

[Appendix to paper by Margaret Masterman,

The Potentialities of a Mechanical Thesaurus]

APPLICATION OF THE LATTICE SYNTAX-PROGRAMME TO THE  
TRANSLATION OF THE ITALIAN PARAGRAPH, SHEWING IN  
DETAIL ITS TRANSFORMATION FROM OUTPUT I TO OUTPUT II

by

A.F. PARKER-RHODES

CAMBRIDGE LANGUAGE RESEARCH UNIT

OCTOBER 1st, 1956.

20, Millington Road,

Cambridge.

METHOD OF OPERATING LATTICE SYNTAX PROGRAMME

1. Each language unit in the input text, having been identified in the bilingual-pidgin dictionary (Dictionary I), is there associated with a binary number called its Lattice Position Indicator. Each L.P.I, contains 6 bits.
  - (a) The L.P.I of any language unit depends upon its grammatical function, as this is initially defined by the bilingual pidgin dictionary (example: L.P.I. for verb, 29).
  - (b) It may be necessary to allow a single input word to be given alternative LPI's in Dictionary I, but it is hoped that this may be avoided.
2. Each structural word-group in the source text is then represented, in the manner explained below, by what is called a lattice group. [It is so called because all such groups can be shown to have the structure of a lattice under a certain linguistically defined partial-ordering relation].
  - (a) The number of possible lattice groups is finite (and well under 1000).
  - (b) Of these, those occurring in each source language form a set specific to that language, containing under 100 members. This is the lattice inventory, (given on pp. 3&4),
  - (c) Each member of the lattice inventory is identifiable by the LPI's which it contains - irrespective of order.
3. As the source text is read, the unit LPI's are collected, and after each new accession the list to date is compared with the Lattice Inventory. As soon as any terminal portion of the list of LPI's is found to coincide exactly with one entry in the L.I., [and as soon as it is checked that the latest LPI is not part of another lattice group], the lattice group thus identified is closed.
  - (a) On closure, a Boolean operation  $\Phi_{39}$  is performed on the LPI's included, and the product of this either (i) represents the function of the Lattice group in the sentence as if it were the LPI of a single word, or (ii) if it ends in ...11 (i.e. = 3 mod. 4) its first 3 digits and last 3 digits are interchanged, and this product becomes the LPI of the group.
  - (b) The resulting LPI thus calculated is put in the list of LPI's in place of those subsumed in it. Thereafter the group is treated just like a single word.
4. When the whole existing list of current L.P.I's forms a lattice group which has the resultant 63 (111 111), a complete sentence has been constructed.
5. When a lattice group has been provisionally identified, and the last LPI in it is found to form another lattice group in combination with subsequent accessions, the latter group has to be closed first. Therefore it is necessary to keep several collections of LPI's, each of which can only yield a lattice group when those beyond it in the series are all empty. When one of these LPI-collections is wholly used up in one lattice group, its resultant L.P.I. is entered in the preceding collection.
6. The operation  $\Phi_{39}$  ( =  $\Phi_{100111}$  ) consists in performing I on the 1st, 4th, 5th, and 6th bits and 0 on the 2nd, and 3rd bits, where I and 0 are defined by the tabular definitions:-

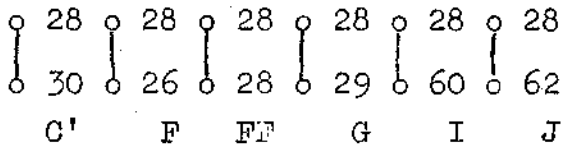
<u>I</u>	0	1		<u>0</u>	0	1
0	0	1		0	0	0
1	1	1		1	0	1

7. Each unit must be labelled, with the LPI's attached to each lattice group of which it is a member. (These groups form a hierarchy culminating in the sentence.) Then, in forming the target sentence, each lattice group can be identified, and to each such group in The Lattice Inventory is attached a special set of instructions for re-ordering its constituents, (cf. Anataxis Procedure, given on p. 8) according to the grammar of the target language.
  
8. The LPI's run from 0 to 63. It is proposed to ignore the 1st bit so that they run from 0 to 31; and then represent each lattice group by a computer-word in which each LPI "n" present is represented by a 1 in the "n"-th bit, other bits being 0. [If any bits are available after this is done, they may be used to indicate whether any LPI is duplicated in the lattice group, or whether any included LIT is > 31.]

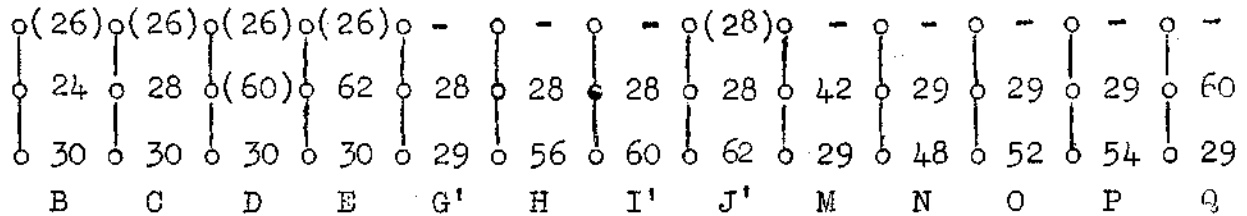
# LATTICE INVENTORY

## A. STRUCTURAL DIAGRAMS OF LATTICES

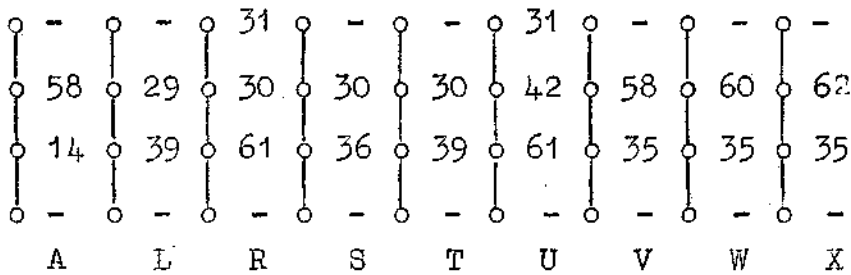
### TWO ELEMENT CHAINS:-



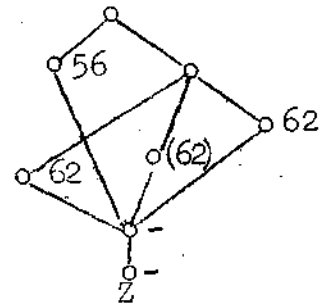
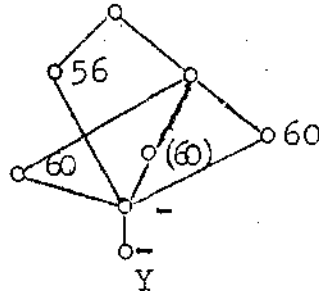
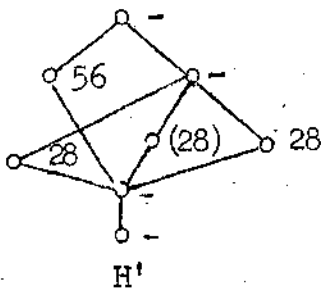
### THREE ELEMENT CHAINS



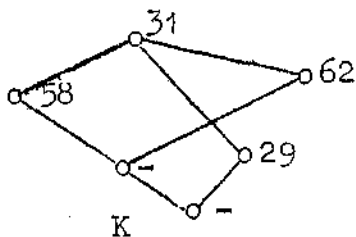
### FOUR ELEMENT CHAINS



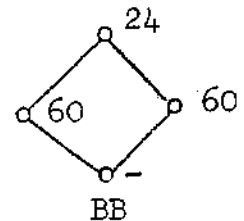
### \* INVENTOR TYPE LATTICES



### DYADIC TYPE LATTICES:



### COMPOUND TYPE LATTICES



(\*) The logical interpretation of these lattices is explained in Progress Report II, Annexe V, (p.12) of the Cambridge Language Research Unit.

B. SUMMARY OF ANATAXIS PROCEDURE FOR ENGLISH

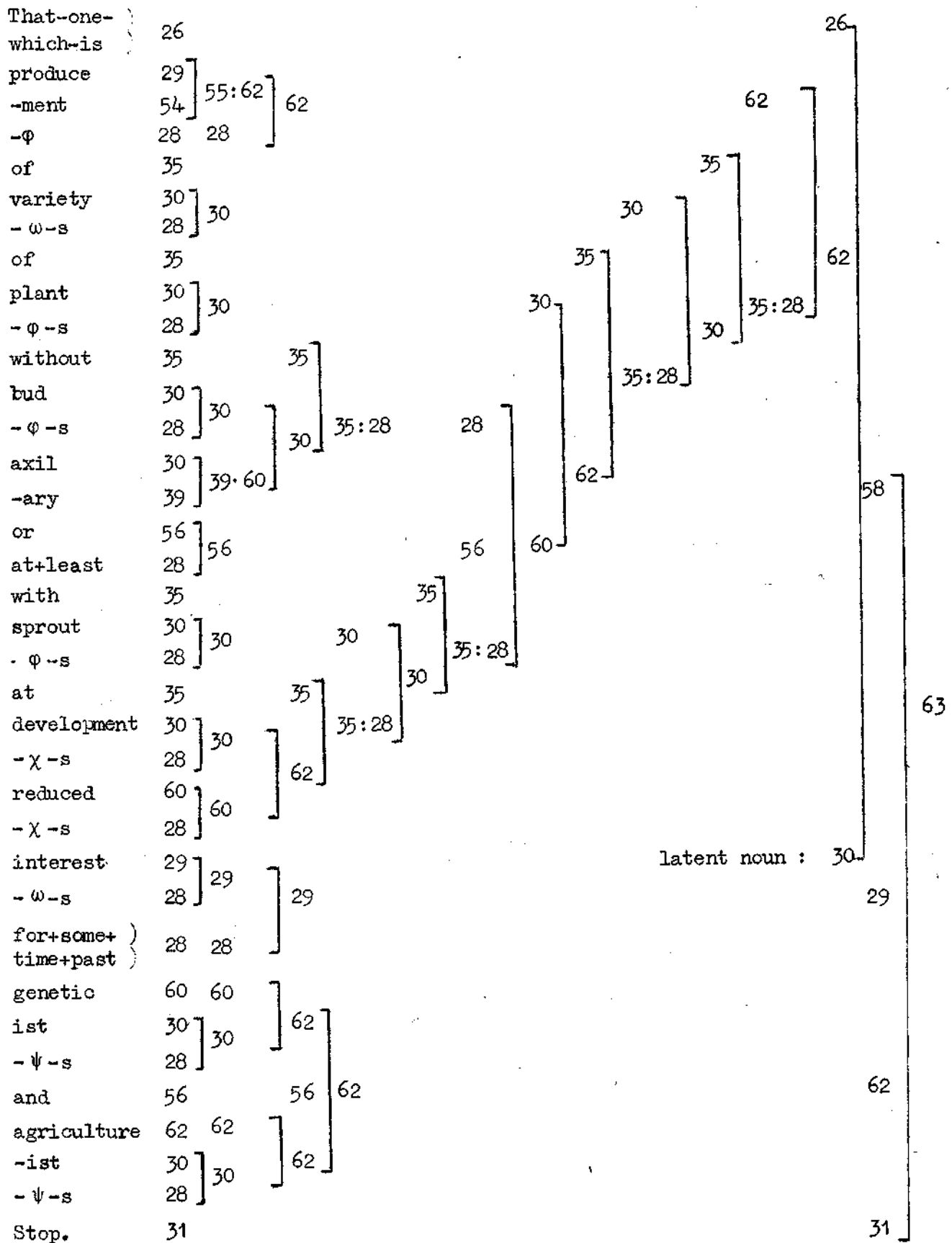
Lattice	Grammatical Form	Anataxis
A	Infinitive clause	14, 58
B	Pro-adjective + noun	24, (26), 30
BB	Intersected nouns	60, (60), 24, 60
C	Adjective + noun	if 28 is single unit : (26), 28, 30 if 28 is a lattice : (26), 30, 28
C	Inflected noun	30, then number-sign if present, otherwise omit 28 altogether.
D	Adjective + noun	if 60 is single unit : (26), 60, 30 if 60 is a lattice : (26), 30, 60
E	Apposed nouns	if 30 manifest in source <u>and</u> target: (26), 62,30 if 30 latent in source and target (26), 62, omit 30 if 30 latent in source but manifest in target: 62, omit 26 and 30
F	inflected pronoun	26, omit 28 (see note 1)
FF	inflected adverb	28, omit second 28
G	inflected verb	29, omit 28 but retain any sign of tense here
GG'	or qualified verb	29, 28.
H	qualified junctor	56, 28
H'	joined adverbs	28, (28), 56, 28 (28's to retain order)
I	inflected adjective	if 28 is inflectional: 60, omit 28
I'	or qualified adjective	if 28 is single unit: 28, 60 if 28 is a lattice : 60, 28
J	inflected noun	62, omit 28
K	dyadic free clause	58, 29, 62 31 (58, 62 to keep the order given 62 58 in source text)
L	participle	29, 39 (but many cases will be unanalysable in English)
M	verb-pronoun compound	29, 42 (but most cases will be unanalysable in English;
N	verb infinitive	"to", 29, omit 48
O	Verb-noun	29, 50 (but most cases unanalysable)
P	verb-noun	29, 54 (but most cases unanalysable;
Q	predicative adjective	proper tense of "be", 60, omit 29
R	monadic free clause	30, 61, 31
S	prepositional clause	if 35 is "'s": 30, 35 otherwise : 35, 30
T	adjective from noun	30, 39 (but many cases unanalysable)
U	monadic free clause (with infinitive as subject)	"it", 61, 42, 31
V	prepositional clause	35, 58
W	adverb from adjective	60, 35 (see note 2)
X	prepositional clause	35, 62
Y	joined adjectives	60, (60), 56, 60 (60's to retain order)
Z	joined nouns	62, (62), 56, 62 (62's to keep order).

Note 1: This is our suggestion, but it might be more convenient to treat English pronouns (other than articles) as declinable.

Note 2: This works when 35 is "-ly", by far the commonest case, as well as for "-wards" &c.; but it may not be possible to avoid programming for alternatives throughout.

LATTICE ANALYSIS  
OF FIRST ITALIAN PARAGRAPH

Sentence 1



Such	24	24							
problem	30			30					30
-o-s	28		30						
self-	42	42							
present	29			43:29				29	
-o-s	28		29						
particular	60								
-ly	35		35:28	28					
interest	29				60				
-ing	39		39:60		60			60	
-o-s	28	28							
for	35					35			
some	26				26				
-o-s	28		26						
essence	30								
-o-s	28		30		30				
forest	30								
-y	39		39:60		60				
-ψ-s	28	28							
and	56			56		60			
fruit	30								
-bearing	39		39:60		60				
-ψ-s	28	28							
for	35					35			
that-which-is	26			26					
plant	30						35:28	28	
-o-s	28		30	30		30			28
of	35	35							
fibre	62								
-o-s	28		62		35:28				
but	56							56	
above+all	28		56						
for	35	35							
that-1-which-is	26								
tobacco	62								
latent noun:	30		58		35:28			28	
Stop.	31								31





The procedures specified in the last column of the preceding table are those which, applied to the elements of the lattice group identified in the first column, will produce therefrom a set of elements corresponding one by one in the right order with the "words" of a grammatical English sentence or other word-group having the same lattice structure. These "words" are not necessarily ordinary dictionary words: for example, phrases such as "at once" may be counted as single words, and unvarying terminations such as the "-ing" of verb forms may be counted as separate words. To show how these procedures work out in practice the best way is to take one or two particular examples from the given sentences.

Example 1: "Such problems"

Here the second "word" problems is itself the representative of a lattice, the first word being a single "chunk" as delivered by the source text; it never matters whether the elements worked with are simple or in any degree complex structures. "Such" comes in with the L.P.I. 24 attached to it, and "problem-s" has for its L.P.I. the figure calculated by the contraction of its lattice, which is 30. Immediately this pair of LPIs has been read the lattice inventory is searched to see if there is a lattice of which these are the elements. In this case there is found to be one and only one lattice of this kind, namely B. It may or may not have a third element 26; strictly speaking we should therefore examine the next "chunk" to see if this is a 26, or if not whether it could be incorporated, after however many intermediate stages, in some bigger lattice with this LPI as its contraction. In this case there is only one such lattice, F, having 28 as its only other element, which is itself the contraction of only one lattice FF made up of two 28's; since the next LEI to be read in happens to be a 42 we could at once dismiss the possibility that our B has an element 26. Actually, we can cut out this process by using the rule (true for Italian but not for all languages) that 26 if it occurs at all comes first in its lattice group. Having decided that "such problem-s" completes a lattice group B, we next form the contraction of the lattice, which is in Boolean notation  $\Phi_{39}(24, 30)$ ; this can be evaluated as 30, the working being set out as below:

Binary notation for 39 :	1	0	0	1	1	1	
operation, <u>0</u> or <u>1</u> , specified							
for each digit by $\Phi_{39}$ :	<u>1</u>	<u>0</u>	<u>0</u>	<u>1</u>	<u>1</u>	<u>1</u>	
Binary notation for operand 24 :	0	1	1	0	0	0	
Binary notation for operand 30 :	0	1	1	1	1	0	
Result of op. in line 2 on these :	0	1	1	1	1	0	= 30

We now turn to the anataxis specified for lattice B : this is given in the table as "24, (26), 30", meaning that we take in the first place the word or group attached to the LPI 24, next that attached to 26 if it is present, and lastly that attached to 30. In our case this means that we have to put "such" first and "problem-s" second.

Example 2 : The first sentence.

As a slightly more complicated example, let us take the whole of the first sentence. It is not more complicated because it is a whole sentence; in fact that makes it easier because when we come to the full stop we know that we have got a complete lattice, or should have, and if there is no lattice in the inventory with the elements which we have got available then one or other of these must be wrong. Actually when we reach this stage we find we have the following elements: "That-one-which-is ... development-s" with LPI computed as 58, "interest ... past" with LPI computed as 29, "geneticist-s and agriculture-ist-s" with LPI computed as 62, and the Full Stop with LPI 31. These four elements correspond to the lattice K in the inventory, and their contraction is easily seen by the working as above to be 63, which is the proper LPI for a free clause.

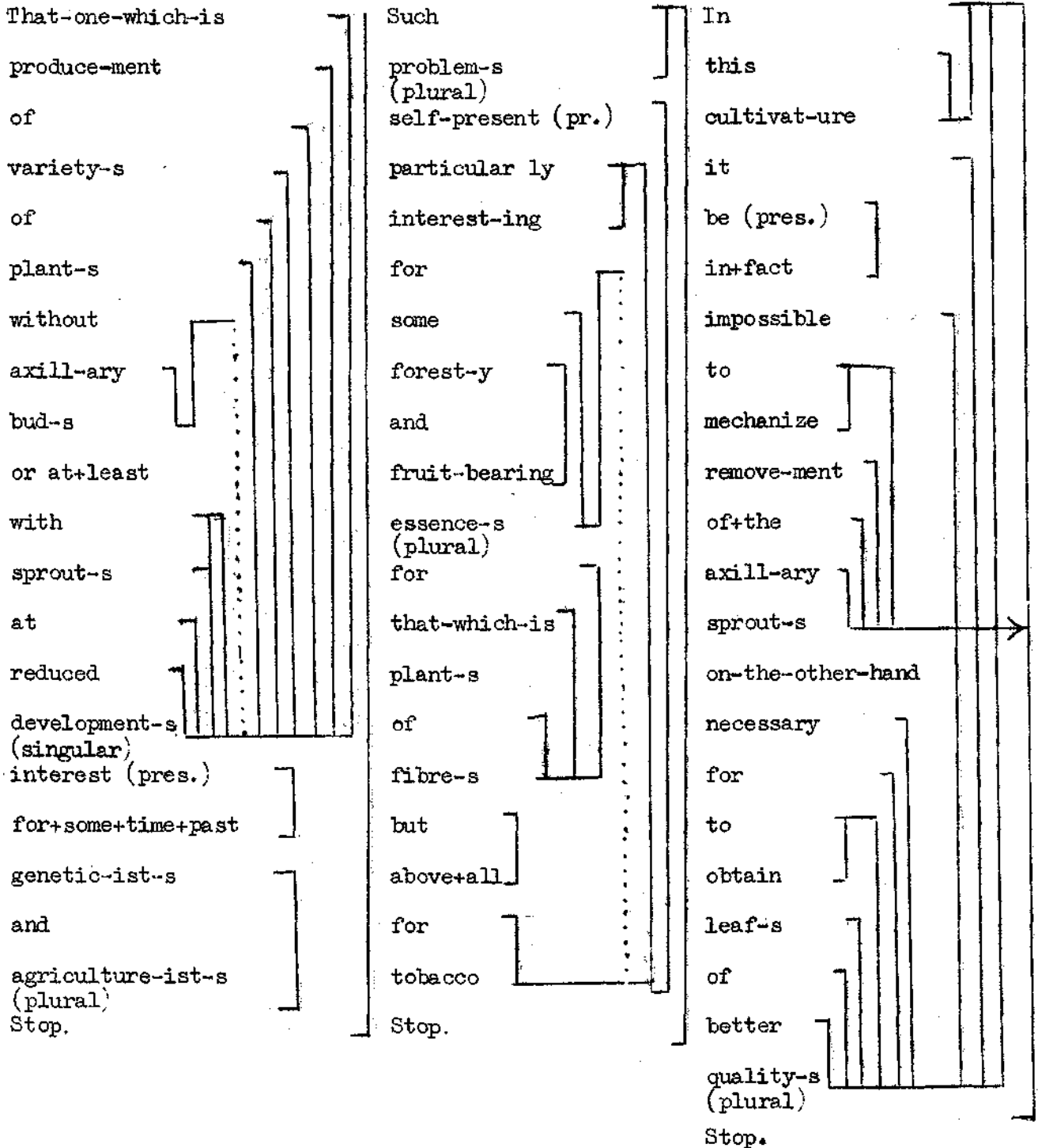
The anataxis specified for K is "58/62, number-sign, 29, 62/58, 31: the 58 and the 62 to retain order as in source the number-sign to be that for the first of these". In the source the 58 comes first, so we have to put first "That-one- ..... development-s", next the number-sign of this lattice-group which by definition is that of the lowest element which has one. The lattice corresponding to this first element will have already been identified as K, in which the lowest element is latent and the next above is itself a lattice J' whose lowest element is the group "producement" with LPI 62. This has been identified as singular; therefore the required number-sign is singular, and in the second place of our K lattice-group we must put this sign. In the third place we put the verb-group 29 "interest (pres.) for some time past", in the 4th place the other noun-group 62 "genetic-ist-s and agriculture-ist-s", and in the fifth place the Full Stop 31. It will be seen that this is the result tabulated in the output.

These examples should be sufficient to show the method of working. Following in this way the instructions given by the anataxis procedures the whole structure of the source passage can be transformed into that of the target language to produce the output as given on page 10.

After applying the lattice programme as here outlined

Before applying the thesaurus programme.

Each sentence is written in column form in order more conveniently to indicate its lattice structure. The brackets are self-explanatory, except for the dotted form which indicates the special property of the inventor lattices (H', Y, and Z) that for context purposes they have to be taken as if each argument were separately associated in the next higher lattice: this means simply that such lattices do not generate contexts at all, but provide data for two (or three) possibly different contexts in the next higher lattice.



Note that in the last sentence the lattice-group "to mechanize remove-ment of+the axill-ary sprout-s" is not included in any other lattice until the complete sentence is reached: this cannot easily be shown with the graphical conventions adopted.