FIRST DRAFT OF AN ENGLISH INPUT PROCEDURE FOR MECHANICAL TRANSLATION *

ERNST VON GLASERSFELD
Centro di Cibernetica
e. di Attività Linguistiche,
Università di Milano
Italia

[Received in March 1962]

Theoretical Basis

The research of the Italian Operational School has shown that the contents of every linguistic expression can be reduced to a network of correlations in which each correlation consists of two elements (first correlatum and second correlatum) linked by a third element, the correlator. Since this principle holds for all languages and, further, since there is a far-reaching correspondence between the correlations expressible in different languages, the "Correlational Net" constitutes a natural basis for translation. (Cf. "Linguistic Analysis and Programming for Mechanical Translation", *Methodos,* vol. XII, N[os] 45-46-47).

The process of translation can thus be divided into three main steps:

1) Reduction of the input text to the correlational nets expressed by it (= understanding the text);

2) Adjustments of these correlational nets (necessary only if and when the input language can express nets for which the output language provides no adequate expression);

3) Expression of the correlational nets in the output language.

Natural languages, however, differ greatly in the ways and means by which they designate the characteristic features of a correlational net. On the one hand there are languages which rely almost exclusively on the

(*) The ideas presented in this paper are part of the results of a research project sponsored by the United States Air Force through their European Office of Aerospace Research (Bruxelles). I am greatly indebted to *Jehane Burns,* who patiently collaborated in the development, testing, and compilation of the flow diagrams.

form (morphological particulars) of single words and disregard their position in a string; Latin is a case in point, and to some extent also Russian and Italian. On the other hand there are languages which put all the weight on the individual position of each word in the string. English belongs to this group, of which Chinese is the extreme representative.

This difference in the structure of natural languages makes it obvious that the procedure for the construction of the correlational net expressed in a given input text, although remaining essentially the same for all languages, can be adapted to the one or the other kind of language in order to reach a certain operational economy.

In a language with a great deal of significant morphological differentiation, the possibility of correlating two elements should first be tested according to morphological criteria, while in a language in which the order of succession is the main correlational indication, the possibility of correlating should first be tested according to criteria of order. In this way a large number of correlation attempts, bound to be discarded later, can be avoided.

In English, morphological differentiation of words is rudimental and, consequently, the order of succession is highly significant.

The input procedure presented here is an attempt to achieve operational economy in the particular case of English input. It covers the main operations of step 1 in the above scheme, that is to say, it provides for the construction of the correlational net. It does *not* include a flow chart for "dictionary search", i.e. identification of the word cards (in the machine's vocabulary) corresponding to the words of the actual input text; several procedures for this have been developed elsewhere and any one of them could be applied here.

Our procedure has been worked out for single sentences, each of which must contain a special "End Sign." (Extension of the procedure to cope with input texts of more than one sentence involves an additional subroutine to introduce the required end sign after each sentence; this should not be difficult, but it obviously calls for a special test for full stops, because not all full stops in a text mark the end of a sentence.)

At any rate, it should be borne in mind that the flow charts given here represent a first draft of the procedure and will be subject to further research and correction.

Correlation

The system used to indicate the individual word's possibilities of correlation is the "Tabellone" System presented by the Centro di Cibernetica at the International Conference on Machine Translation, held at the National Physical Laboratory, Teddington (England), in 1961.

Having isolated and indexed the correlators expressible in English, each word card WM of the machine's vocabulary is completed with a string of indices I_c indicating the correlators with which the particular word can enter into correlation. The indices I_c consist of three-figure numbers; a fourth figure, added to them, indicates the "Saturation Function" SF, that is, the place or places the word can occupy in a correlation made with that particular correlator. Besides I_c the WM bears a string of morphological and semantic indices* and is marked with its "Input Number" as it enters the combination cycles.

Words having more than one grammatical or semantic function (polisemy of form or of content) are represented by WM bearing as many I_c strings as they have functions, and for each of these there is an individual I_g string (indexed by different values of P_m) which characterises the particular grammatical or semantic function to which the I_c string refers. In principle it is irrelevant whether the different functions of one and the same word are represented on as many WM or by as many sections on one WM; but as the first tests for the present procedure had to be done by hand, a single word card, or matrix, no matter how cumbersome owing to the multiplication of sections on it, was still easier to handle than several matrices for one word.

The possible Saturation Functions, indicated in conjunction with individual I_c, are:

SF1 = 1st correlatum
SF2 = 2nd correlatum
SF3 = correlator
SF4 = correlator + 1st correlatum
SF5 = correlator + 2nd correlatum
SF6 = correlator + 1st + 2nd correlatum (closed correlation)

Correlators are divided into two kinds:

a) Explicit Correlators (bearing indices between I_c 001 and I_c 109), designated by particular words;

(*) A table of the morphological indices I_g is shown on p. 79; the system of semantic classification is still at the experimental stage and will be published only when it has been tested within the ambit of a vocabulary considerably larger than the one so far used for tests.

b) Implicit Correlators (bearing indices between I_c 110 and I_c 499) designated either by words that are at the same time correlata, or designated merely by position.

Correlations made with *explicit* correlators consist of at least three words. Hence, as the machine proceeds by combining two and never more than two elements, correlations made with explicit correlators have to be arrived at in two steps: first one correlatum and the correlator (giving rise to an element with SF4 or SF5) and then this product and a second correlatum. The intermediary product gives rise to a "Temporary Product Matrix" TPM.

Combinations made with *implicit* correlators consist of two words and are arrived at in one step. They give rise to "Product Matrices" PM, having SF6. As part of the process of production these are *reclassified* as new, single elements for further combination with other elements. Hence all PM are treated in the same way as the original WM and, like these, they contain a string of I_c and SF, indicating their further correlational possibilities, a string of I_g compiled during reclassification from the I_g strings of the component elements, and a string of Inp. N^{os} representing the Inp. N^{os} of all the component elements.

The production of PM, however, is subject also to two kinds of condition:

a) morphological (grammatical) conditions,

b) semantic (contents) conditions.

These conditions are given on a special table, the "Correlation Control Matrix" CCM, for each correlator. CCM are numbered by means of the I_c of the correlator they concern.

On each WM the morphological and semantic characteristics of the word it represents are indicated by means of strings of indices I_g and I_s. Since the CCM procedure for morphological and for semantic conditions is identical, the indices I_s are formally the same as I_g. While I_g are represented by numbers between 100 and 599, the indices I_s will be represented by numbers higher than 700.

Each CCM contains, in a first section S1, the I_g and I_s which an element must show in order to be eligible as a first correlatum in the correlation governed by that CCM; in a second section S2 it contains the Ig and I_s an element must show in order to be eligible as a second correlatum in that correlation.

If the linguistic conditions are such that an element, although eligible as first correlatum of the particular correlation, cannot form this correlation with *every* element meeting the requirements of S2, but can form the correlation only with elements bearing further specific indices, these specific requirenients of correspondence are indicated on a separate "Line" of the CCM.

In a third section, RC, the CCM indicates the "Reclassification", that is, the strings of I_c, I_g, and I_s that are to be registered on each PM produced in accordance with the requirements of the particular line of the CCM. Since correlator designations, in English, differ insofar as some allow for an inversion of the normal word order, and others not, a separate combination cycle has been evolved for elements eligible for correlations with inverted order. In this way a considerable reduction of the combinations that have to be tried by the machine was achieved.

Whenever, in the text, the input order of elements that form a correlation is inverted, we speak of "Maintainment"; that is to say, in the word sequence of the input sentence we may find a word classified as 2nd correlatum (SF2 or SF5) of a particular correlation, *before* the respective 1st correlatum (SF4 or SF1) has arrived.

e.g. 1) he went quickly (normal order)

2) he quickly went (inverted order)

In some cases of inverted word order the elements remain eligible for correlation, even if they are separated in the text by an intervening element ("interval").

e.g. 1) the tall man walked quickly across the street (normal order);

2) the tall man quickly walked across the street (inverted order);

3) quickly the tall man walked across the street (inversion + interval).

In examples (2) and (3) "quickly", the 2nd correlatum of "walked", is *maintained* (or, in colloquial terms, put on ice) until an element that can function as the required 1st correlatum arrives in input and enters the cycle of correlating operations.

Maintainment is also possible in the case of explicit correlators (e.g. "in + the winter + it snows", where the actual input sequence is: correlator + 2nd correlatum + 1st correlatum) and the procedure provides for this in the combination cycle in which, among other elements, the previously produced TPM bearing SF5 are treated in the same way as WM or PM bearing SF2 or SF5, that is, these TPM are treated like any word or combination of words representing a correlator plus 2nd correlatum.

CCM of correlators that allow for "maintainment", naturally, have—besides the part (SN) indicating requirements for correlata in normal sequence — a second part (SM) indicating requirements for correlata entering in an inverted order; this second part, of course, also contains the section RC which gives the reclassification of the respective products.

Preliminary research has shown that, in English, only correlations involving "maintainment" can be made over an interval; consequently, in a sentence in which the correlanda (the elements to be correlated) appear in the order 1st correlatum, correlator, 2nd correlatum, only adjacent elements can correlate (it should be remembered, in this context, that elements can be single words as well as correlations).

This natural restriction developed in English usage has made it possible to divide the construction of the correlations as expressed by an English sentence into two combination cycles, one for "normal" correlations (Diagram N) not allowing any intervals, and another for correlations involving maintainment correlations (Diagram M) and allowing for intervals.

On this basis the correlation attempts to be made by the machine in the course of the construction of the correlational net, could be guided and reduced. On the one hand combination attempts with I_c of correlators that do *not* permit maintainment can be excluded from the maintainment correlation cycle; and this is implemented in the present procedure by operations M2 and M3 (Diagram M). On the other hand, elements bearing SF1 and SF4 (i.e. possible 1st correlata) need not enter the maintainment cycle with elements that succeed them in the input text; the procedure assures this by the fact that in one of the actual correlation cycles (Diagram N) only correlations with normal order are produced, while in the other (Diagram M) only maintainment correlations can result, which may or may not bridge an interval.

The program in its present provisional form allows and provides for maintainment correlation over intervals of any kind and any length within one sentence; it is clear, however, that English usage embodies definite restrictions on word order with regard to maintainment correlations. Operations A8 and A9 and operations C13 and C14 (see pp. 10 and 15, and Diagrams A and C) incorporate one of the simplest of these linguistic restrictions into the present procedure. Research is going forward to establish other rules of this sort, and thus to reduce the number of correlations made by the machine which are meaningless for the reason that current English word order precludes them.

Correlation Cycles

The combination cycles in which the various elements are tested for their correlational possibilities and in which correlations are produced, can best be visualised if one imagines them as an arrangement of two "Tracks" on which elements enter into action in such a way that there is always one and only one element operative on each of the tracks. Thus we speak of "Lists" of elements entering a track for the subsequent insertion of their elements, one by one, into the cycle of combinatory operations.

Track "I" is the path taken by elements constituted by or containing the most recent input word, and track "H" that taken by all other elements that, owing to the input numbers of the items they contain (i.e. owing to their position in the input sentence) can possibly correlate with the element "EI" on track I (see Diagrams N and M).

The different "Lists", or collections of elements found in one store, have been introduced above all to avoid correlation attempts with elements whose position in the input text precludes correlation; the system of lists, or stores, of course, takes into account the fact that only correlations allowing for maintainment can operate across an interval.

All lists contain a permanent element, which is their "End Sign" (see "Abbreviations", p. 78).

Conditions of Correlation

A correlation is made and accepted if:

a) the same correlation index I_c figures on both elements E_I and E_H;

b) the correlation index I_{cI} has a saturation function SF complementary to that of I_{CH};

c) E_I and E_H bear the I_g and I_s required for the correlation and indicated respectively in S2 and S1 of the CCM bearing the particular I_c.

Conditions a and b are examined by the procedure sketched in Diagrams N and M. Condition c is examined by the CCM procedure sketched in Diagram CCM.

Stores, Lists, and Insertion Cycles

To implement the system of limited combination cycles a number of different stores has been introduced, in each of which a "List" (or collection) of elements is assembled which have to undergo the same procedure with regard to correlation attempts.

After input of a new word, the WM corresponding to this word enters on track I and remains there throughout the Insertion Cycles A and B (see Diagram T, p. 74); Cycle A leads it to meet those elements, from preceding levels, which, owing to their adjacent position in the input sentence are eligible for "normal" correlation with the most recent WM (Store 9). These correlations are attempted and, if admissible, made in Correlation Cycle N set off by Cycle A.

Next, Cycle B leads the new WM to meet those elements, from preceding levels, which are eligible for "maintainment" correlation (Store 10) these correlations are made in Correlation Cycle M set off by Cycle B.

The products issuing from Correlation Cycles N and M are at once inserted into different stores according to their SF-reclassification:

Elements with I_c bearing SF1 or SF4 go to Store 1,

Elements with I_c bearing SF2 or SF5 go to Store 2.

Temporary Product Matrices need not enter "Recirculation Cycles" (see below) on the same level on which they have been produced (since the sequence of words corresponding to the order 2nd correlatum + correlator + 1st correlatum is impossible in English) and can, therefore, be kept apart from other products. Hence

TPM4 go to Store 6 (Diagram N),

TPM5 go to Store 7 (Diagram M).

With Insertion Cycle C begins the process of "Recirculation", that is, the attempts to correlate the products so far produced on that level, with elements from previous levels. Again there are two cycles, one for normal and one for maintainment correlation (Cycles D and C respectively). Cycle C leads the products stored in Store 1 to meet the elements eligible for maintainment correlation (Store 10) in Correlation Cycle M. Cycle D does the same for the lists from Stores 3 and 11, leading them to the "normal" Correlation Cycle N.

The Recirculation Cycles C and D continue to alternate as long as new products are produced in either of them. When there is no more new production (see "Production Tests CET and DET), the "End of Level" is reached (Diagram FL, p. 64). At this stage the products of the level, stored in Stores 6 and 7, are transferred into the respective permanent stores (9, 10) and the next level begins with the input of a new word.

Description of Diagrams

DIAGRAM A (p. 56)

Operation

A1 The procedure begins with the extraction (from the machine's vocabulary) of the Word Matrix WM, corresponding to the first word of the input sentence, and its insertion in "Track I". The input sentence must be completed with an "End Sign". This end sign "E.I." has the same form as the word matrices, of which it is, in fact, the last (see page 2).

A2 If the WM entered on track I is "E.I." (which is the case after the last WM representing an input word has been entered), the flow proceeds directly to the operations required at "End of Input" (Diagram FI).

A3 If the WM is not "E.I.", the input counter QI is advanced one unit
A4 and the resulting Input Number "n" is attached to WM.

A5 Next, the elements from Store 9 are inserted in track H and the
A6 first of these elements is examined.

All lists (see page 53), or stores, have a permanent end sign "E.L.", which concludes the collection of elements contained in them. An element of a list may be:

a Word Matrix	WM,
a Product Matrix	PM,
a Temporary Product Matrix	TPM,
or an End Sign	"E.L.".

A7 If the element examined is "E.L." (which means that no or no more elements are in that list), the flow proceeds to the operations shown
A15 in Diagram B, and the elements of List 9 — if there were any — are transferred to Store 11.

(Note: when a list is transferred from one store to a track or to another store, its elements cease to be present in the first store and are present only in the second. The end sign, however, being "permanent", remains also in the empty store.)

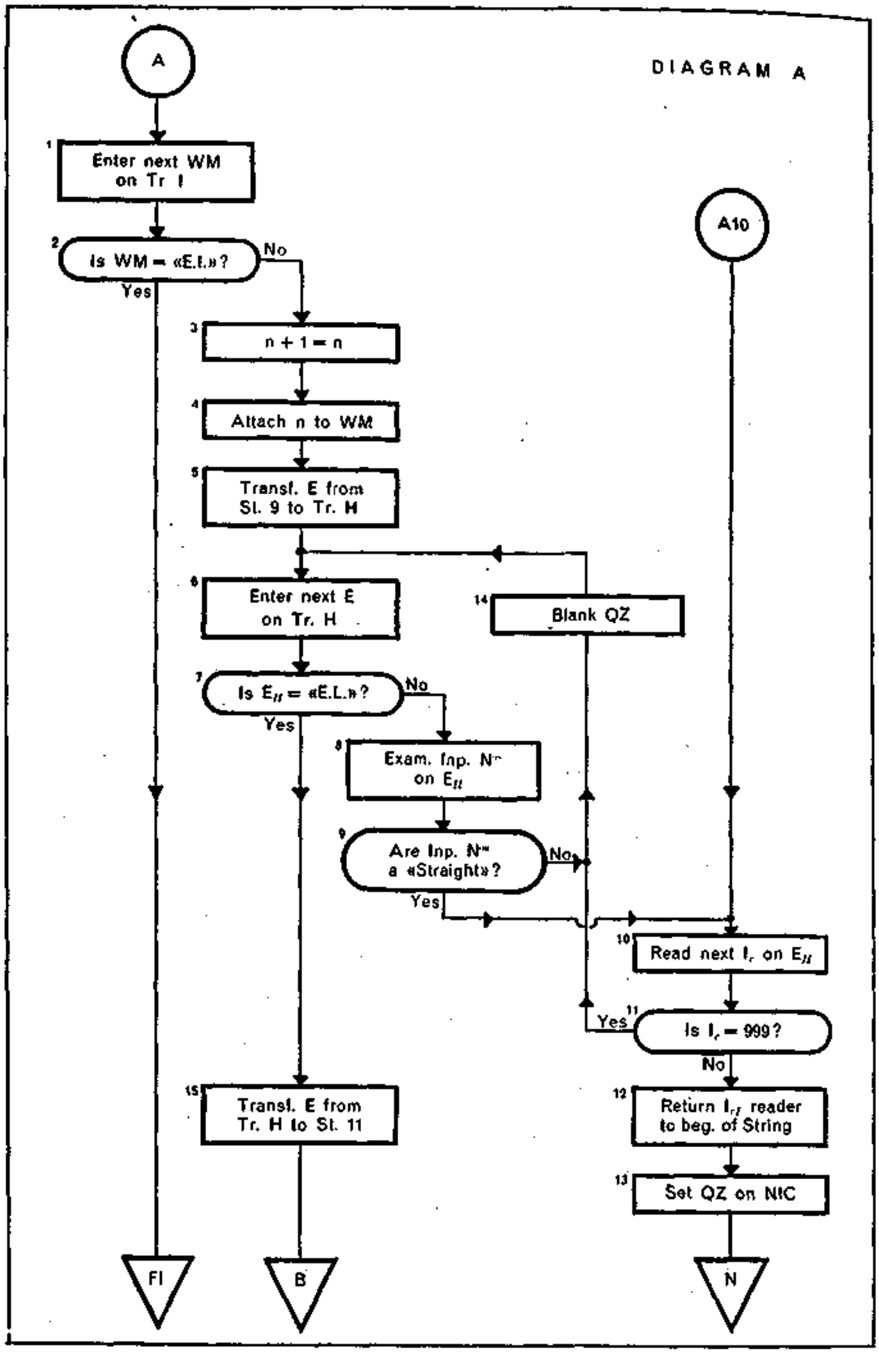


DIAGRAM A
A
1 Enter next WM on Tr I
2 Is WM = «E.I.»?
No
Yes
3 n + 1 = n
4 Attach n to WM
5 Transf. E from St. 9 to Tr. H
6 Enter next E on Tr. H
7 Is E_H = «E.L.»?
No
Yes
8 Exam. Inp. N° on E_H
9 Are Inp. N^{os} a «Straight»?
No
Yes
A10
10 Read next I_r on E_H
11 Is I_r = 999?
Yes
No
12 Return I_r reader to beg. of String
13 Set QZ on NIC
14 Blank QZ
15 Transf. E from Tr. H to St. 11
FI
B
N

A8 If the element is not "E.L.", its string of Inp. N^{os} is examined to
ascertain whether or not they constitute a "Straight", i.e. a conse-
A9 cutive sequence of numbers n, n + 1, n + 2, n + 3,... Only elements
whose Inp. N^{os} are a "Straight" (representing a sequence of adjacent
words in the input sentence) are admitted on track H for combination
in this cycle. If E_H does not show a "Straight", the flow returns to
op. A6 and repeats op. A6-A9 for the next E of Store 9.

Al0 If E_H does show a "Straight", the first I_c of E_H is examined. (The
I_c strings on all elements are concluded by the end sign "999".)

All If Ic_H is "999", which means that there are no more I_c on this E
to be examined, the next E is entered on track H (return to op. A6)
and op. 6-11 are repeated.
A13 If Ic_H is not "999", the flow proceeds to the Correlation *Cycle N*
(Diagram N), setting the Cycle Indicator QZ on "NIC".

A12 Operation A12 has no significance at the first run through Cycle A;
if, however, some Cycles N have already been run through, this in-
struction makes sure that the new I_{cH} will be tested with *all* I_{cI}, and
not only with those left over from the previous Cycle N.)

(Note: the Cycle Indicator QZ has the purpose of directing the flow at the end of subsequent Cycles N, M, and CCM; as these subsequent cycles are the same regardless of the Insertion Cycle they follow, a record has to be made at their beginning in order to direct the flow at their end. Subsequently, this record has to be cancelled whenever

A14 the material of one combination cycle is exhausted.)

DIAGRAM N (p. 58)

Cycle N is the Correlation Cycle for "normal" correlations, that is, for correlations not involving "maintainment" (see page 51). At the beginning of Cycle N both tracks already contain an element and it has been ascertained that neither E_I nor E_H are end signs. Further, an I_c has been isolated on E_H (op. A10 or D15).

N1 I_{cH} is now given the recognition index "j" and its Saturation Func-
N2 tion is examined.

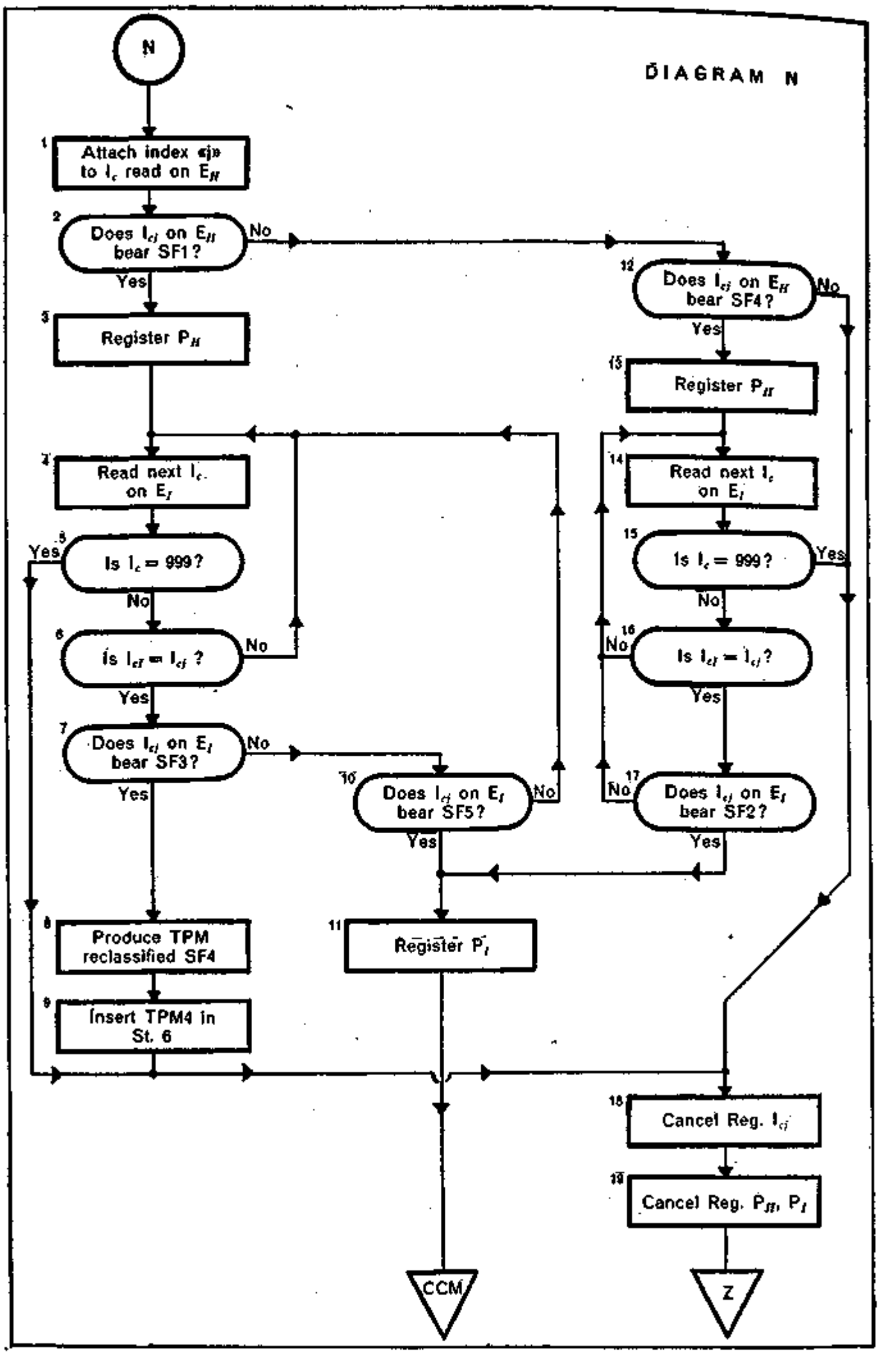


DIAGRAM N
N
1 Attach index «j» to I_c read on E_H
2 Does I_{cj} on E_H bear SF1?
Yes
No
3 Register P_H
4 Read next I_c on E_I
5 Is I_c = 999?
Yes
No
6 Is I_{cI} = I_{cj} ?
Yes
No
7 Does I_{cj} on E_I bear SF3?
Yes
No
8 Produce TPM reclassified SF4
9 Insert TPM4 in St. 6
10 Does I_{cj} on E_I bear SF5?
Yes
No
11 Register P_I
12 Does I_{cj} on E_H bear SF4?
Yes
No
13 Register P_H
14 Read next I_c on E_I
15 Is I_c = 999?
Yes
No
16 Is I_{cI} = I_{cj}?
Yes
No
17 Does I_{cj} on E_I bear SF2?
Yes
No
18 Cancel Reg. I_{cj}
19 Cancel Reg. P_H, P_I
CCM
Z

If its SF is 1, E_H could correlate with an E_I bearing the same I_c (I_{cj}) and SF3 or SF5 (see SF-table, page 49).

N3 In this case the index P (corresponding to the particular I_g string belonging to the I_c string in which I_{cj} has been isolated) is registered for later reference during CCM procedure.

N4 The I_c on E_I are now examined in order to ascertain whether I_{cj}
N6 is among them, and, if so, whether it bears an SF complementary
N7,10 to SF1 shown on E_H.

N8 If SF3 is found (indicating an explicit correlator), a TPM is at once produced and reclassified as a product of I_{cj} bearing SF4 (1st correlatum plus correlator).

N9 Any TPM4 produced is inserted into Store 6.

N10 If the outcome of op. N7 is negative, but I_{cj} on E_i bears SF5, the
N11 index P corresponding to the particular I_g string relevant on E_i is registered for later reference during CCM procedure, and the flow proceeds to op. CCM1. If the outcome of op. N10 is negative, the next I_c on E_i is examined and op. N4-N10 are repeated for this I_{c1}.

N2,12 If I_{cj} on E_H does not bear SF1 but SF4, E_H could correlate with an E_i bearing I_{cj} and SF2; this possibility is tested in operations sim-
N12- ilar to op. N3-N10 and if these tests turn out positive, the flow
N17 proceeds, as before, to op. N11 and thence to op. CCM1.

N18 If the outcome of op. N12 is negative (which is the case if I_{cj} on E_H bears neither SF1 nor SF4), or if in operations N5 or N15 the end sign of the I_c string on E_I has shown up, this Correlation Cycle
N18 has been finished, the registration I_{cj} can be cancelled, and also any
N19 registrations of PH or PI that might have been made can be cancelled; the flow, then, proceeds to operations *Z* where it is re-directed according to the indication given by the Cycle Indicator *QZ*.

DIAGRAMS CCM S1 and CCM S2 (pp. 60, 62)

CCM S1

op. 1 The Control Matrix Procedure begins with the extraction of the relevant CCM, which is the matrix governing correlations made by I_{cj} (see pages 50-2). Diagram S1 shows the procedure concerning Sec-

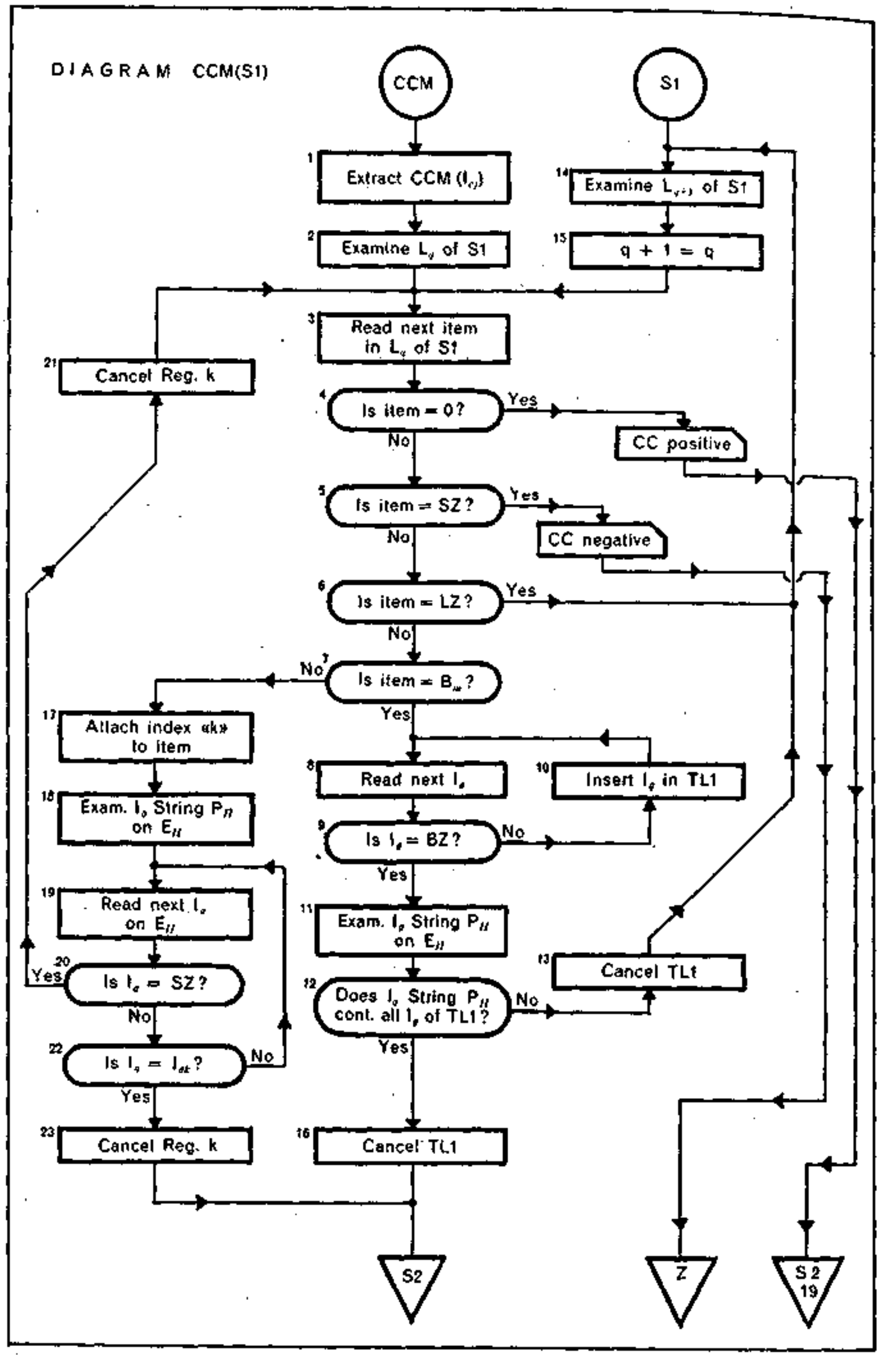


DIAGRAM CCM(S1)
CCM
S1
1 Extract CCM (I_{cj})
2 Examine L_q of S1
14 Examine L_{q+1} of S1
15 q + 1 = q
3 Read next item in L_q of S1
21 Cancel Reg. k
4 Is item = 0?
Yes
No
CC positive
5 Is item = SZ?
Yes
No
CC negative
6 Is item = LZ?
Yes
No
7 Is item = B_w?
No
Yes
17 Attach index «k» to item
8 Read next I_e
10 Insert I_e in TL1
18 Exam. I_e String P_{II} on E_{II}
9 Is I_e = BZ?
No
Yes
19 Read next I_e on E_{II}
11 Exam. I_e String P_{II} on E_{II}
20 Is I_e = SZ?
Yes
No
13 Cancel TL1
12 Does I_e String P_{II} cont. all I_e of TL1?
No
Yes
22 Is I_q = I_{ek}?
No
Yes
23 Cancel Reg. k
16 Cancel TL1
S2
Z
S2 19

tion 1 of the CCM (i.e. the section indicating the requirements for the 1st correlatum); Diagram S2 shows the subsequent procedure concerning Section 2 (i.e. the section indicating the requirements for the 2nd correlatum).

op. 2-3 The first line of S1 is examined and its first item is isolated.

The requirements for an element, to function, as 1st correlatum in the particular correlation, may be indicated by single I_g or by "Blocks" of I_g. An I_g block consists of several single I_g all of which have to figure on the element, if the element is to be accepted. Each block is introduced by the recognition sign B_m and ends with the end sign BZ. Single I_g and I_g blocks, on the CCM, are arranged in lines, so that one line in S1 corresponds to one line in S2. The present procedure provides for correspondences between single I_g and I_g blocks as well as, of course, correspondences between single I_g and single I_g and correspondences between block and block.

op. 4 If the item isolated in S1 is "0", it means that the CCM contains no restrictions of acceptability and that the correlation can be accepted at once. (Diagram S2, op. 19-21).

op. 7-13 constitute the correspondence test if the item found in L_q of S1 is an I_g block;

op. 17-23 constitute the correspondence test if the item found is a single I_g;

op. 14-15 shift the examination to the next line of the CCM if the block test has not yielded a correspondence.

If either of the correspondence tests has given a positive result, the flow proceeds to the set of operations shown in Diagram CCM S2, where the same tests are made with regard to section 2 of the CCM and E_I (the 2nd correlatum).

CCM S2

op. 1-11 correspond to operations 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 16 in Diagram CCM S1;

op. 12-18 correspond to operations 17-23 in Diagram CCM S1.

op. 19 If the outcome of the tests is positive, two product matrices are produced and, according to the indications found in section EC corresponding to the line of the CCM that has produced them, the reclassification is attached to the PM. One of these PM will bear

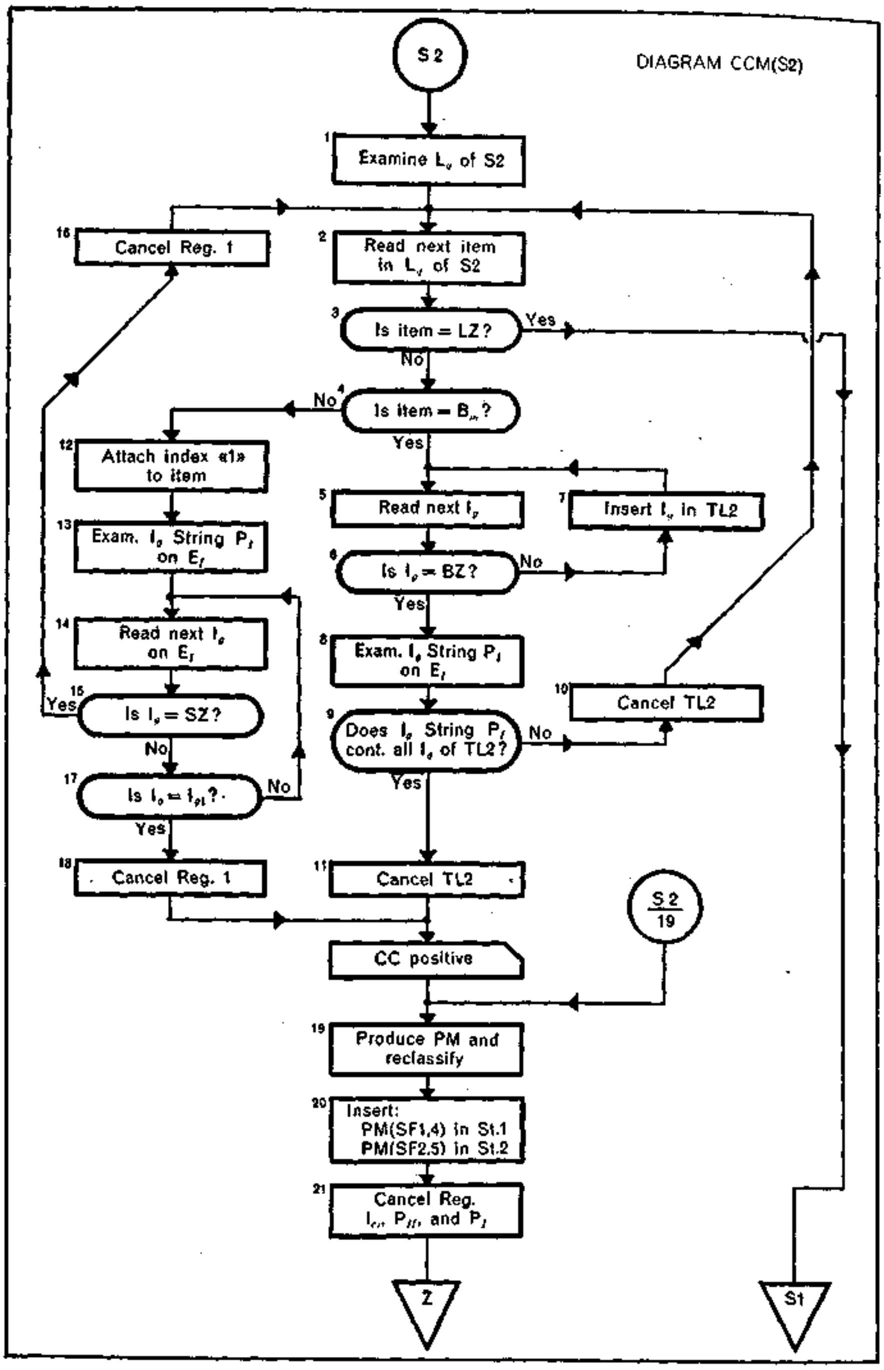


DIAGRAM CCM(S2)
S2
Examine L of S2
Read next item in L of S2
Is item = LZ?
Is item = B?
Cancel Reg. 1
Attach index «1» to item
Exam. I String P on E
Read next I on E
Is I = SZ?
Is I = I?
Cancel Reg. 1
Read next I
Insert I in TL2
Is I = BZ?
Exam. I String P on E
Does I String P cont. all I of TL2?
Cancel TL2
Cancel TL2
CC positive
S2 / 19
Produce PM and reclassify
Insert: PM(SF1,4) in St.1 PM(SF2,5) in St.2
Cancel Reg. I, P, and P
Z
S1
Yes
No

all I_c that show SF1 or SF4, while the other will bear all I_c that
op. 20 show SF2 or SF5. According to their SF, the PM are then inserted either in Store 1 or in Store 2.

op. 21 All temporary registrations (I_{cj}, P_H, P_I) are cancelled and the flow proceeds to operations Z for re-direction.

op. 3 If, however, none of the items in line q of S2 yield a correspondence with E_I and the end sign of the line, LZ, is reached, the same CCM procedure is repeated for the next line of the CCM; that is to say, the flow returns to op. 3 of CCM S1, shifting the examination to the next line by means of op. 14, 15 of Diagram S1.

DIAGRAM Z (p. 64)

shows the re-directions imparted to the flow by means of the Cycle Indicator QZ.

Assuming that the cycle just finished was a CCM Cycle set off by the first Cycle N after the input of a new WM, the indicator QZ will read "NIC" (see op. A13), and this re-directs the flow to operation A10, in order to repeat Cycle N and, if it should be required, Cycle CCM to test the correlational possibilities of a new I_{cH}.

When the last I_{cH} has been dealt with, the end sign of this I_c string directs the flow back to operation A6 and the same procedure is repeated for the next element on track H.

When the last E_H has been dealt with, the end sign of the list present on track H turns up in operation A7, and the flow proceeds, as we have seen, to the operations shown in Diagram B.

DIAGRAM B (p. 65)

This cycle concerns the same WM_n that entered at the beginning of
B1 Cycle A. This WM_n remains in track I, while on track H enter the elements collected in Store 10.

B2-B9 are the same operations as A5-A12;

B10 provides a new setting for the Cycle Indicator QZ, if the tests of Cycle B have given positive results, and from there the flow proceeds to the operations shown in Diagram M.

DIAGRAM Z

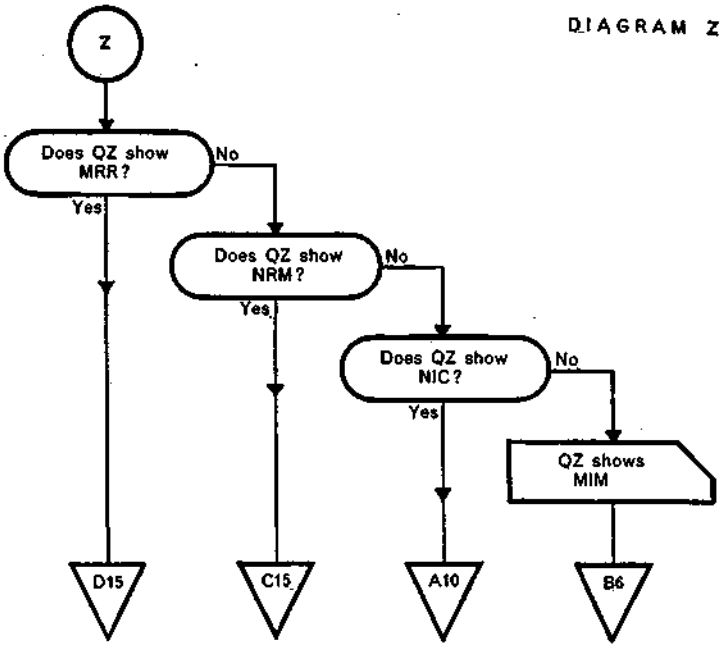
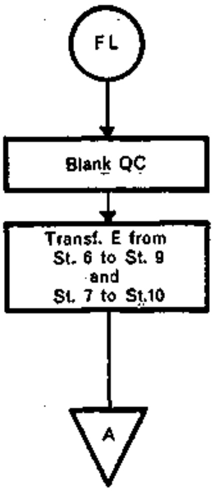


DIAGRAM FL



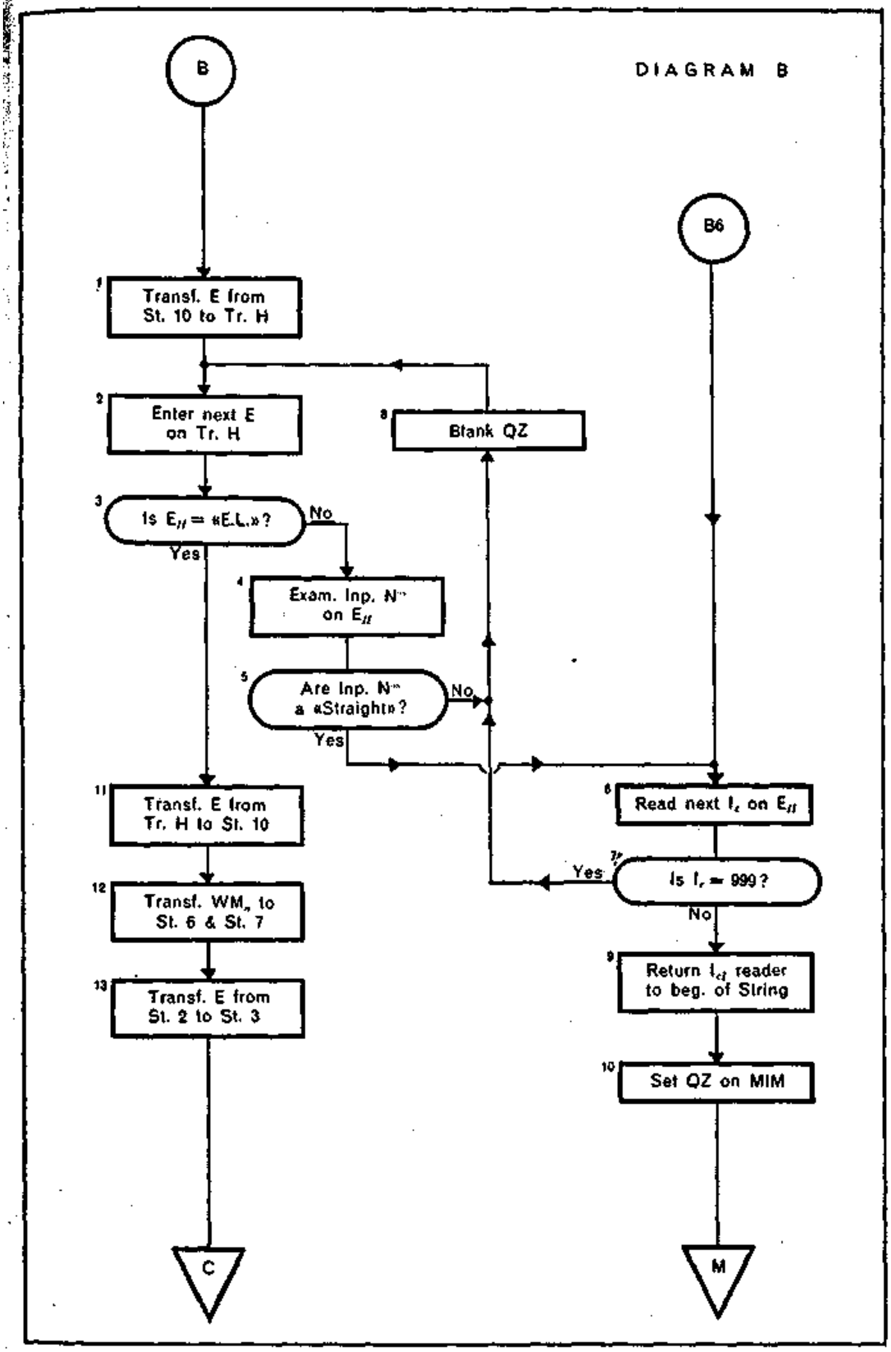


DIAGRAM B
B
B6
1 Transf. E from St. 10 to Tr. H
2 Enter next E on Tr. H
8 Blank QZ
3 Is E = «E.L.»?
No
Yes
4 Exam. Inp. N on E
5 Are Inp. N a «Straight»?
No
Yes
6 Read next I on E
7 Is I = 999?
Yes
No
11 Transf. E from Tr. H to St. 10
12 Transf. WM to St. 6 & St. 7
13 Transf. E from St. 2 to St. 3
9 Return I reader to beg. of String
10 Set QZ on MIM
C
M

B3 When the last element on track H has been dealt with and the
B11 respective end sign turns up in operation B3, the list present in track
B12 H is transferred to Store 10, while the WM_n present in track I is duplicated and inserted both in Store 6 and in Store 7;

B13 concurrently the elements present in Store 2 are transferred into Store 3, and the flow proceeds to operation C1 (Diagram C).

DIAGRAM M (p. 67)

shows the operation of Cycle M, which is the Correlation Cycle for "maintainment" correlations (see page 51). The operations indicated here correspond exactly to those of Cycle N, except that they, of course, concern other saturation functions. Besides, there is an
M2, initial I_c test to ascertain that the I_c under examination does, in
M3 fact, allow for "maintainment". If this test has negative outcome, the entire cycle is by-passed and the flow proceeds to the re-direction operations shown in Diagram Z.

M4 If, however, the present I_c allows for "maintainment", I_{cH} is tested
M11 for SF2, SF3, and SF5, and if any of these tests is positive, EI
M18 is examined to ascertain whether it contains the same I_c (I_{cj}), and,
M9,15 subsequently, whether I_{cj} on E_I bears the complementary saturation
M23 function, which would be SF4, SF2, and SF1 respectively.

M10 If the correspondence SF2/SF4 is found or the correspondence
M24 SF5/SF1, the flow proceeds to the Correlation Control procedure shown in Diagram CCM.

M16 If a correspondence SF3/SF2 is found, a TPM is at once produced and reclassified as bearing I_{cj} and SF5 (correlator plus 2nd correlatum);
M17 Any TPM5 produced is inserted into Store 7.

If a TPM5 has been produced, or if the SF tests concerning E_H have all been negative, all temporary registrations are cancelled (like in op. N18, N19) and the flow proceeds to the re-direction operations shown in Diagram Z.

DIAGRAM C (p. 68)

Cycles C and D insert into the Correlation Cycles (N and M) all elements that have been produced on the present "Level". A "Lev-

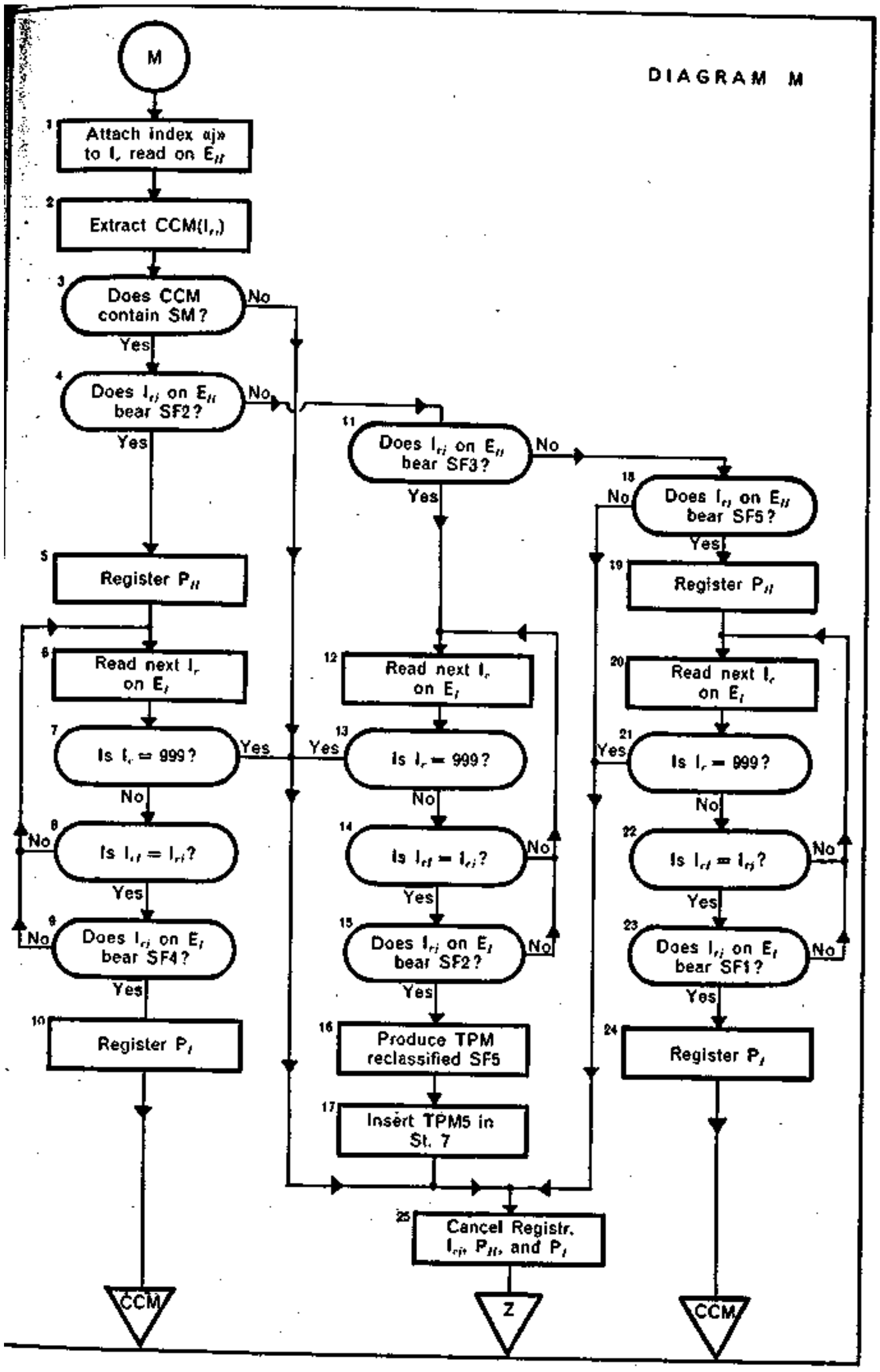


DIAGRAM M
M
Attach index «j» to I_r read on E_{II}
Extract CCM(I_r)
Does CCM contain SM?
Does I_{rj} on E_{II} bear SF2?
Register P_{II}
Read next I_r on E_I
Is I_r = 999?
Is $I_{rI} = I_{rj}$?
Does I_{rj} on E_I bear SF4?
Register P_I
Does I_{rj} on E_{II} bear SF3?
Read next I_r on E_I
Is I_r = 999?
Is $I_{rI} = I_{rj}$?
Does I_{rj} on E_I bear SF2?
Produce TPM reclassified SF5
Insert TPM5 in St. 7
Does I_{rj} on E_{II} bear SF5?
Register P_{II}
Read next I_r on E_I
Is I_r = 999?
Is $I_{rI} = I_{rj}$?
Does I_{rj} on E_I bear SF1?
Register P_I
Cancel Registr. I_{rj}, P_{II}, and P_I
CCM
Z
CCM

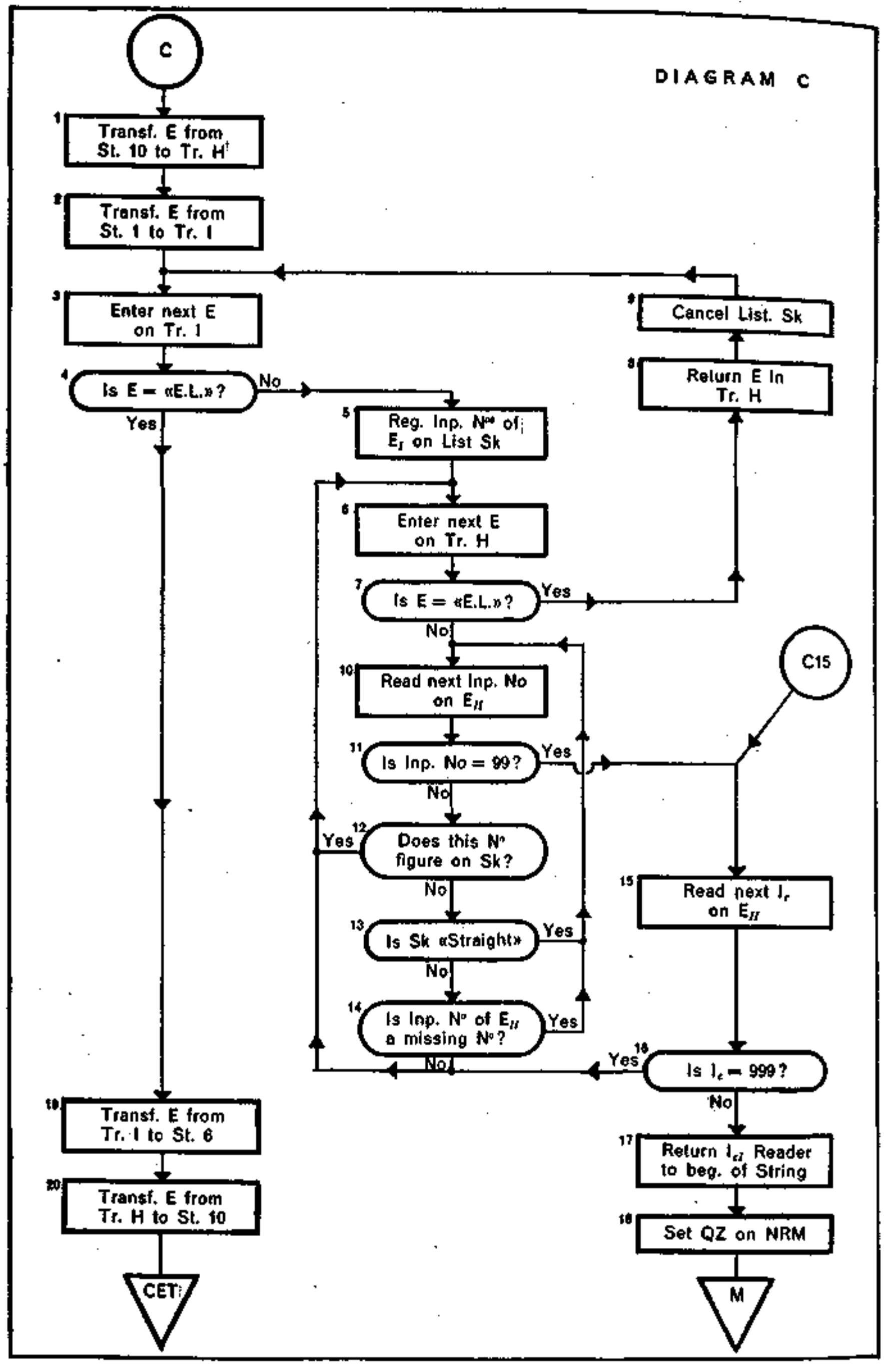


DIAGRAM C
C
1 Transf. E from St. 10 to Tr. H
2 Transf. E from St. 1 to Tr. I
3 Enter next E on Tr. I
4 Is E = «E.L.»?
No
Yes
5 Reg. Inp. Nos of E_I on List Sk
6 Enter next E on Tr. H
7 Is E = «E.L.»?
Yes
No
8 Return E in Tr. H
9 Cancel List. Sk
10 Read next Inp. No on E_II
11 Is Inp. No = 99?
Yes
No
12 Does this N° figure on Sk?
Yes
No
13 Is Sk «Straight»
Yes
No
14 Is Inp. N° of E_II a missing N°?
Yes
No
C15
15 Read next I_r on E_II
16 Is I_r = 999?
Yes
No
17 Return I_cI Reader to beg. of String
18 Set QZ on NRM
M
19 Transf. E from Tr. I to St. 6
20 Transf. E from Tr. H to St. 10
CET

el" is the entire sequence of operational cycles from the input of one WM to the input of the next WM. (They correspond to what, . in previous input procedures, has been called "Recirculation".)

C1 The elements from Store 10 are entered on track H, and on track1 I enter the elements from Store 1, which are the PM that have been reclassified with SF1 or SF4.

C5-C12 Since it is necessary to exclude from the correlation procedure elements which contain a common WM, there is an Inp. N^{os} test. Following this, there is a further "Word Order" test incorporating a
C13-14 restriction on word order in actual English usage.

(Note: if an element is composed of WM whose Inp. N^{os} do not form a "Straight", that is, if they form a sequence of numbers that are not consecutive, such an element can correlate only with an element whose Inp. N^{os} fit into the gap of the sequence. For example, if E_H is composed of $WM_2 + WM_5$, it can correlate only with an E_I constituted by WM_3 and/or WM_4.)

C15-C18. If the input number tests yield a positive result, the next I_c on E_H is examined in operations that correspond to op. A10-A13 and to op. B6-B10, except that the Cycle Indicator is set differently.
C19 If the Inp. N° tests are negative, and the end sign turns up in the list on track I (op. C4), the list present in track I is inserted into
C20 Store 6, while the list present in track H is returned to Store 10, and the flow proceeds to the operations shown in Diagram CET.

DIAGRAMS CET and DET (pp. 70, 71)

Since "Recirculation" continues on one level as long as new PM are produced, a production test has to be made before proceeding to the next level. Diagrams CET and DET show the operations of this test for Cycles C and D respectively.

CET 1-3 Since PM are at once stored in stores 1 and 2, these stores are examined and a Counter QE indicates, by showing "2", that both
CET7 stores have been found empty. A further Counter QC in turn records this result.. If, then, after the subsequent Cycle D (at the end of which
DET there is again a production test — Diagram DET), there are again
1-8 no new PM in the relevant stores, the counter QC reaches position

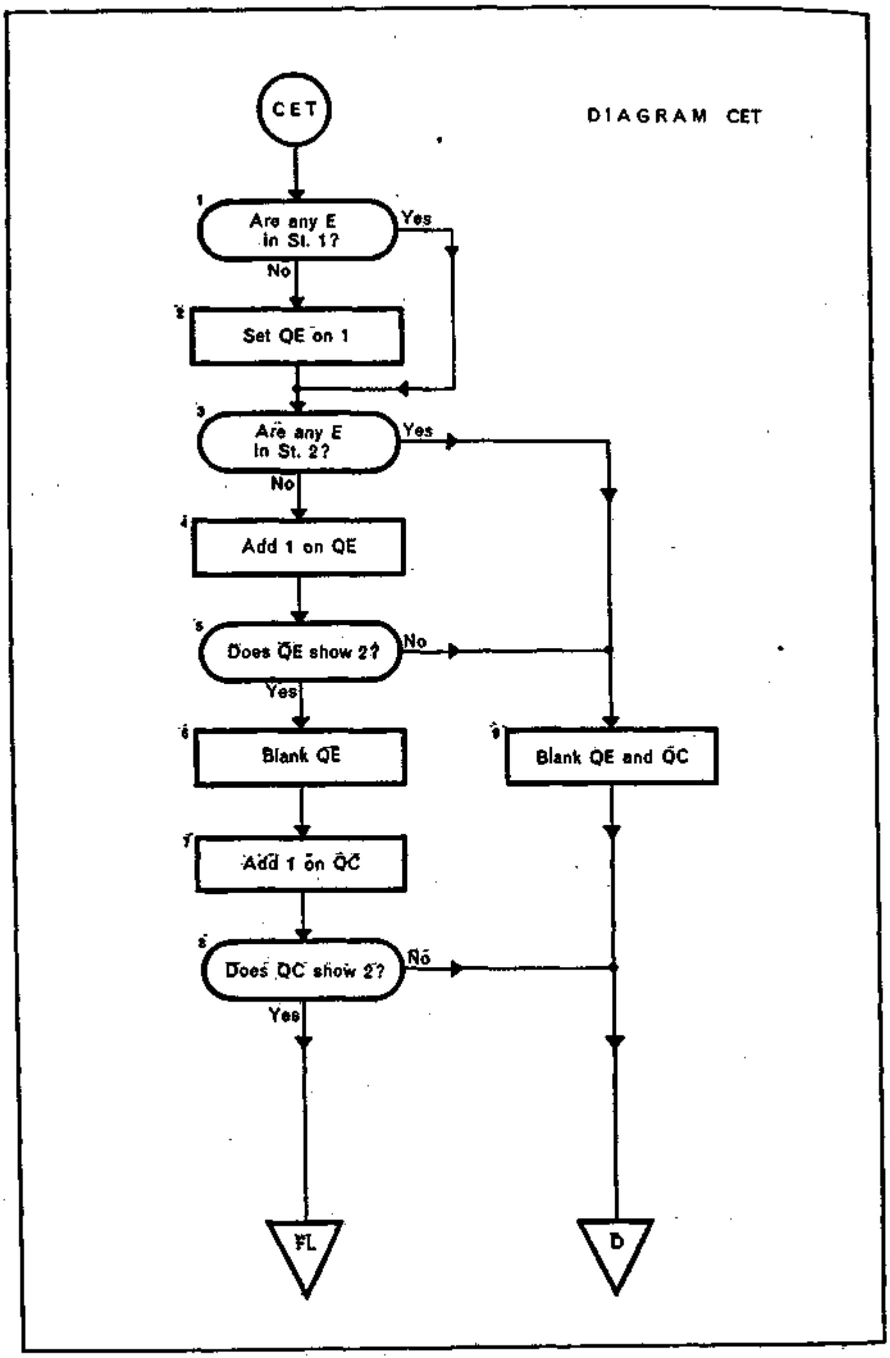


DIAGRAM CET
CET
1 Are any E in St. 1?
Yes
No
2 Set QE on 1
3 Are any E in St. 2?
Yes
No
4 Add 1 on QE
5 Does QE show 2?
No
Yes
6 Blank QE
9 Blank QE and QC
7 Add 1 on QC
8 Does QC show 2?
No
Yes
FL
D

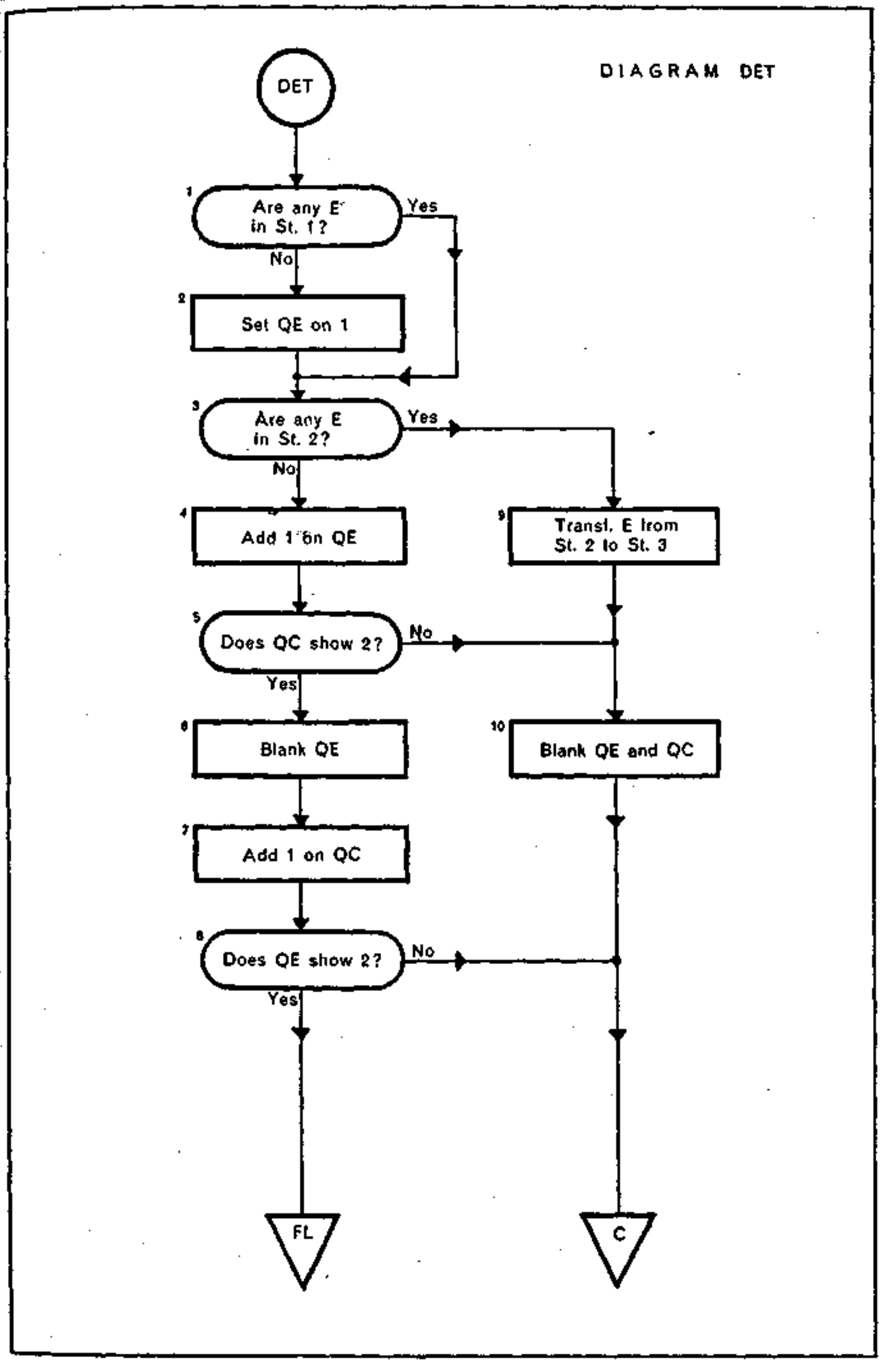


DIAGRAM DET
DET
1
Are any E in St. 1?
Yes
No
2
Set QE on 1
3
Are any E in St. 2?
Yes
No
4
Add 1 on QE
9
Transl. E from St. 2 to St. 3
5
Does QC show 2?
No
Yes
6
Blank QE
10
Blank QE and QC
7
Add 1 on QC
8
Does QE show 2?
No
Yes
FL
C

"2" and the flow proceeds to the operations that constitute the end of the level (Diagram FL). On the other hand, if either in Cycle C or in Cycle D new PM are produced, QC returns to its zero position and the flow proceeds to the other of the two alternating Cycles C and D.

DET9 The difference between Cycles CET and DET lies in the storage order contained in the latter. If, when DET occurs, there are elements in Store 2, these elements are transferred into Store 3.

DIAGRAM D (p. 73)

Cycle D contains the same operations as Cycle C, but they are carried out with other elements, and therefore the insertion instructions at the beginning and the storage instructions at the end are different.

D1 In Cycle D the list from Store 11 enters on track H, while on track I
D2 the elements from Store 3 introduced; and at the end, the list pre-
D19 sent in track I is transferred to Store 7, while that present in track
D20 H is returned to Store 11.

DIAGRAM FL (p. 64)

When the Cycle Counter QC reaches position "2" (which is the case when two successive cycles C and D, or D and C, have yielded no new products), the "end of Level" operations are carried out. They consist in blanking the Cycle Counter (returning it to its zero position) and in transferring the E from Store 6 to Store 9 and the E from Store 7 to Store 10.

The flow then returns to Cycle A in order to repeat the entire procedure for the next WM which enters in operation A1. If this WM is identified (op. A2) as the end sign of the input sentence, the "End of Input" operations are carried out (Diagram FI).

DIAGRAM FI (p. 75)

The "End of Input" operations consist in a simple test of the products collected in the permanent stores 9, 10, and 11. This test,

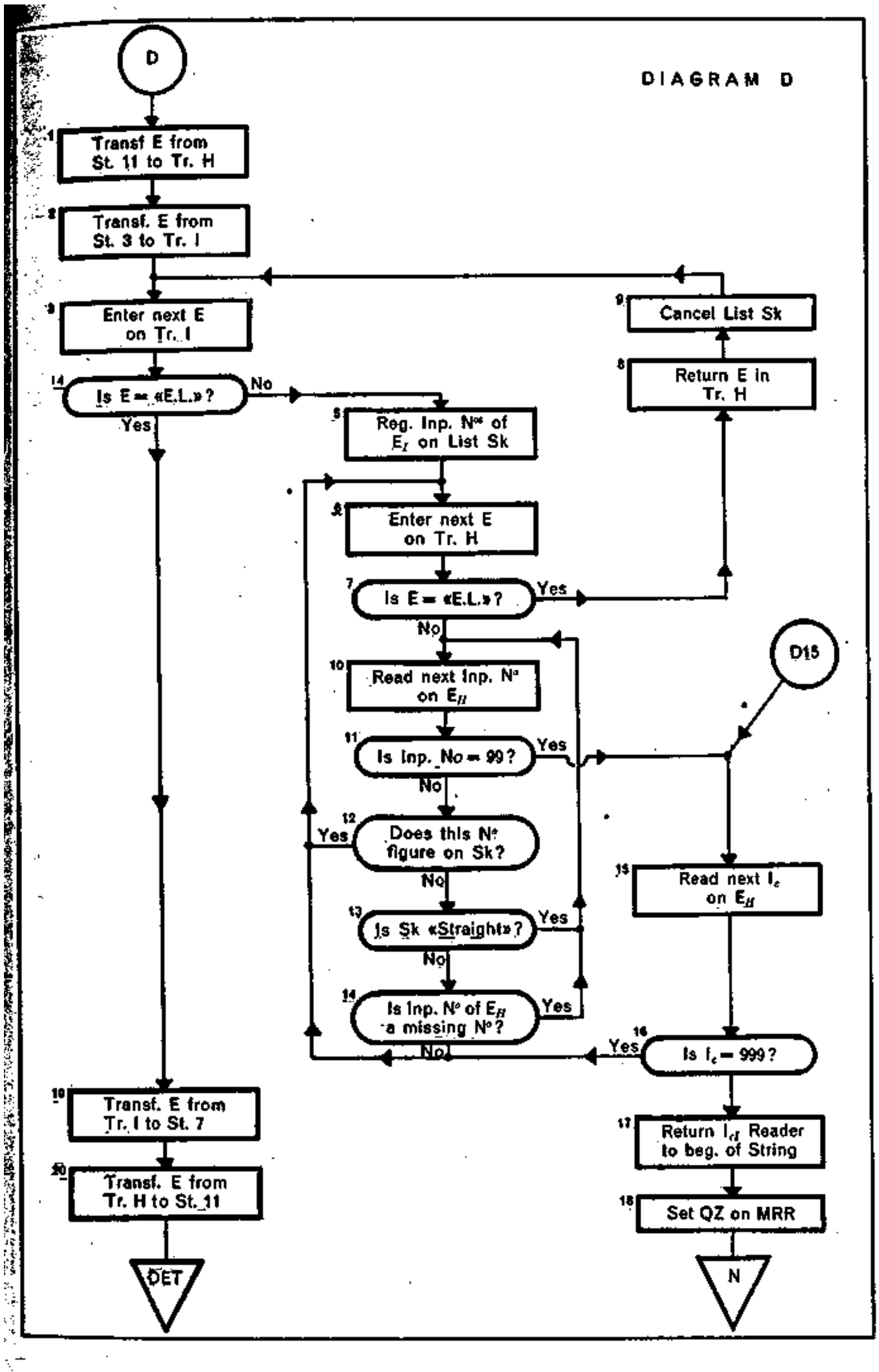
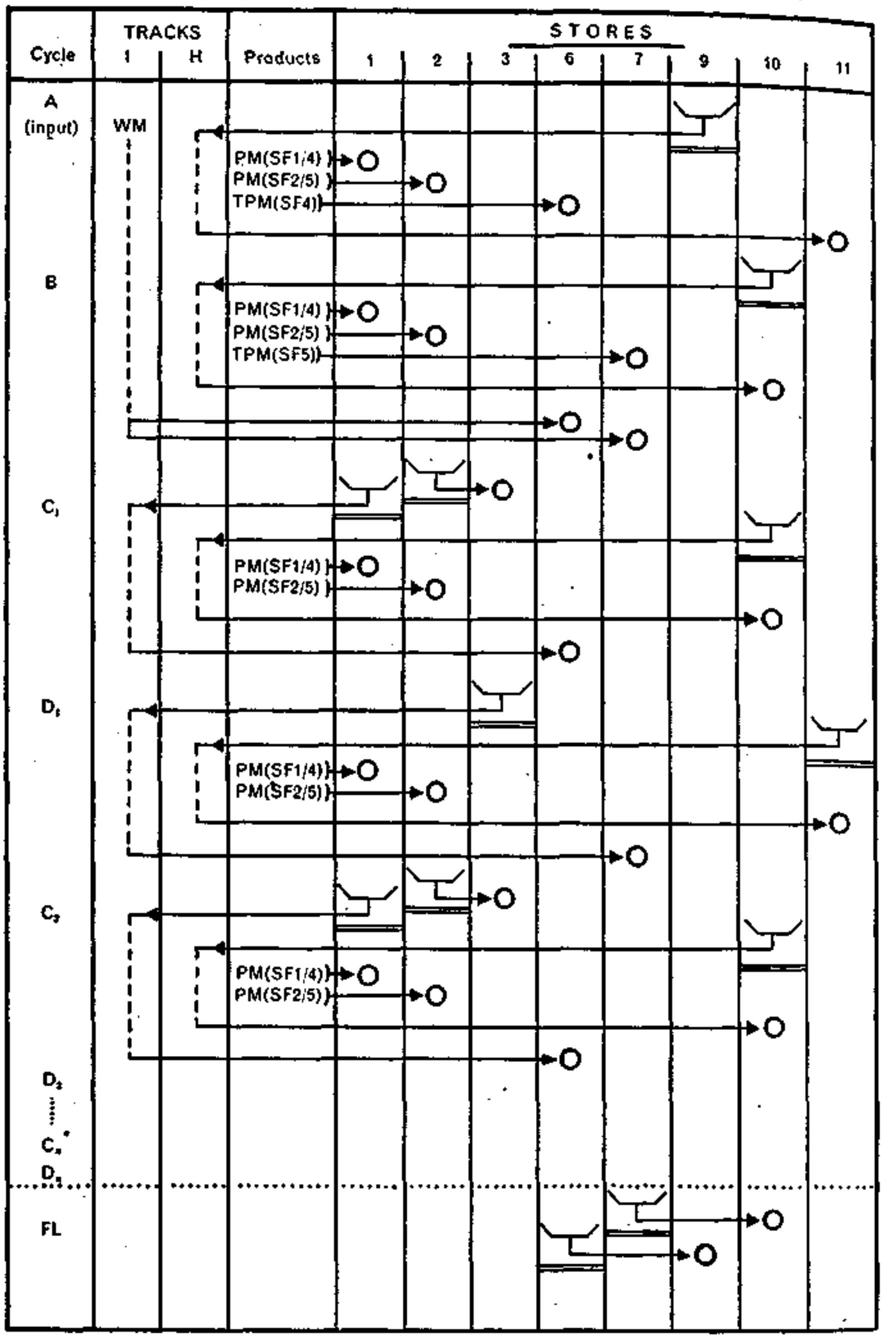

D
DIAGRAM D
Transf E from St. 11 to Tr. H
Transf. E from St. 3 to Tr. I
Enter next E on Tr. I
Is E = «E.L.» ?
No
Yes
Reg. Inp. N°s of E_I on List Sk
Enter next E on Tr. H
Is E = «E.L.» ?
Yes
No
Return E in Tr. H
Cancel List Sk
Read next Inp. N° on E_H
Is Inp. No = 99 ?
Yes
No
D15
Does this N° figure on Sk?
Yes
No
Is Sk «Straight» ?
Yes
No
Is Inp. N° of E_H a missing N° ?
Yes
No
Read next I_c on E_H
Is I_c = 999 ?
Yes
Return I_{cI} Reader to beg. of String
Set QZ on MRR
Transf. E from Tr. I to St. 7
Transf. E from Tr. H to St. 11
DET
N

DIAGRAM T


Cycle
TRACKS
I
H
Products
STORES
1
2
3
6
7
9
10
11
A (input)
WM
PM(SF1/4)
PM(SF2/5)
TPM(SF4)
B
TPM(SF5)
C1
D1
C2
D2
Cn
Dn
FL

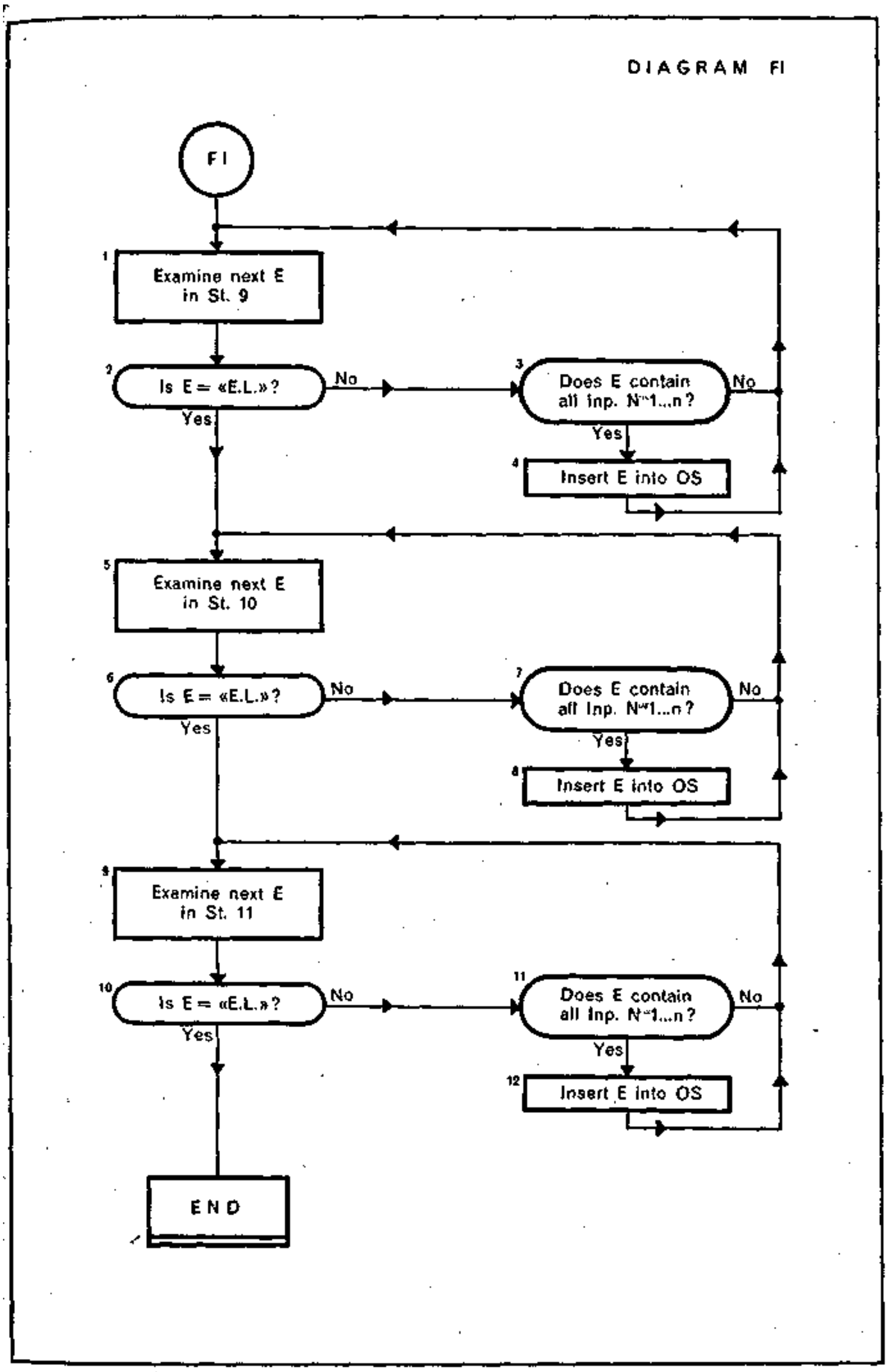


DIAGRAM FI
FI
1 Examine next E in St. 9
2 Is E = «E.L.»?
No
Yes
3 Does E contain all inp. N° 1...n?
No
Yes
4 Insert E into OS
5 Examine next E in St. 10
6 Is E = «E.L.»?
No
Yes
7 Does E contain all Inp. N° 1...n?
No
Yes
8 Insert E into OS
9 Examine next E in St. 11
10 Is E = «E.L.»?
No
Yes
11 Does E contain all Inp. N° 1...n?
No
Yes
12 Insert E into OS
END

applied in the same way to the three stores (op. FI1-FI4, FI5-FI8, and FI9-FI12) selects the products that contain *all WM of the input sentence* and inserts them into the "Output Store".

Conclusion

The criterion of "Completeness" applied in the "End of Input" procedure is certainly the most important one for the selection of the produced material that is to enter output procedure; but it is neither the only one — in the sense that it will assure sufficient selection of the produced material — nor will it assure, under all circumstances, that every input sentence actually yields a correlational net for output.

With regard to the first point — its insufficient selectiveness — we plan to supplement it by means of a further test procedure concerning the semantic aspect of the produced nets. In the present procedure the semantic indices I_s are consulted only with regard to the legitimacy of formally possible correlations; it is obvious that they can be used also to determine a choice between several complete nets, a choice which would take into account the semantic characteristics of *previous* sentences, in other words, of *context.* But, as I mentioned at the beginning (see pages 47-8), I am here presenting a first draft of the procedure applicable only to single sentences.

With regard to the second point — the fact that the above completeness test will prevent certain sentences from reaching output — the problem arises when a sentence, although perfectly comprehensible for the human reader or listener, does not explicitly contain all the elements required to construct the correlational net expressed by it.

For instance, the sentence "the Prime Minister's mistress dies at the end of the last act, his wife at the beginning of the first" cannot yield a complete correlational net unless at least the verbal form "dies" is repeated in the second part of the sentence. This requires what I should call a "Retrieval Procedure". Preliminary research has reassured us that subroutines for this kind of retrieval are not quite as impossible as they might seem at first sight. Since authors, as a rule, are fairly careful not to overtax their readers' capacity of correlation, they avoid formulations that require random retrievals and limit themselves to expressions in which retrieval, if it occurs at all, is governed by certain definite rules based on formal or semantic indications actually present in the sentence. Research in this direction is being continued and we are confident that a more complete input procedure containing also subroutines for retrieval of implicit elements can be developed.

In the course of research we have also seen that a number of further subroutines can be developed in order to reduce even more the number of correlation attempts whose results have to be discarded at later stages (see page 52). The more subroutines of that kind can be applied to the basic procedure, the faster, the more reliable and, above all the more manageable it will become; and in programs involving, as does translation procedure, not hundreds but hundreds of thousands of operations, any reduction of the operational volume is worth being applied.

ABBREVIATIONS

A	Diagram A
B	Diagram B
B_m	Block of I_g on CCM
BZ	End sign of I_g block
C	Diagram C
CCM	Correlation Control Matrix, Diagram CCM
CET	Production Test, Cycle C, Diagram CET
D	Diagram D
DET	Production Test, Cycle D, Diagram DET
E	Element
E_H	E on track H
E_I	E on track I
"E.I."	End of Input
"E.L."	End of List
FI	End of Input Diagram
FL	End of Level Diagram
H	Track leading to combination procedure
I	Track leading to combination procedure
I_c	Index of correlation
I_{cH}	I_c on Element H
I_{cI}	I_c on Element I
I_{cj}	I_c registered for operations
I_g	Grammatical Index
I_{gk}	I_g registered for operations
I_{gt}	I_g registered for operations
Inp. N°	Input Number
I_s	Semantic Index
L	Line of CCM
LZ	End sign of line
M	Diagram M
MIM MRR NIC NRM	Indication Symbols of combination cycle
N	Diagram N
OS	Output store
P	Particular I_g string
P_I	I_g string on E_I
P_H	I string on E_H
PM	Product Matrix
q	Index of CCM line
QC	Cycle Counter
QE	Element Counter
QI	Input Counter
QZ	Cycle Indicator
RC	Reclassification Section on CCM
S	Section of CCM
S_1	S corresponding to first correlatum
S_2	S corresponding to second correlatum
SF	Saturation Function
SM	CCM section for "maintainment" correlations
SN	CCM section for normal correlations
"Straight"	Sequence of consecutive N°s
St.	Store
SZ	End sign of I_g string
T	List Transfer Diagram
TL	Test List
TPM	Temporary Product Matrix
Tr.	Track
WM	Word Matrix
Z	Re-direction Diagram
99	End sign of Inp. N°s
999	End sign of I_c string

GRAMMATICAL CLASSIFICATIONS

Recognition Index	2nd figure		3rd figure	
1	*Person*		*Number*	
	irrelevant	0	irrelevant	0
	1st person	1	singular	1
	2nd person	2	plural	2
	3rd person	3	sing. & plural	3
	1st & 2nd	4		
	1st & 3rd	5		
	1st, 2nd, & 3rd	6		
2	*Gender*		*Case*	
	irrelevant.	0	irrelevant	0
	masculine	1	nominative	1
	feminine	2	genitive	2
	neuter	3	dative	3
	masc. & fem.	4	accusative	4
	masc. & neuter	5	dat. & acc.	5
	fem. & neuter	6	nom., dat. & acc.	6
	masc., fem. & neut.	7		
3	*Aspect*		*Tense*	
	irrelevant	0	irrelevant	0
	indefinite	1	present	1
	continuous	2	past	2
	perfect	3	future	3
	perfect contin.	4	future in past	4
			present & past	5
4	*Voice*		*Mood*	
	irrelevant	0	irrelevant	0
	active	1	infinitive	1
	passive	2	indicative	2
	active & pass.	3	subjunctive	3
			participle	5
			imperative	6
			supine	7
			imper. & supine	8
5	*Modifier*		*Degree*	
	irrelevant	0	irrelevant	0
	adjective	1	positive	1
	adverb	2	comparative	2
	adj. & adv.	3	superlative	3
	possessive demonstrative interrogative distributive	4		
	pronoun	5		