inclusion of instructions (assuming that they could be formulated) which would enable it to dis-tinguish these formally identical tenses; consequently for the time being one translation must be made to serve for both, and it is naturally desirable that this translation should be determined by the tense in which this disputed form has the wider occurrence. The ending -is, for example, represents the present indicative in *finir* type verbs, the past historic in *finir, vendre* and *sentir* type verbs. Therefore we take it as characteristic primarily of the latter tense; it is just a question again of utility.

Clearly, however, the fewer there are of these identical tense forms, the lesser the confusion will be. Hence the stem *fin-* instead of *fini-* for the *finir* type verb in the present scheme. If the stem had been *fini-* the following difficulties would have arisen:

1.	1st, 2nd, 3rd pers. sing. pres. indic.	=	1st, 2nd, 3rd pers. sing. past historic	(ending -s)
2.	1st, 2nd, pen. sing. pres. indic. and past historic	=	perf. particip. masc. plural	
3.	Perf. Particip. masc. sing.	=	3rd pers. sing. pres. indic. of *vendre* type verbs	(ending = stem)
4.	Perf. particip. fem. sing. and plural	=	1st and 2nd pers. sing. respectively pres. indic. *donner* type verbs (endings e and es)	

As it is, with the stem *fin-,* only the first two of these are encountered. *Fin-,* therefore, is the best stem, but only, it must be repeated, in the present scheme. As the dictionary is enlarged and takes in more irregular verbs we may find that there are ambiguities which can be solved by making the stem *fini-* instead of *fin-* (e. g. the verb *saisir* will have a stem identical in form to the 1st and 2nd pers. sing. pres. pres. indic. of *savoir,* unless the 'i' is added). The important thing to remember is that there can be no one optimum system for all translation programs: their general principles may very well be the same, but the details must always vary according to the size of the dictionary and the nature of its contents. This applies particularly to stems of verbs, which will continually have to be modified in order to avoid ambiguities (e.g. 1st, 2nd, 3rd pers. sing. pres. indic. of *vivre* 1st. 2nd. 3rd. pers. sing, past historic of *voir*, which in its 1st and 2nd persons may again be confused with the stem of the verb *viser).* Since correspondences of this type are far from being uncommon it may prove advisable in a larger dictionary to revert to the conjugational classification each with its separate procedure, in order to reduce the confusion.

Returning to our explanation of the table of endings, the reason for the forms which represent the present indicative being written as they are is to be found in the rule previously mentioned, ac-cording to which the machine identifies any incoming word or part of a word with the nearest equivalent (by which is meant the nearest shorter equivalent) in its dictionary. Therefore, if we enter, for example, -s in the dictionary as a present indicative ending, the machine will identity with it not merely the bare -s of *vendre* and *sentir* type verbs, but also the -es of the 1st and the -ons of all conjugations, simply because there is no other entry in its dictionary with which it can identify them. It is for this reason that the lst/2nd and 3rd pers. sing. pres. indic. of *recevoir* type verbs *(-ois, -oit)* has to be entered in full, since otherwise they would be identified not with the pres-ent forms -s and -t, but the past historic -is and -it.

The system described above covers not only the regular but also the irregular verbs in the major-ity of their forms. The only difference is that many of the latter require several stems instead of only one; *aller,* for example, the following: *vais, va-, vont, all-, i-.* Of course difficulties do arise, just as they do in the regular verbs, but their number is relatively small.

In the method described above the tense of a verb is discovered, after its meaning has been obtained from the stem, from the ending. This is only one way. There is another, as the reader may already have observed for himself in perusing the table of endings, namely by a particular letter in the ending instead of by the whole ending. The most obvious example is the letter '*r*', which is sufficient to indicate that the tense is either conditional or future indicative. In place of an ending dictionary, therefore, we can have a series of instructions like this:

1. If the ending ends in -*r* or -*re*, then the verb is an infinitive, and (to) should be inserted before the translation of the stem. (The parenthesis denotes that the 'to' will not always be needed to make good English in the translation e.g. in '*puis-je sortir?'*, whereas '*il va partir*' does require it. When larger storage capacities are available, it will be possible to make this distinction.)

 2. If the ending ends in -*ant,* we are dealing with the present participle and '-ing' should be added to the translation of the stem. (Again it would be quite easy with a larger storage at one's disposal to include an instruction requiring the machine to deduct the mute -e from words ending so before adding the '-ing'.)

 3. If the ending contains the letter '*r*' and this is not followed by -*ent* (in which case the tense is the past historic), then the verb in question is either future or conditional. If now the letter 'i' occurs after (but not necessarily immediately after) this '*r*' and is not the last letter of the ending, the verb is conditional, and the word 'would' must be inserted in front of the stem meaning. Otherwise it is the future tense, and the word to be inserted is 'will'. (The English idiom of 'shall' with the 1st person and the interchange of this and 'will' with a change of emphasis could likewise be included in a machine with a larger dictionary space.)

4. If the ending begins with '*e*' (not followed by '*a*'), '*s*' (not followed by '*a*' or '*i*'), '*t*' (not followed by anything), *x, o, n, l, d,* or if there is no ending (i.e. the stem is bare) and *avoir* or *être* does not precede, then the verb is in the present indicative, and the stem meaning is an adequate translation. (The same can be said of the suffixed '-s' in the 3rd pers. sing. of the English present as of the other refinements mentioned above.) If there is no ending and *avoir* or *être does* precede, the verb is in the perfect tense (see no. 5).

 The reason for *e*, *s, x, o, d, t,* being indicative of the present tense will be clear: why *n*, and *l* should be, and why '*e*' should not be followed by '*a*', '*s*' by '*a*' or '*i*', or *t* by any letter is this:

 (i)   n — enables certain irregular verbs which double the final consonant of their stem in the 3rd pers. plur. (e. g. *venir viennent* to *be comprehended under one stem* (i. e. *vien —*) instead of two.

 (ii) 1 — does the same for 1st conjugation verbs in — *eler* (e. g. *atteler — attelle).*

(iii) '*t*' not followed by another letter — since — *tes* is an ending characteristic of the past historic. (Note: this means that 1st conjugation verbs in -*eter* (e. g. *jeter-jette*) need two stems).

(iv) '*e*' not followed by '*a*' — because 1st conjugation verbs in -*ger* (e.g. *manger}* retain the '*e*' before '*a*' and '*o*,' so that but for this precaution the imperfect and past historic of these verbs would be translated as the present tense.

 (v) '*s*' not followed by '*a*' or '*i*' — as happens in the imperfect of *finir* type verbs, if the stem *fini*- is chosen instead of *fin*- (i.e. -*ssais, -ssait, -ssions, -ssiez, -ssaient.)*

 (Note: with the present method the stem *fini*- is preferable to *fin*- for this reason: if the 'i' were not included in the stem, it would come at the beginning of all forms of the present indicative, which would consequently be translated as a past historic. As it is, only the singular forms of the past historic are confused with those of the present indicative).

 5. Everything not accounted for by the previous instructions is in the past tense, and -(e)d should be added to the translation of the stem. (Note: this will result in some odd looking forms in the case of the English strong verbs, but as we have said so often before, it is only the restricted storage capacity of the present machine which forbids the incorporation of these strong forms (e.g. sang, sung) alongside the regular form (sing) together with instructions for their correct application).

These are two ways of ascertaining the tense of a verb, which we may call the stem-ending and the eliminating method respectively. The question now arises of which is the better, but this is not easily answered, depending as it does to a great extent on the machine, the size of the dictionary employed, the nature of the text to be translated and the quality of translation required. The eliminating method, for instance, must assume that the person receiving the translation will be content to have all past tenses translated simply as such and prepared to use his own common sense and judgement to distinguish where the sense requires a perfect, imperfect or past historic. If the more precise translation is desired, then there is no doubt that the stem-ending method is to be preferred.

As regards the nature of the text, if the tense in which it is written is the past, the stem-ending method will probably be quicker than the eliminating method, which arrives at this tense only at its last stage; if, however, the tense is the present and future, then it may well be that the eliminating method will prove quicker. All this smacks too much of 'perhaps' and 'maybe', however, as would any remarks which we might hazard on the amount of space which each method would occupy in the machine's dictionary. A definite answer to all such questions will after all be forth-coming when the

machine starts on its program, since this will be tried with both methods. In the meantime two facts are quite certain:

a) the *eliminating method* is superior to the stem-ending method when it comes to translating perfect participles, because it needs to consider only the first letter of the ending, which is the one distinctive of the perfect (i.e. *é, i, u,*) whereas the stem-eliding method has to have all the forms (fem. and plur.) written down for it to choose from.

b) the *stem-ending method* is more economical for treating the irregular verbs, because the fact that it can ignore the middle part of some such verbs enables it to account for them with one stem less in each case than the eliminating method. The verb *asseoir,* for example, requires three stems in the latter method i.e. *assied-* (for the sing. of the pres. ind.), *assey-* (for the imperf. and plural of the pres. indic), and *ass-* (for future, past historic and conditional), but only two in the former method, which can ignore the -ied- in the sing. pres. indic. and so use *ass-* for this too. Out of about seventy irregular verbs examined this economizing applied to seventeen, (i.e. *acquerir, boullir, tenir, sentir*, s*ortir, partir, servir, asseoir, savoir, pouvoir, écrire, joindre, craindre, mettre, battre, suivre).* To these must be added their compounds and any similarly conjugated verbs, though some of the latter are misleading because a verb with a similar stem prevents the reduction of the number of stems. For instance, the only stem that would be needed for *dormir* would be *dor-,* but for the exist-ence of a verb *dorer:* so the stem has to be *dorm-,* as for the eliminating method, and the sing. forms of the pres. indic. *(dors, dort)* have to be entered in the dictionary as they stand.

That completes our treatment of verbs, and we can pass on now to explain the program for dealing with the other parts of speech. Nouns and adjectives are taken together, since both may have feminine and plural forms, and these can be satisfactorily accounted for by the same stem-ending procedure, thus:

Type 1. If the feminine form is created merely by adding extra letters to the masculine form without involving any change in the latter, then the masculine form (so marked) is entered in the stem dictionary, the extra letters denoting the feminine form (so marked) in the ending dictionary. (e.g. *chien -ne, joli -e).*

Type 2. If the feminine form is created by first altering then adding to the masculine form, (e.g. *chameau, chamelle),* then the letters which are common to both forms (i.e. *chame-*) are entered in the stem dictionary, those peculiar to the masculine and feminine in the ending dictionary (i.e. *-au. -lle).* The gender of a noun which has only one form for both masculine and feminine forms must be determined by other meant (e.g. by reference to the definite or indefinite article or an adjective, if it is accompanied by any of these).

The plural form of nouns and adjectives, both masculine and feminine, is easily recognised — as long as there is in fact a distinct plural form to be recognised — inasmuch as the ending will have added to it either *-x* or *-s.* Apart from irregulars the only class of nouns and adjectives which does not comply with this "rule" is that ending in *-al,* which drops the "*l*" before adding *-ux* (e.g. *cheval, chevaux).* This is resolved in the same way as the Type 2 feminine form above, that is, by putting the letters common to both singular and plural (in this case *cheva-*) in the stem dictionary, the letter "*l*" which is dropped in the ending dictionary. So we have the following three types of regular nouns and adjectives:

	Stem	Ending	Plural		Stem	Ending,	Plural
Type 1	noun chien (m)	-ne (f)	-s	Type 3	noun cheva- (m)	-1 (m)	-ux
	adj. joli (m)	-e (f)	-s		adj, principa-(m)	-1 (m)	-ux
Type 2	noun chame-	-au(m)	-s			-le (f)	-s
		-lle (f)	-s				
	adj. bre-	-f (m)	-x				
		-ve(f)	-s				

116

In short, we can identify all plurals which are genuine forms and not identical with the singular by the three suffixes *-s, -x, -ux.* The only difficulty is in getting the machine to recognise where these suffixes occur after the ending and where directly after the stem. The latter is true of the masculine forms of Type 1 nouns and adjectives and masculine adjectives and all nouns of Type 3, the former of Type 2 nouns and adjs. and Type 3 feminine adjectives. It can be done quite easily, if the machine proceeds according to the following instructions, which do not even require it to distinguish between the three suffixes:

    1. The machine looks in its stem dictionary. If:

        a)   the incoming noun or adjective can be identified exactly with one of the stems, then it is a Type 1 masculine noun/adjective and can be translated immediately.

        b)   the incoming noun or adjective cannot be identified exactly with a stem, because an extra letter or letters is left over, then it cannot be translated immediately, and

    2. the machine looks in its ending dictionary. If

        a)   the extra letter(s) can now be identified exactly with any ending contained in the dictionary, then the incoming word is either a feminine singular noun/adjective of Type I, a masculine or feminine or singular noun/adjective of Type II, or a masculine or feminine singular noun/adjective of Type III, and can be translated accordingly (see below)

        b)   the extra letter(s) can be identified with an ending contained in the dictionary, but if there is still an extra letter left over, then the incoming noun/adjective is the plural of the stem plus ending word and its gender that marked after the ending. The plural is denoted by adding -s to the English translation.

        c)   the extra letters) cannot be identified with an ending contained in the dictionary, then the incoming noun/adjective is the plural of the stem word and its gender that marked after the stem. (Note: the identification of number in adjectives is, of course, irrelevant for their actual translation, but as was noted above, it can be useful for determining the gender and number of nouns with identical forms for the masculine and feminine or the singular and plural).

This scheme will provide rapid identification of all regular adjectives and nouns, but it has the fault of being somewhat uneconomical as regards storage space, since the ending dictionary needs to be of the same size as the stem dictionary, and in fact larger, most stems having one and often two endings. A possible modification offers itself, however, when we realise that the number of different endings for regular adjectives and nouns is only 15 (5 masculine and 10 feminine) i.e. masculine: *-eur,-au, -l, -f, -x;* feminine: *-euse, -rire, -lle, -ne, -te, -sse, -e, -ve, -se.* It would seem feasible, therefore, to have an ending dictionary consisting of only these 14 endings which would be capable of producing a correct translation of all regular French adjectives and nouns. And, in fact, it would be, but for English often having a completely different where French has only a partially different feminine form of a noun (e.g. *fils-fille*, son-daughter), or an irregular plural where French has a regular one (e.g. *homme — s,* man — men). A reasonable solution is to divide all nouns into two classes.

    a) those which have a separate word for masculine and feminine (e.g. *cheval — jument),* those which are without gender in English (e.g. *le village,* abstract nouns), — all, in fact, which do not require a distinction of gender in translation, (consequently all adjectives can be included in this class),

    b) those which have a feminine or plural form that does require a different translation from the masculine or singular form.

The first class, in which the ending is required only for identification not translation, can be dealt with in the manner suggested above, that is by having an ending dictionary for the whole class of only 14 distinct endings (or 28, if we wish to include the plurals and so get immediate recognition of these, rather than provide the machine with instructions). This will retain the speed of identification while at the same time improving the economy.

For the second class the unmodified system will be retained, and each noun will be accompanied by its feminine and plural endings, alongside each of which will be its appropriate translation where different from that of the masculine singular. When the feminine form is merely a lengthened

form of the masculine, the two translations will appear with the stem and ending respectively, for example, stem: *chien* = dog, ending *-ne* = bitch (dog). When it involves a change in the masculine form, the two translations will accompany the two endings, thus: stem: *act-* (no translation), endings: *-eur* = actor, *-rice* = actress.

In closing the section on nouns and adjectives the observation made previously with verbs must be repeated, namely that if the stem-ending method anywhere involves a confusion of two words with different meanings, the stems of those words should be altered, where possible, to create a distinction. Where it is not possible, the offending forms will have to be regarded as irregular and entered in the dictionary in full.

The only other part of speech in the program which requires separate instructions for its identification is the pronoun, to be more explicit, the personal (unstressed) and possessive pronoun. Four difficulties arise, namely those of distinguishing the accusative of the 3rd personal pronoun *(le, la, les)* from the definite article, the dative plural of the 3rd personal pronoun (*leur*) from the identical possessive form, and the nominative *nous* and *vous* from the accus./dative, and lastly that of removing the oblique forms of the personal pronoun from their French position before the verb to their normal English position after the verb. The first two difficulties are surmounted by assuming that whenever *le, la, l', les* are followed by a pronoun or verb, they are forms of the personal pronoun and not the definite article, and similarly that wherever *leur* is followed by a verb it is the personal pronoun. The second two are also solved simultaneously in the following way:

Type 1. the sequence is — pronoun^1 pronoun^2 pronoun^3 verb if *nous* or *vous* is the first pronoun, then it is nominative translate in order — pronoun^1 verb pronoun^2 pronoun^3 (e.g. *nous le leur donnons* — we give it them)

Type 2. the sequence is — pronoun^1 pronoun^2 verb

a) if pronoun^1 is a nominative form *(je, tu, il, elle, ils, elles)*, pronoun^2 is oblique. Therefore translate in order: pronoun^1 verb pronoun^2 e.g. *je le donne* = I give it)

b) if pronoun^1 is oblique, pronoun^2 is also oblique. Translate in order: verb pronoun^1 pronoun^2 (e.g. *le berger le lui donne* = the shepherd gives it him)

e) if pronoun^1 is a form which is not recognisable per se as nominative or oblique (i.e. *nous* or *vous*, then the verb must be examined to see whether it agrees with *nous* or *vous*.

Thus: 1) if the verb ends in *-ons* or *-mes*.
    nous is nominative ⎫ Translate in order
    2) if the verb ends in *-z* or *-tes*. ⎬ given in a)
    vous is nominative ⎭

Otherwise *nous* and *vous* are oblique and the translation order is that of b).

Type 3. the sequence is — pronoun verb. The procedure is exactly the same as for Type 2.

This, then, is the program which will be used for translating from French mechanically. The objection might well now be raised that it is devoted only to the grammatical side, the parsing of words, and gives no attention to syntax. Apart from the fact that an instruction will, as it happens, be included to reverse in translation the order of adjectives following their nouns in French, this objection is quite valid. In defence we must make the plea made so often already that the restrictions imposed by the present limited storage capacity of the computer renders any comprehensible scheme which might be drawn up for dealing with syntactical problems redundant. When the storage capacity is enlarged, as it soon will be, there will be the possibility of making use of such a scheme.

Nevertheless, it is probably true to say that, since its word order is more or less identical with that of English, the main difficulty in translating French lies precisely in what we have endeavoured to achieve in the program described, namely *successful identification of the various forms of words,* and this being so the program will prove capable of producing an adequate translation of such French prose as has no literary pretensions — for example, scientific publications, for which indeed it is primarily intended. Whether this belief is justified or not will be decided by the actual results of the. experiment, which will be published.