HOW DOES SYSTRAN TRANSLATE?

A brief description of the workings of the
European Commission's English-French machine translation system.

Paper prepared for a meeting of the
Natural Language Translation Specialist Group
of the
British Computer Society
23 April 1980

Ian M. Pigott
Commission of the European Communities
Luxembourg

HOW DOES SYSTRAN TRANSLATE?

A brief description of the computational and linguistic aspects of the European Commission's English—French machine translation system.

INTRODUCTION

Many hundreds of pages have been written on the technical details of machine translation systems and almost as many more have been devoted to the results of evaluations of MT performance.

On this occasion, in view of the mixed audience of linguists and computer specialists, I shall try to explain as simply as possible how Systran actually produces translations without dwelling on any specific aspects of the system or its performance.

However, before describing the system's workings, I should like to say a few words about its historical background and how it compares with other MT developments.

The first large-scale machine translation development project took place in the United States at Georgetown University in the late fifties and early sixties when the U.S. authorities spent some 20 million dollars on developing a Russian—English MT system. Funding was however discontinued in 1966 with the publication of the ALPAC report which concluded that the results obtained were not sufficient to warrant further development. The Georgetown system was nevertheless used both users found the output adequate for purposes of information gathering.

Not everyone agreed with the outcome of the ALPAC report. Indeed, several of those involved in the Georgetown project decided to go it alone and continue development on a commercial basis. The most successful was undoubtedly Peter Toma who developed Systran which represented a considerable improvement over Georgetown owing to its dictionary structure on the one hand and its more sophisticated parsing capability on the other. The Russian-English Systran system became operational in 1970 and has since been used extensively by the U.S. Air Force and other American government agencies.

By the time the English-French system was being developed in 1973, further sophistications had been introduced and in. 1975, when the European Commission undertook a survey of free-syntax MT systems in existence at the time, Systran came out on top. The Commission thus decided to purchase and further develop the English-French system and has since extended coverage to French-English and English-Italian.

Yet Systran was by no means the only M.T. development in the 1970's. Linguistic research led to many new systems based on new approaches to grammar, perhaps the most successful being Logos (used mainly for English-Vietnamese) and Meteo (used for translation of Canadian weather forecasts between English and French). However most work on new approaches has remained at an experimental level and it still remains to be seen whether recent linguistic theory is in fact ideally suited to machine translation.

The European Commission is actively encouraging research along these lines and has plans for making full use of European expertise in Eurotra, a multilingual MT system planned to cover all the EEC languages.

Last but not least we have recently seen the effects of hardware miniturization on automatic translation with the advent of the Weidner MT. system and a number of so-called pocket translators.

OBJECTIVES

Initially, the Commission undertook development of Systran for two rather different applications. The first of these was to assist in-house translators in their work by producing machine-translated versions of texts for post-editing'. However, despite considerable improvement in quality over the past four years, the general feeling among translators seems to be that the potential gains in overall cost-efficiency are not sufficient to outweigh the considerable amount of post-editing which needs to be done.

For this reason, top priority has now been given to the second objective which is to use the system for producing raw translations (i.e. without post-editing) of abstracts accessed via Euronet, the recently inaugurated European documentary data base network.

Evaluations have shown that highly intelligible output can be obtained from Systran even if there are often shortcomings in style and idiomatic usage.

However, for purposes of information gathering these constraints appear to be of secondary importance, success depending far more on accurate terminology. Our present efforts are therefore being directed first and foremost at extending our dictionaries in order to be able to produce acceptable translations of abstracts in a variety of technical fields.

OVERVIEW OF THE SYSTEM

Systran is a free-syntax system. This simply means that it is designed to accept and translate any text in the source language (i.e. without any restrictions on the grammatical structure of the sentences or on the terminology used) and, without any human intervention, to produce a translation in the target language. Input of the source text can either be handled by keying the text in either on punchcards or, more usually, on magnetic tape or by automatically converting texts already in machine-readable form (abstracts from data bases, photocomposition tapes, etc.) into the specific format required for the system. The translated text appears in hard copy form side by side with the source text; the final printout is complete with upper and lower case characters and accents.

For certain applications post-editing or revision of the machine output may be required. Until now translators have written in their post-edit ing changes on paper but we are investigating the use of word processors *at* this stage.

BASIC COMPUTING ASPECTS

The program is written basically in IBM assembler language with special macro instructions to facilitate linguistic processing. It was originally designed to run on IBM computers (IBM. 370/158) but has now been converted to operate on the Commission's Siemens computer, a 7760 *w*ith operating system BS 2000.

Translations are normally run in the batch mode which can accommodate up to 300,000 words per CPU hour although in practice translations do not usually average more than 10,000 words.

In addition to the translation component, the system also includes dictionary management and updating facilities, a number of support programs for text handling and printing, and facilities for obtaining KWIC indexes and word frequencies .

THE TRANSLATION PROCESS

The linguistic part of the system contains two rather distinct components, he dictionaries and the translation programs, although there is of course constant interplay between both in the translation process.

The dictionaries are the major source of static information about words and expressions while the translation programs make dynamic use of this information in order, ultimately, to arrive at the correct selection of terms and grammatical relationships to provide a meaningful translation.

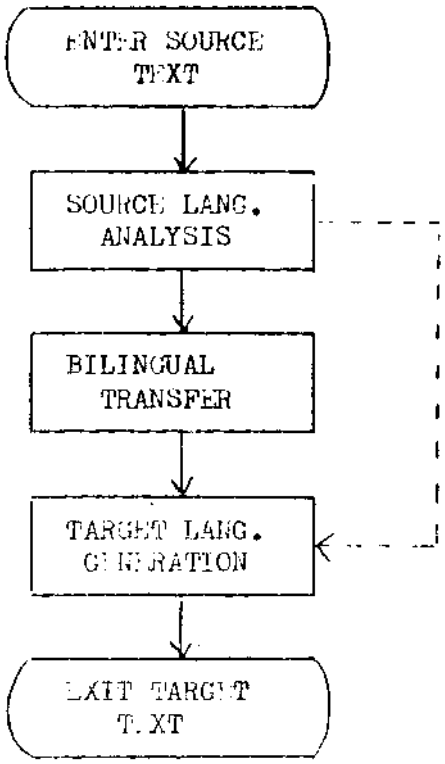
Three rather different operations need to be carried out to convert the source language text into the target, or in our case English into French. First, the English text must be parsed, that is the grammatical function of each word and its relationships with other words must be established. In Systran this is done principally on a sentence-by-sentence basis and is referred to as analysis. As the success of the remainder of the translation process is largely dependent on correct results at this level, a great deal of care needs to be taken to ensure the best possible output here.

At the other end of the process, creation of the French translation, much the same operation is done in reverse. On the basis of the grammatical information obtained from the analysis of English, the corresponding French sentence is structured on the basis of acceptable French word order, agreements and inflections. This process is known as target generation.

However, as any translator knows, successful translation depends to a large extent on selecting the correct meaning of a word in context and on using the idiomatic structures of the target language.

Thus, something more is needed than just analysis and generation. An intermediate stage which inserts the correct contextual terns, structures and phrases in translating a given source language (English) into a given target (French) is essential. This stage is commonly known as the transfer stage and deals with those aspects of machine translation which are truly bilingual, that is which go beyond the basically monolingual capabilities of source and target.

The MT process in Systran can therefore be represented by the following (greatly simplified) flowchart.



The dotted line between source analysis and target generation is simply to indicate that in some cases, it is possible to produce an acceptable word-for-word target rendering without going through the transfer box, although more often than not, high-quality output will require a fair amount of bilingual transfer treatment.

Although we are concerned here mainly with the English-French system, I should point out at this stage that to a considerable extent the analysis and generation components can be used for various language combinations. For example, the same English analysis programs are used by the Commission to translate into both French and Italian and are also used in other Systran developments for translation into German, Arabic, Spanish, Portuguese and even Persian.

ENGLISH ANALYSIS

Now that we have seen the general flow of the translation process, we can look a little more closely at the workings of each of the three main components in practice.

The first, most essential and indeed the most complicated part of the process is analysis. This may seem strange to those who are native speakers of English. Despite our lack of inflection and formal syntax, small children have little difficulty in intuitively understanding the structural relationships in a sentence and indeed, an intelligent child of nine or ten will be able to pick out the nouns, verbs, prepositions, adjectives and adverbs in an English text and may will be able to identify the subject and object of a verb.

Why then is it so difficult to handle this operation by computer ?
The simple answer to this question is that no one has yet been able to provide a fully dependable mathematical description of how the human brain manipulates and understands language. It is not clear, for instance, to what degree comprehension of a sentence is based on its formal structure and vocabulary and to what extent it depends on overall context, that is not only on the surrounding textual or verbal material but on the author's or speaker's preconception of what the reader or listener is likely to interpret.

A simple illustration of this is the English word LEAD.
Out of context this is utterly ambiguous yet in the following sentences its meaning becomes quite clear :

- Steel and lead are useful materials (l)
- U.S. production was in the lead (2)
- The lead from the earth terminal was defective (3)
- This may lead us to disaster (4)
- These changes should lead to considerable improvements (5).

However, even in a complete sentence such as :

- Lead connections must be avoided (6)

an ambiguity still exists.

As can be seen, the problems here are twofold. On the one hand, the meaning of the word is partly dependent on its grammatical function (noun in 1,2 and 3, verb in 4 and 5) and partly on structural relationships with other words in the sentence. In (1) we recognize that lead is a metal because it is in enumeration with steel and is stated to be a material.
In (3) it is a cable as we recognize the syntactic and semantic relationships between lead and terminal. In (6) however, it cannot be established whether LEAD refers to the metal or to cables.

The reason I have gone to such lengths to explain these examples is simply to show why a sentence needs to be correctly analysed if an acceptable translation is to be produced. Target meanings and structures based on part of speech, syntax and semantic relationships can obviously not be inserted unless dependable parsing information is available.

DICTIONARIES FOR ANALYSIS

In analysing any sentence, the first step is to obtain sufficient static information about each of its component words for subsequent processing. Each English word therefore needs to be accompanied by a description of its basic syntactic characteristics such as part of speech (noun, verb, etc.), gender, number (singular or plural), person and tense. For example door will be described as a common noun, neuter, singular. Translates will appear as a finite verb, third person singular, present tense.

In cases where a word can function as more than one part of speech, distinct dictionary entries for each part of speech are made. Clerical will therefore be entered twice, once as a noun and once as an adjective, with an indication in the noun entry that it may also function as an adjective, and a similar indication in the adjective entry stating that it may function as a noun. The correct choice will be made by a series of tests in the translation process.

Where applicable, information is also given about the potential grammatical government of a word. Verbs, for example, always have information on transitivity (usually transitive, usually intransitive, always intransitive and, in rare cases, always transitive) in order to help establish whether, in certain positions, noun groups are acting as objects or not. In addition, they may be described as being able to govern an infinitive (remember to come), a predicate adjective (it appears useful), a present participle (avoid doing something), two direct objects (they elected him president) and. so on. Nouns, too, may often need to carry syntactic information to indicate, for example, that the noun plus certain prepositions may frequently govern a gerund (a method of writing reports) or that it can have a noun clause in apposition (the fact that it is difficult). To a lesser extent adjectives, adverbs and conjunctions require information on their possible syntactic relationships.

Finally, the syntactic information appended to a word may describe restrictions. Some present participles act as adjectives, some do not. The latter can be described accordingly to ensure that a phrase like redefining criteria is interpreted as the redefining of criteria rather than criteria concerning redefinition.

Very frequently syntactic information alone is not sufficient in itself to provide the translation programs with all the information they need to structure a sentence. For this reason, a certain amount of semantic information in the description of a word proves useful, particularly for nouns, as these often occur in enumerations where pure syntax gives little clue to the linkups.

At a fairly general level, basic characteristics stating whether, for example, a noun is concrete or abstract, countable or non-countable, can prove very useful and are widely employed. At a more specific level, semantic descriptions stipulating that a given noun is, say, a chemical compound, a container, a device or a food can provide additional help in sentence structuring.

These, then, are the various types of information available for individual words in the English source dictionary which, for those interested in statistics, now contains about 75,000 terms.

The second most important source of static information for analysis is a dictionary which enables any sequence of words with a basic syntactic function to be strung together and treated as a single word. A phrase like in_order to can, in this way be reduced to the equivalent of a one word entry functioning as an infinitive particle. Once this has been done, the translation programs will no longer have to examine whether, for example, order is functioning as a verb or a noun, in as a preposition or adverb. In practice, the likelihood of any use of this string other than that of the infinitive particle is so remote that it need not even be considered although, admittedly, perhaps once in a million pages of text one could come across *a* sentence such as :

- The points were presented in order to the members of the House.

This would be misanalysed and mistranslated but the vast majority of cases would be resolved correctly.

This dictionary facility is extremely useful in in dealing with prepositional phrases, compound conjunctions and (adverbial expressions as one-word equivalents.

The last dictionary file which helps analysis along is the dictionary of noun phrases. The fact that two or more terms appear in this listing clearly establishes that their functions are basically nominal or adjectival, any other grammatical possibilities (e.g. verbal) being ruled out. The insertion of machine translation as an expression of this type will prevent any interpretation of machine as a verb in this context and will force its resolution as a noun or adjective.

There is, then, quite a variety of dictionary information available to the translation programs for use in defining the structuring of a sentence. Let us now examine how this is achieved.

ANALYSIS PROGRAMS

Particularly when translating from English with its tremendous lack of inflections, the establishment of the correct part of speech for each word in the sentence is an essential operation in the translation process. We saw earlier with the example of lead, to what extent meaning is dependent on part of speech. However, even more important than precise meaning at this stage in the process, is the provision of correct syntactic information for sentence structuring.

The following sentence :
"Some states wish that more funds could be provided for work on term banks"
seems perfectly clear to all of us. However, each of its component words could, in other contexts, function as a different part of speech. Some could be used as a pronoun rather than adjectivally, states as a verb, wish as a noun, that as a demonstrative, and so on.

Each of these words will therefore have two or more entries in the basic dictionary (one for each part of speech) but only one of these in each case will be correct for the above sentence.

The process of deciding which particular part of speech is required for the specific function of a word in a sentence is technically referred to as homograph resolution and comes right at the beginning of dynamic analysis.

The homograph resolution program is certainly the most complex of all the linguistic processes in the English-French translation system. It consists of 83 separate routines, each of which is designed to choose between a particular set of parts of speech (noun or adjective; finite verb, infinitive or noun; preposition or adverb, etc.). There is of course some branching between routines but to a large extent most of the routines are self-contained.

Each routine works on a series of contextual tests. In our example, the test concerning some could, for example, ask whether the word to the right (states) can be a third person plural verb. If, as is the case here, it can not then the resolution of some as a pronoun may tentatively be judged as unlikely and a further test, examining whether states can be a plural noun may be made. States could be, and is of course a plural noun, and it could therefore be concluded that some has an adjectival function.

This is, in fact, a tremendous over-simplification of the testing process. In practice anywhere up to 900 different tests may have to be made on a given word before its part of speech can be clearly established. Perhaps rather surprisingly, the most difficult case to resolve is the potential ambiguity between the true past tense of an English verb and the past participle.

All part-of-speech homographs are resolved sequentially from left to right as they appear in the sentence. This approach has the advantage of always having reliable information at hand about words to the left, when making tests on any word, within a sentence. For example, once it has been established that some is adjectival, it is fairly easy to decide that states is a plural noun. And with these two pieces of information available, it is not too difficult to decide that wish is a verb.

Nearly all the tests used in the homograph resolution program, as indeed those in the other translation programs, are written largely in linguistic macro instructions specially compiled for the 'Systran system. There is, for example, a macro TESTC (test character) which, in combination with dictionary information provides, among other things, an easy way to ask questions about the grammatical function of words surrounding any word whose part of speech has yet to be established.

If, as rarely happens, a homograph routine is unsuccessful in clearly establishing the part of speech of a word, then, on the basis of dictionary information, the part of speech corresponding to its most frequent usage is employed.

Once the part of speech of each word in the sentence has been set, it is possible to go on to more interesting aspects of linguistic analysis.

The next program in aimed at breaking down the sentence into a main clause and any subordinate clauses. In our example, that will already have been resolved by the homograph program as a conjunction introducing a noun clause. The clause boundary definition program will then insert a marker on the word wish to show that it is the last word in the main clause and, after a further series of tests will set another marker on the word banks to indicate it is the last word in the subordinate clause. The words
"Some states wish"
will then each be marked as belonging to the main clause while
"that more funds could be provided for work on term banks"
will each be marked as belonging to the subordinate noun clause.

These markers are in fact what computer specialists refer to as hexadecimal representation and are stored in one of 160 bytes (or boxes in the computer memory) for later access.

After establishing clauses, which is not always quite as easy as I may have make it seem, we now have a series of syntactically complete sentence units which can be given a true grammatical structure.

The first task here is to establish primary syntactic relationships between nouns and their modifiers (other nouns or articles and adjectives), between verbs and their objects, between words governing infinitives and gerunds, and so on. This is done by scanning the sentence from right to left and setting pointers between words governed and governing words.

For example, in our main clause
"Some states wish"
we have established that states is a noun and some an adjective. On the basis of word positions (some is to the immediate left of states and the next word is a verb), we can use information about the probable structure of an English clause to ret a pointer between the noun states and the adjective some to indicate that some is an adjectival qualifier of states .

This relationship can then be stored in the analysis byte area on each of these words, some having an adjectival marker pointing to states, and states having a "modifier-by-an-adjective" marker which points back to some.

This process is repeated in much the same way for the other basic syntactic relationships in the sentence.

The next, and rather more difficult task, is to establish the relationships between words in enumeration. This can often be done on the basis of syntactic information. For example, because of the plural verb are, it is clear in the sentence

- Smog and pollution control are important factors

that both smog and pollution _contro1 are subjects. Smog can therefore be marked as being in enumeration with control.

However, in

- Smog and pollution control is under consideration

the singular verb shows that control is the only subject and that smog and pollution are in enumeration, both modifying control.

Very often such syntactic information is not available and enumeration can only be established on the basis of semantics. For instance in the heading

- Zinc and aluminium components

it is obvious to all of us that we are referring to components made of zinc and components made of aluminium.

Providing both zinc and aluminium carry the same semantic marker (chemical element), their correct enumeration will also be established at this level of programming.

The final stage of analysis is concerned with establishing the deeper grammatical structure of the sentence. For example, the surface subject of a passive construction in English such as

- Pressure was released

is in fact the deep object of the verb released (someone or something released pressure) . Marking pressure here as the deep object not only assists in the selection of its correct translation at target level but also facilities restructuring of the entire target sentence when, for example, an impersonal active construct ion is required.

Other deep relationships of this kind are established in order to correctly identify the subjects and predicates of all the finite verbs in analysis.

Once all these stages of analysis have been completed, a clear picture of the sentence structure will be obtained. In the computer memory, this information is stored in the 160 bytes available for each word in the analysis area. However, for error correction and system development, it is possible to print out this digital information, which although rather meaningless to those not familiar with the system, is fairly easy to interpret after a period of practical involvement. In our experience, though, translators are not always prepared to devote the necessary tine to interpreting such highly technical output and so efforts are being made to convert digital information from the computer language into meaningful natural-language equivalents.

FRENCH GENERATION

The next process in the translation sequence is in fact transfer but as the workings of transfer are to some extent similar to those of generation, it will be simpler to consider generation first.

At target level, three main operations need to be carried out.
First, the words in the source language must be allocated target language meanings. Secondly, the words chosen to provide such meanings must be correctly inflected in accordance with the rules of the target language and thirdly, the structure or word order of the target translation must be adapted to the accepted norms of the language in question. Here again the main source of static information in the target dictionary.

FRENCH TARGET DICTIONARY

For each one word entry in the English source dictionary a basic meaning is allocated in the target dictionary.
As this meaning will be used in all cases where a more specific choice cannot be made on the basis of contextual coding (transfer), an attempt is made to allocate a term which will most often give the basic meaning of the source equivalent. A good example here is the word station whose basic French meaning is that of poste, which is understandable in most contexts, whereas equivalents such as gare, station, base or centrale are far more specific and can only be introduced dependably on the basis of context.

Side by side with the French meaning of a word, information must also be provided about inflections, particularly for verbs, adjectives and nouns, gender, for nouns and pronouns, and the governmental and structural behaviour of that word in French. The inflections corresponding to all regular and irregular forms of verbs, adjectives and nouns are listed in tables and are triggered by digital codes in the dictionary. Gender is simply indicated as masculine or feminine for nouns and pronouns, adjectives and other adjectival forms being inflected during the generation process on the basis of their structural links with the nouns concerned.

The governmental requirements of the target word might be quite different from those of the source equivalent. In particular, the choice of the infinitive particle in French (de, pour, à or zero) is largely dependent on the word governing the infinitive. Markers can therefore be appended to the governing word in the target dictionary to ensure, for example, that propose to do is rendered proposer de faire while like to do is translated simply aimer faire.

Various other codes of this type are used in connection with nouns and verbs to ensure that the correct French preposition is inserted. Salt content will be translated as teneur en sel rather than teneur de sel as a result of a marker attached to teneur. A pound of bread becomes une livre de pain with no article owing to a marker on livre whereas the normal resolution of a sequence of this type would require the full article form (e.g. the manufacture of bread would be translated la fabrication du pain).

Rearrangement markers can be used with adjectives and verbs in order to position them correctly in relation to the nouns and verbs they modify. Most adjectives in French come after the noun; special rearrangement information is therefore appended only to exceptions like bon which usually come before the noun.

Informationr is also given in the target dictionary to deal with cases where, say, a French verb is reflexive when the English equivalent is not (rise = se lever), or where a French verb is conjugated with être in the past tense rather than with avoir.

GENERATION PROGRAMS

Drawing on the information provided by source analysis as well as on target dictionary information, the primary aim of the generation programs is to ensure that the target word is correctly inflected in accordance with the grammar of the target language.

For French, this does not only mean that an adjective must be in correct agreement (gender and number) with the noun or nouns it qualifies but extends to more complicated synthesis to make sure, for example, that a past participle in a verb complex agrees with a preceding direct object, that the imperfect rather than the perfect tense is used in conditional clauses introduced by si and even covers some quite complicated transformations as when an English plural has to be reduced to a French singular and an additional negative has to be inserted. Thus

- No comments were made

becomes

- Aucun commentaire n'a été fait.

rather than

- *Pas de commentaires ont été faits.

The other main function is to undertake the rearrangement of word order. Quite appart from the obvious exceptions which must be treated at target dictionary level, normal French word order is rather different from English. Adjectives must usually be rearranged, after the noun (black telephone → telephone noir), a noun modifier will usually be transformed into a prepositional phrase after the noun it modifies (investment bank → banque d'investissement) , pronoun objects come before the verb (I saw it → Je l'ai vu(e).), and negatives behave quite differently (He could not see anyone → Il ne pouvait voir personne).

Finally, substantial restructuring is required when an English passive is to be translated as an impersonal French active form:

- Experiments carried out on rats and mice are described in detail.
- On decrit en detail des experiences effectuees chez les rats et les souris.

TRANSFER

I have left transfer to the end as it not only draws on analytical information from source but also makes use of many aspects of generation. In that it is largely responsible for transforming a fairly literal translation into a most sophisticated one - both as far as meanings and structures are concerned - it is an essential part of the overall process.

It can be regarded as consisting of two main features, a contextual dictionary and a set of lexical routines to deal with words which behave very differently between source and target.

CONTEXTUAL DICTIONARY

The contextual dictionary is a very powerful and useful tool for selecting the "mot juste" in the translation process.

Up to now, we have seen how the system can choose between the meanings of a word which depend on its part of speech, yet little or nothing has been said about words which require a range of different meanings for the sane part of speech.

One of the simplest, and indeed most frequent ways of dealing with the problem is by allocating special meanings to noun phrases or expressions. For example, assigning the meaning traduction automatique to the expression machine translation will ensure that the literal translation traduction de machine is not obtained. Similarly, to go back to our example of station, we can enter the correct translations of railway station, power station, service station and filling station as gare, centrale, station de service and poste d' essence respectively.

As may be imagined, this facility is particularly useful in obtaining consistently correct translations of technical expressions.

Very often, however, the meaning of a word does not depend so much on its appearance in an expression but rather on its syntactic and/or semantic relationships with other words in the sentence. In our earlier example:

- Steel and lead are useful materials.

we can clearly establish the meaning of lead by writing a contextual dictionary rule based on results of analysis - which of course are still fully accessible at this stage - stating that if lead is in enumeration with steel it is to be translated plomb.

However, this rule is rather restrictive as it will only work when the word steel occurs. Many more situations of this type can be covered by drawing on the semantic information available. For example, a rule stating that whenever lead is in enumeration with any chemical element or with any material (both of which characteristics are documented in the form of semantic markers) , then the correct meaning plomb will be provided in many situations of this type.

Similarly, we could have a rule stating that if lead is the subject of the verb to be, and if the predicate of that verb carries a semantic marker material (as of course is the case with material itself), then, once again, lead should be translated plomb. This would not only cover the example

- Lead is a useful material

but would also cover such occurrences as:

— Lead is a substance which often causes water pollution.

Sometimes very elementary syntactic information can be used to establish the meaning of a word. The noun leads in its plural form is far more likely to have the meaning câble than either plomb or, say, avance. It could therefore simply be stated that if lead is a plural noun, its meaning is câble.

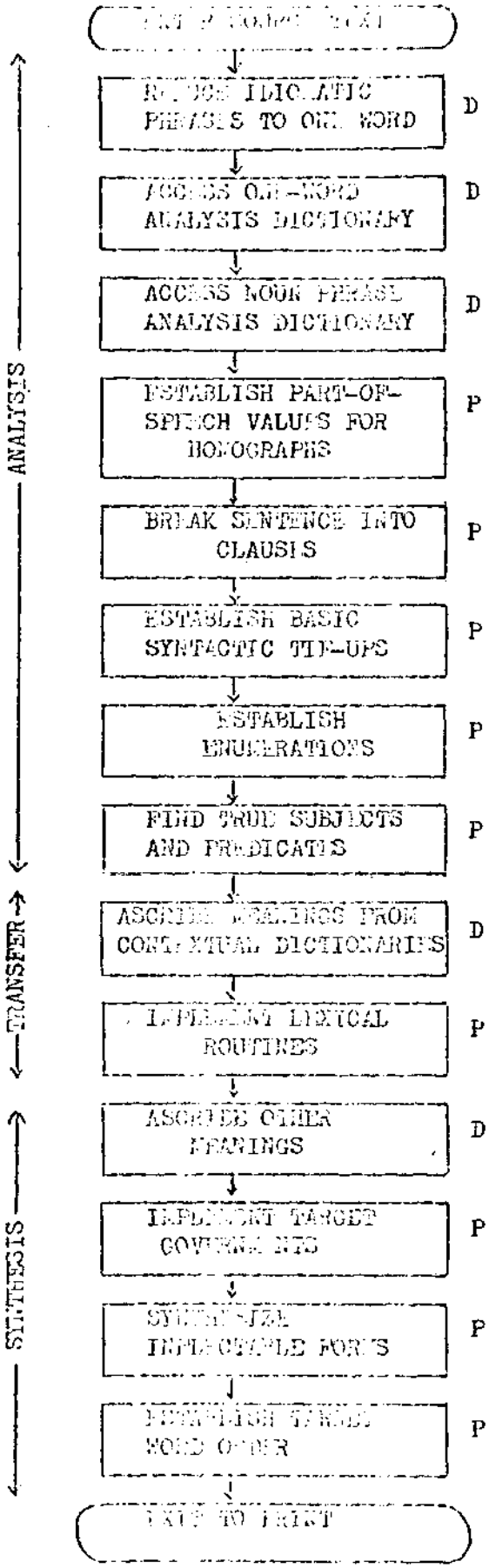
Any of the syntactic and semantic relationships established during analysis can be used for ascribing meanings in context. Among the most useful are perhaps subject → verb, verb → object, noun modifier → noun modified, adjective → noun, adverb → verb, preposition → object and enumerations between any parts of speech. Most of the rules tend to be very simple but some are extremely complex, complexity depending to some extent on the limitations imposed by the linguist. For instance, some 70 different dictionary entries have been written around the word in to obtain the correct translation in context. These cover a whole range of simple and complex syntactic and semantic relationships.

The other means of obtaining the correct transformations of structure and meaning at the transfer stage is by lexical routines. These are, for the most part, fairly short programs which are written on specific words or classes of words requiring treatment which goes beyond the limitations of the dictionaries.

The date routine, for example, is a program based on all words carrying the semantic marker month (i.e. all months and their abbreviations) and on combinations of fig-ares (the day of the month and the year) to correctly reproduce dates in French. December 1, 1979 will thus be translated le 1er décembre 1979 rather than Décembre 1, 1979.

One of the longer lexical routines concerns the word as. On the basis of information from analysis, the routine tries to establish which particular translation and structure is required for each of its parts of speech in context. The conjunction can, for example, be variously translated as comme, pendant que, à mesure que, or puisque. Moreover, depending on which these is chosen, the sequence of tenses in the subordinate clause will vary between perfect and imperfect. This can be successfully handled by lexical routine but could not be triggered from the dictionary alone.

SUMMARY Now that we have seen the main features of each stage in the translation process we can summarize the correct sequence of operations in flowchart form.



In point of fact, there are a number of other program which have not been mentioned in this description as they would add little or nothing to the understanding of the overall process. They concern such matters as establishing sentence boundaries, assigning information to words not in the dictionaries, dealing with abbreviations and sorting words into three different categories of length for dictionary lookup.

The M.T. process consists then of quite a complicated interweaving of programs and dictionaries. Its success, that is the quality of translation produced, depends far less on the computational functioning of each operation — although this must of course be ensured — than on the store of linguistic information in the system.

Unfortunately, provision of such information can only be achieved through a long process of trial and error. Language is such a complicated process that it is virtually impossible to predict exactly how any word or phrase will behave in all contests or exactly what its various translations will be. Quality improvement in the future will therefore depend largely on the introduction of more and more contextual rules to obtain the correct meanings and structures for the target language.

The system itself will probably also undergo some changes in order to accommodate linguistic information in the most efficient way. Some of the aspects under consideration here are an algorithm to identify automatically subject field and document typology, to extend some aspects of transfer to "force" some of the operations carried out in analysis, to increase the use of semantics and, last but not least, to investigate seemingly more dependable ways of dealing with the all-important problem of assigning part-of-speech values in grammatical homographs.

The system has certainly not reached its ceiling, even in its present form, and we can, I believe, look forward to a great deal of improvement in the future. That is not to say that the present system is not mature enough for use. For certain types of text and certain subject fields, the quality of output is already good enough for purposes of information gathering. Within two to three years, we can therefore expect it to be used fairly extensively for providing a multilingual service in Euronet. And other applications will probably emerge as psychological barriers are overcome..

Since it can be fairly assumed that all three candidate countries will become Members of the Community during the lifetime of the next Commission (1981-1985), the principle that the Commission should be composed of not more than one Member from each Member State should be applied from its inception in January 1981, which coincides with the entry of Greece.

229 Puisqu' on peut raisonnablement supposer que chacun des trois pays candidats deviendront des membres de la Communauté pendant la vie de la prochaine Commission (1981-1985), le principe que la Commission devrait se composer de pas plus d' un membre de chaque Etat membre devrait être appliqué à partir de son commencement en janvier 1981, qui coïncide avec l' entrée de la Grèce.

Thus, for the period before further enlargement, the Presidency and the eight portfolios will be divided among ten Members.

230 Ainsi, pendant la période avant davantage d' agrandissement, la présidence et les huit portefeuilles seront divisés parmi dix membres.

When Portugal and Spain join the Community, and the number of Members of the Commission rises to twelve, the number of portfolios will have to be increased by two by detaching some sectors of activity from what will by then be eight portfolios.

231 Quand le Portugal et l' Espagne rejoignent la Communauté, et le nombre de membres de la Commission se lève à douze, le nombre de portefeuilles devra être augmenté de deux en détachant certains secteurs d' activité de ce qui sera d'ici là huit portefeuilles.

Indeed, the likely increase in the workload after this further enlargement will justify increasing the number of portfolios to ten without impairing the quality of the administration and without entailing an imbalance in sharing out duties among Members of the Commission.

232 En effet, l' augmentation probable de la charge de travail après cet agrandissement ultérieur justifiera accroître le nombre de portefeuilles jusqu' à dix sans altérer la qualité de l' administration et sans impliquer un déséquilibre en partageant des fonctions parmi des membres de la Commission.

The choice of the sectors to be detached should be made in the light of the development of the Commission's work and of

233 Le choix des secteurs à détacher devrait être fait à la lumière du développement du travail de la Commission et de

MEG.PAT.DESIE.

A tool for coal planes and similarly acting mining machines.

The invention relates to a tool for coal planes and similarly acting mining machines;

a removable tip is connected by a tenon and mortise joint to a tool carrier and secured to the latter by fastening means.

Coal planes make it possible to extract the coal by a planing process in which they are drawn along and simultaneously pressed against the face by an endless cable or chain.

A coal plane is known in accordance with the prior art portion of Claim 1 (German utility model No ...) in which the head has a dove-tailed mortise the sides of which are uniformly spaced along their length.

A tool head (removable tip) with a correspondingly shaped tenon is inserted into this mortise until its face contacts

1 MEG.PAT.DESIE.

2 Un outil pour des rabots à charbon et des machines d' abattage fonctionnant de manière analogue.

3 L' invention concerne un outil pour des rabots à charbon et des machines d' abattage fonctionnant de manière analogue;

4 Une tête démontable est reliée par une liaison tenon-mortaise à un support d' outil et retenue à ce dernier par des dispositifs de fixation.

5 Les rabots à charbon permettent d' extraire le charbon par un processus de rabotage où ils sont entraînés et simultanément pressés contre la face par un câble ou une chaîne sans fin.

6 Un rabot à charbon est connu selon le préambule de la revendication 1 (modèle d'utilité allemand no ...) où la tête a une mortaise en queue d'aronde dont les côtés sont uniformément espacés sur leur longueur.

7 Une tête d' outil (tête démontable) avec un tenon de forme correspondante est insérée dans cette mortaise jusqu'à ce

```
93  IMPULSE PRINTOUT COUNTER                                    1 K 0 D'IMPULSIONS
                                                                    CDS-NOPREP
                                                                2 K 0 IMPRIMEUR
                                                                    CDS-NOPREP
                                                                3 K 0 COMPTEUR
91  IN $C-B0 $C-B1,E,A0 $C-B0,+,1 TIME                          1 0 0 DANS
                                                                5 0 0 $$$
                                                                    CDS-ENDPRP
91  IN $C-B0 $C-B1,E,A0 $C-B0,M MIN                             1 0 0 EN
91  IN $C-B0 $C-B1,E,27 $C-B0,M TURN                            1 0 0 $$$
                                                                    CDS-END2A
91  IN $C-B0 $C-B1,NE,3C $C-B0,-,1 $C-B18 $C-SCINQ              1 0 0 EN
                                                                    CDS-DELART
91  IN $C-B0 $C-B8,E,20 $C-B0 FORM $C-B1,E,10 $C-B4,D,01        1 0 0 EN
        $C-B30,E,DC
                                                                    CDS-DELART
                                                                5 0 0 FORME
91  IN $C-B0 $C-B8,E,27 VIEW                                    1 0 0 2A
                                                                4 0 0 AVIS
91  IN $C-B0 $C-MU                                              1 0 0 EN
                                                                    CDS-DELART
91  IN $C-B0 $C-UNABR                                           1 0 0 EN
                                                                    CDS-DELART
91  IN $C-B0 FRENCH $C-B16,E,00                                 1 0 0 EN FRANCAIS
91  IN $C-B0 LUXEMBOURG                                         1 0 0 $$$
                                                                    CDS-END2A
                                                                3 0 0 C$LUXEMBOURG
91  IN $C-B0 PARTICULAR $C-B16 $C-B4,Z,02                       1 0 0 EN PARTICULIER
91  IN $C-B0 THAT $C-B0 $C-B1,E,3C                              1 0 0 PARCE
                                                                3 0 0 QUE
91  IN $C-B0 THAT $C-B0,M IT                                    1 0 0 PARCE
                                                                3 0 0 QUE
91  IN $C-B0 TRACE $C-B4,D,02                                   1 0 0 EN TRACES
91  IN $C-B0,M OVER                                             1 0 0 DANS
91  IN $C-B0,M VARIOUS $C-B16 $C-B4,D,02                        1 0 0 DANS
                                                                    CDS-DELART
91  IN $C-B105,D,02 $C-B0,-,1 $C-B1,E,DD $C-B0,+,2 BILLION      6 0 0 C$BILLION
        OF                                                      7 0 0 $$$
                                                                    CDS-DELART,ENDDE
91  IN $C-B103,D,02 THE CHAIR                                   1 0 0 C'EST L'ENDROIT
91  IN $C-B10?,D,02 THE FIRST PLACE                             1 0 0 C'EST EN PREMIER LIEU
91  IN $C-B105,D,02 THE MAIN $C-B0 $C-B1,E,6D                   1 0 0 C'EST PRINCIPALEMENT
91  IN $C-B105,D,02 THE SECOND PLACE                            1 0 0 DEUXIEMEMENT
91  IN $C-B18 $C-B0 EMERGENCY                                   1 J 0 EN CAS D'URGENCE
91  IN $C-B18 $C-B1,E,AD                                        1 0 0 EN
91  IN $C-B18 $C-B11,D,04 $C-B0,-,1 $C-B1,E,AD                  1 0 0 DANS
91  IN $C-B18 $C-B152,D,PD                                      1 J 0 PENDANT
91  IN $C-B18 $C-B[illegible],D,04 $C-B1,E,AD                   1 0 0 EN
91  IN $C-B18 $C-B8,E,1F                                        1 J 0 EN
91  IN $C-B18 $C-ENPRIS                                         1 0 0 $$$
                                                                    CDS-END2A
    IN $C-B18 $C-INFO                                           1 0 0 $$$
```