A POSSIBLE APPLICATION OF ELECTRONIC COMPUTERS TO THE BLOCK ANALYSIS OF GREEK SENTENCES

BY

ARISTOTELIS D. STATHACOPOULOS

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering, University of Washington

A POSSIBLE APPLICATION OF ELECTRONIC COMPUTERS

TO THE BLOCK ANALYSIS OF GREEK SENTENCES

BY

ARISTOTELIS D. STATHACOPOULOS

I. INTRODUCTION

This thesis describes the research conducted towards the development of a means of isolating and identifying certain word structures, called word blocks, by means of a general purpose digital computer. The method of isolating and identifying these word blocks is called block analysis. It was devised in such a way that specialized equipment for its accomplishment could be constructed using a minimum of logical computer elements. The block analysis described in this thesis is intended for use as an integral part of future sophisticated mechanical translators.

A study of the literature on Machine Translation shows that an automatic word-for-word translation is generally considered feasible. The output of such a translation process will appear in source-language word order and will contain multiple translations for source-language words which, in isolation, require a plurality of target-language alternatives. In other words, in situations where English is the target language, such an output is not in conventional English and requires considerable editing before it can compare with the output of a competent human translator.

In view of the above conclusions, efforts were concentrated on the investigation of methods to improve the output. Early in these studies it was recognized that improved translation could be provided if the translating process was based on larger units than single words. Such units have to be linguistic constructs, i.e., constructs of linguistic forms that can be established on the basis of what is known about the structure of language in present-day linguistic science. A particular type of linguistic construct will be defined and utilized in this thesis and will be termed "word block." Some method for the automatic recognition of these word blocks is required before actual sophisticated translation procedures are begun.

Initial work was done on the automatic analysis of English noun blocks to investigate the feasibility of block recognition. The promising results of this early study were published in "The Trend in Engineering at the University of Washington," in April, 1957.[1] Since block analysis proved to be feasible, at least, attention was turned to the elaboration of a method for the block analysis of Greek sentences. Greek was chosen primarily because it is the author's native language and secondarily because, as a relatively highly inflected Indo-European language, it exhibits many of the syntactic problems inherent in the ultimate goal, the Russian language. This study on the Greek language makes no pretense of completeness. It is limited to a fair-sized sample of Greek journalistic prose, and the present method of block analysis has an efficiency of about 90 per cent in this stylistic area.

Chapters II through IV are devoted to the definition of the word block and its relationship to the word rearrangement problem of machine translation. The block analysis method is presented in Chapter V. Chapter VI describes the adaptation of this method to a general-purpose digital computer. The results, limitations and possible applications of the method are discussed in chapters VII through IX.

II. CONCEPT OF THE WORD BLOCK

Since word-for-word translation is inadequate it can be concluded that translation should be based on larger units than words. An easily recognized and almost ideal unit is the sentence. It can be isolated easily by the very fact that it is included within certain punctuation marks. Sophisticated translation, on the other hand, is absolutely dependent in the ability of a machine translation program to determine the syntactic links existing among the various subunits within each sentence. These linkages are semantic and grammatical in nature.

The semantic linkages cannot be exploited at the present because they must be converted to grammatical linkages.

The grammatical linkages can be exploited. They include form classes, inflections of various types and word order. Of these, the easiest to exploit are form classes and word order.

These two elements, form classes and word order, are the criteria used to isolate word groups that function as a unit. These word groups are called word blocks.

The majority of word blocks function as simple units with one essential element; the head word of the block. For instance, the noun block functions as the noun alone. As such, the noun block can be replaced by its head word with no change in the basic structure of a sentence. In some cases the head word is the only constituent of the block.

---

[1] Stathacopoulos, A. D., "Noun Phrase Analysis Using the IBM 650 Computer," The Trend in Engineering at the University of Washington, April, 1957, p. 20

The prepositional block, however, functions as a complex unit with two essential elements. These elements are the preposition and the noun.

III. USE OF WORD BLOCKS IN WORD ORDER REARRANGEMENT

Most bilingual persons recognize that there are significant differences in the phrase, clause, and sentence structure of any two different languages. These differences in some cases are so pronounced that if no provision is made for structural transformations during the translating process, considerable ambiguity will often exist in the translated material. The problem can be best defined by the presentation of a series of examples of French, German, Greek and Russian sentences with their word-for-word translations as well as their translations into conventional English.

French example:

Je lui ai donné le livre bleu.
I him gave the book blue.
I gave him the blue book.

German example:

Wir haben gestern Äpfel gegessen.
We have yesterday apples eaten.
We have eaten apples yesterday.

Greek example:

Τό ἐδῶ ὑπάρχον μνημεῖον εἶναι 30 ἐτῶν.
The here existing monument is 30 of years.
The monument existing here is 30 years (old).

Russian example:

Из приведенного выше примера видно, что такие условия ненормальные.
From cited above example, it is evident that such conditions are abnormal.
From the example cited above it is evident that such conditions are abnormal.

Although the above examples were selected to emphasize the problem, the fact remains that word rearrangement is a significant part of the translating process. If one adds to the word-for-word translation of the above examples the ambiguities of multiple grammatical and nongrammatical meaning, the resulting translation will be nearly unreadable. On the other hand, if the translated sentences appear in the target-language word order, they can be more readily understood in spite of the presence of multiple alternatives.

Study of the problem indicates that, in mechanical translation, word-order rearrangement can be accomplished more easily by first rearranging the words within uniquely defined word blocks, then rearranging the word blocks in the light of the environment of a clause or sentence.

For example, the French sentence above can be divided into word blocks as follows:

I	/	him	/	gave	/	the book blue	/	.
1		2		3		4		

In word block 4 the adjective (blue) should in English be placed in front of the noun (book) and then block 2 should be placed after block 3. The sentence will then appear in the following form:

I	/	gave	/	him	/	the blue book.
I		3		2		4

The German example can be segmented into word blocks as follows:

We	/	have	/	yesterday	/	apples	/	eaten	/	.
4				2		3				
						1				

Here there is a case of a split verbal block which is unusual in English word order and should therefore be eliminated. The two parts of the verbal block should be placed together and the sentence will be:

We	/	have eaten	/	yesterday	/	apples	/	.
4		1		2		3		

By rearranging blocks 2 and 3 the final output is:

We	/	have eaten	/	apples	/	yesterday	/	.
4		1		3		2		

The same process of rearrangement can be applied to the remaining two examples.
From the above examples it is evident that three main steps are required for the word rearrangement:

1. Block analysis of the source-language sentence
2. Word rearrangement within each block
3. Block rearrangement within a clause or a sentence

Steps 2 and 3 require considerable linguistic research and are directly dependent upon step 1. The block analysis of sentences is then the most basic step in the solution of the entire problem of word rearrangement.

IV. WORD BLOCKS ISOLATED IN THE GREEK SENTENCE

Since the structures of various languages differ, a single language must be selected as the pilot source language in the development of a block analysis process. Although some of the structural rules of the pilot language may apply to other languages, a specific block analysis process will be required for each particular source language at least to the extent of its details. The general principles of analysis will be common to a number of languages, but the details may differ.

Greek has been selected as the pilot source language for the development of the block analysis process because the writer is more familiar with it and as a rather highly inflected language it presents several interesting problems common to many European languages including Russian.

The word blocks for the particular pilot language must be defined uniquely by their possible constituents and specifically by their initial constituents.

The word blocks for this process were defined in cooperation with the linguists of the University of Washington Mechanical Translation research group as follows:

1. Noun Block: The noun blocks are introduced by articles, adjectives or nouns and their constituents are articles, adjectives, adverbs, nouns and coordinating conjunctions.
   Noun blocks may appear before or after verbs as subjects or objects, after nouns, and as parenthetical expressions.

2. Prepositional Block: The prepositional blocks are always introduced by prepositions and their constituents are prepositions, articles, adjectives, adverbs, nouns and coordinating conjunctions, e.g., preposition preceding a noun block. Prepositional blocks may occur as parenthetical phrases, after verbs as objects or after nouns.

3. Verbal Block: Verbal blocks consist of any verbal form.

4. Adverbial blocks consist of single free forms which either introduce an adverbial phrase (i.e. ἐπάνω εἰς τὸ πιάνο , up on the piano) or qualify a verb which they then follow (i.e. ἡ συζήτησις προχωρεῖ κανονικῶς, the conversation is continuing normally). Adverbs appearing within noun or prepositional blocks are not considered adverbial blocks but constituents of the noun or prepositional block (i.e. τὸ ἐδῶ ὑπάρχον μνημεῖον , the monument here).

5. Coordinating Conjunction Block: The coordinating conjunction block consists of a single coordinating conjunction and can occur between words, clauses or phrases.

V. DESCRIPTION OF THE BLOCK-ANALYSIS METHOD

1. Prepositional and Noun Block Analysis

The operational principle of the prepositional and noun block analysis is the determination of the word-order classes of the block constituents from which certain logical products (intersection and union sums) can be formed.

In order to define adequately word-order classes the general concept of word-order class will be discussed.

The principal problem of Machine Translation is the solution of the problem of multiple grammatical and non-grammatical meaning. The solution demands first the elaboration of gross categories of words. These entities usually correspond to the parts of speech. The basis for the determination of these word categories is their distributional patterns or their occurrences, i.e., all words occurring in the same syntactic situations are members of the same gross category. The syntactic occurences shared by the members of a category may be regarded as the elements of a set, and the category itself becomes the set.

Form classes and word-order are two grammatical features common to all the word blocks considered in this study, and thus they become the criteria for the determination of the elements of every set. The gross categories or sets whose elements are determined on the basis of both form classes and word-order are called word-order classes for word-order is the variable criterion and as such it becomes the principal means for the isolation of word blocks. The elements of a word-order class represent the occurrence of the word-order class in question plus its possible cooccurrences with other word-order classes.

For example, a preposition (from) can occur in a sentence and it can occur in conjunction with an article (the) in a given word sequence (from the); it can occur in conjunction with an adjective (from big); or it can occur in conjunction with a noun (from observation). All these are accepted word sequences according to the word-order rules of the English and Greek languages. Thus the word-order class of a preposition, class PR,

has four members a, b, c and d; or a -, its own occurrence; and the possibilities of its occurrence in conjunction with:

b - an article
c - an adjective
d - a noun

The word-order class of the article, class AR, has three members b, c, and d. That is, b -, its own occurrence; and the possibilities of its occurrence in conjunction with:

c - an adjective
d - a noun

Additional possibilities of the occurrence of a preposition or an article in conjunction with words which belong to different grammatical classes will be discussed later.

The logical products (intersection and union sums) are defined as follows:

Intersection Sum: The intersection sum ($X \cap Y$) of the classes X and Y is a new class which is the set of all common elements in both X and Y. For example, if the elements of a class X are a, b, c and d, and the elements of a class Y are a and b, then the intersection sum ($X \cap Y$) is a new class with elements a and b.

Union Sum: The union sum ($X \cup Y$) of the classes X and Y is a new class which is the set of the elements in either X, or Y, or both. For example, if the elements of a class X are a, b, c and d and the elements of a class Y are a and b, then the union sum ($X \cup Y$) is a new class with elements a, b, c and d.

In consideration of the word-order classes of the preposition and the article as they have been defined previously, a new word-order class can be formed by the intersection sum of the classes PR and AR, or the word-order class $PR \cap AR$. The elements of this new class are uniquely determined by the definition of the intersection sum and are the common elements of both classes, PR and AR, b, c and d. This means, therefore, that the sequence, PR - AR, is a permissible one within the prepositional block structure and that it can occur in conjunction with an adjective or a noun. This can be observed from the following examples:

. . . . . of ↓ Prep. the ↓ Art. big ↓ Adj. . . . .

. . . . . of ↓ Prep. the ↓ Art. University ↓ Noun . . . .

Consideration of this simple case demonstrates that the new word-order classes formed by the intersection sum have at least a nonincreasing membership. This leads to consideration of the possibility that after a sufficient number of intersection sums a word-order class can be formed which will have only a single member. The formation of such a word-order class could facilitate the solution of our main problem, the recognition of the word sequence of a noun or prepositional block.

To demonstrate the process of forming a single-member word-order class let us consider a prepositional block:

He appeared / at the new large entrance/ . .
Prepositional Block

A schematic description of the word-order classes of the constituents of the prepositional block under consideration is shown in Fig. 1. From the schematic description it can be seen that the word-order class of the preposition (at) has a membership consisting of the elements a, b, c and d. The word-order class of the article (the) consists of the elements b, c, and d.

We designate the preposition class as PR, the article class as AR, the adjective class of the word (new) as ADJ_1, the adjective class of the word (large) as ADJ_2 and the noun class as N.

Now the class $PR \cap AR$ will have the members b, c and d. After the formation of the class $(PR \cap AR) \cap ADJ_1$, the members will be c and d. The class $(PR \cap AR \cap ADJ_1) \cap ADJ_2$ will have as members the elements c and d. Finally, the word-order class $(PR \cap AR \cap ADJ_1 \cap ADJ_2) \cap N$ will have the single member d. Thus, upon reaching a single-member class, the processing of the constituents of the prepositional block has been completed. Under certain unique conditions, however, it is possible to obtain a single-member class without having completely analyzed the prepositional or noun block. These conditions will be discussed later.

The example of a prepositional block just discussed is used only to demonstrate the basic function of the block analysis. It does not represent the general form of a Greek prepositional block.

To establish the more general form of a Greek prepositional block, possible variations of the example already discussed should be examined.

The fact that a noun has been encountered during the process of analysis does not uniquely indicate that this noun is the last constituent of the prepositional or noun block under consideration. It is possible in Greek word-order structure that a noun may be followed by a word which can be a constituent of the prepositional or noun block being processed. The word-order classes that follow a noun and belong to its prepositional or noun block are:

1. A coordinating conjunction which syntactically links the preceding noun with another noun.
2. An article, adjective or noun in the genitive case actually governed by the preceding noun.

Note that Greek articles can be conjugated.

To include the two foregoing cases in the process of analysis, the union sum of logical algebra must be included in the method. The new word-order class, formed by the union sum of the noun with the next word class, contains as its elements all the information about both classes. Examination of this information facilitates the determination of the possibility that the word class which follows the noun is a constituent of the prepositional or noun block.

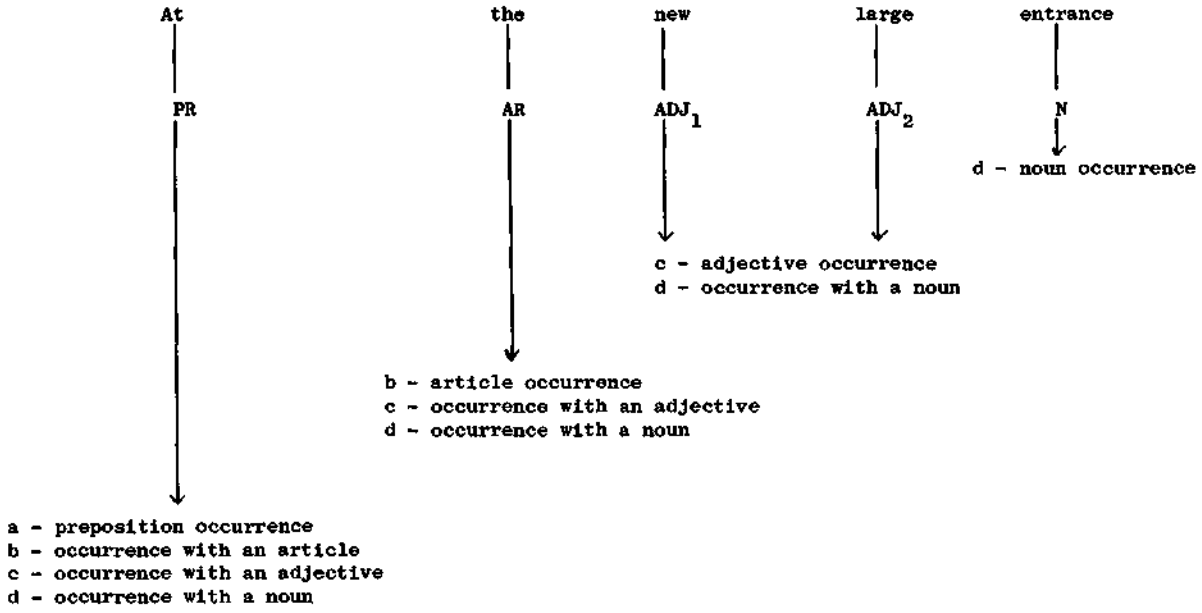


FIG. 1

In order to demonstrate the use of the union sum as a part of the block analysis method let us examine the following Greek prepositional block:

/Εἰς τό Πανεπιστήμιον τῆς Οὐάσιγκτων/ εὑρίσκεται ....
Prepositional block

the translation of which into English is:

/ At the University of Washington / there is . . . .

In the above Greek sentence the notion of possession is not expressed by a preposition as it is frequently in the English translation but by the use of the genitive case of the article.

A schematic description of the word-order classes of the constituents of the prepositional block under consideration appears in Fig. 2.

The word-order class PR∩AR, which has the elements b, c and d, can be formed by the previously discussed process. The class (PR∩AR)∩N has the single member d which might indicate the end of the block. Since linguistically this is not necessarily so, the next word-order class must be examined to determine whether it is a conjunction, a genitive article, a genitive adjective or a genitive noun. To consider these possibilities the union sum of the last single-member word-order class with the next word-order class should be formed, or the class (PR∩AR∩N) ∪ GAR (GAR means article occurring in the genitive case). The membership of this new class will be b, c, d and e, showing that the next word is a genitive article and consequently that the previous noun is not the last constituent of the prepositional block. At this point the information that the article (τῆς) is in the genitive case may be eliminated because it is no longer needed and the members of this last class GAR (GAR means genitive article after the genitive case information has been eliminated) will be b, c and d. Next it

is observed that the word-order class $(PR \cap AR \cap N \cup \overline{GAR}) \cap GN$ (GN means genitive noun) will have the single member d. This means that another noun has been encountered, but still does not indicate uniquely that the noun under consideration is the last constituent of the prepositional block. To determine whether the last noun is the final constituent of the prepositional block, the union sum with the next word-order class, $(PR \cap AR \cap N \cup \overline{GAR} \cap N) \cup V$ must be formed. Since the elements of this new class are d and f, the last word is neither a conjunction nor a word in the genitive case; thus the last genitive noun is the final constituent of the prepositional block under consideration.

In the previous example it is demonstrated that the extended prepositional block can be isolated by the block-analysis process.

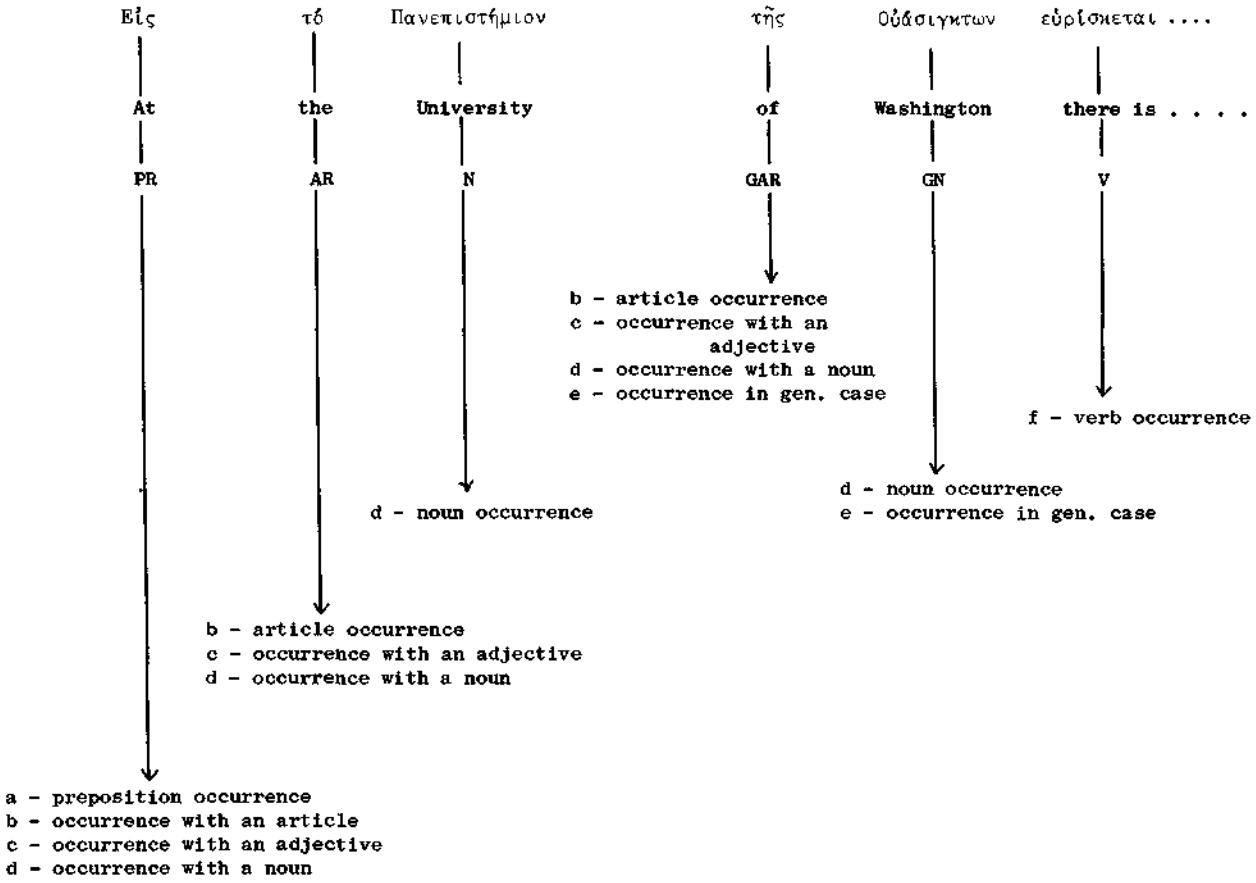


FIG. 2

In order that the process of the prepositional and noun-block analysis be mechanized, it is necessary that each word-order class be represented by some kind of symbol or tag, which must have the following two properties: (1) it must contain all the structural information about each word-order class and (2) it must be applicable to the logical operations required by the method. The method proposed is the use of intersection and union sum of binary numbers.

The following tag-form has been arbitrarily selected for the mechanization of our process:

N	ADV	ADJ	AR	PR	GC	PV	V	P	N
1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0
d		c	b	a	e		f		d

where:

N: Noun GC: Genitive Case

ADV: Adverb	PV: Principal Verb
ADJ: Adjective	V: Verb
AR: Article	P: Participle
PR: Preposition	N: Noun

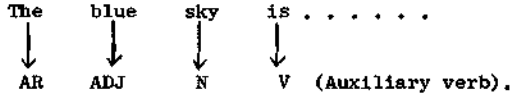
In the tag-form for the mechanization of the block-analysis process, nine digits will suffice to describe all the information required by the method. The tenth digit is redundant information. The digit order of the tag-form is arbitrary.

If a word has a given syntactical property, the tag will have a one in the respective digit position. Otherwise it will have a zero. For example, a noun in the genitive case will have ones in digit positions 1, 6 and 10, with zeros in the remaining digit positions.

All verbs will have a one in the 8th digit position of the tag; however, if the specific verb is used only as a principal verb (not in an auxiliary sense), the tag will have a one also in position 7.

The selected tags have the property that the intersection sum of any two permissible word classes is the word-class tag of the final member, e.g., $A \cap B = B$. This logically indicates B as a subclass of A.

In order to demonstrate the use of binary tags as an integral part of the analysis let us consider the following noun block:



The membership of the word-order class of the article has been discussed previously and consists of the following members:

b - Article occurrence
c - Occurrence with an adjective
d - Occurrence with a noun
g - Occurrence with an adverb

The occurrence of an article in conjunction with an adverb is a special isolated case which will be discussed later.

The binary tag representing the word-order class of the article will have the form:

N	ADV	ADJ	AR	PR	GC	PV	V	P	N
1	1	1	1	0	0	0	0	0	1

The adjective (blue) will be represented by the following tag-form:

1110000001

The word-order class of the noun will be represented by the tag:

1000000001

By definition, the intersection sum of two binary numbers will result in a ONE whenever we add two ONES in the same column, while it will yield ZEROS in any other case. The intersection sum of the first two word-order classes of the block yields:

1111000001 ←——— AR word-order class

1110000001 ←——— ADJ word-order class

1110000001 ←——— New word-order class,
AR∩ADJ = ADJ

Next, the intersection sum of the new class with the following word-order class (sky) is:

1110000001 ←——— AR ADJ

1000000001 ←——— N

1000000001 ←——— (AR∩ADJ)∩N = N

By the presence of ONES only in the first and last digit positions, this tag indicates that a noun has been encountered by the process. Following the identification of a noun, the union sum with the next word-order class is always used to determine whether this noun is the final constituent of the noun block or not. For this purpose the tag of the word-order class of the auxiliary verb must be defined as having the follow-

ing form:

0000000100

The union sum of two binary numbers is defined to yield a ONE whenever two ONES or a ONE and a ZERO are added in the same column, while a ZERO will result only when two ZEROS are added. By taking the union sum with the next word tag, the result is:

1000000001 ⟵ AR ∩ ADJ ∩ N

0000000100 ⟵ V

1000000101 ⟵ (AR ∩ ADJ ∩ N) ∪ V

This new word-order class indicates that the next word (is) is not in the genitive case since there is a "0" in the sixth digit position. It is not a conjunction, either. The conjunction tag has been selected to be:

1010000001

because in the environment of the prepositional or noun block it can appear with a noun or an adjective. Therefore, the new word-order class is not a conjunction.

In the above example, the word following the noun is neither in the genitive case nor a conjunction; the word (sky) is therefore the final constituent of the noun block under consideration. By this process the noun block,

/ The blue sky / . . . .

is isolated and can be uniquely identified since it is introduced by an article.

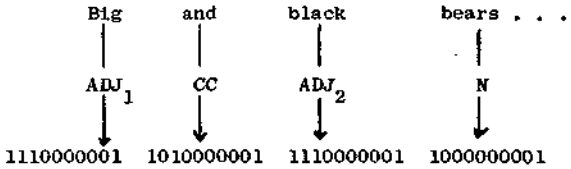
The tags representing the word-order classes which are required by the block-analysis method herein described are listed in Table I.

2. Special Cases of Prepositional and Noun-Block Structures

There are two special word-order variations in the structure of the prepositional and the noun blocks of the Greek language which have not been discussed. A presentation of the analysis of the prepositional and noun block would be incomplete without some discussion of these variations. They occur infrequently, but a complete analysis requires that provision be made that the following two cases be recognized and analyzed correctly.

A. Occurrence of a Coordinating Conjunction Following an Adjective

A coordinating conjunction may appear after an adjective and link it with another adjective which follows the conjunction. In this case, after the conjunction has been recognized, the union sum of the conjunction word-order class and the next word-order class must be formed. The process then continues as previously discussed. To demonstrate this consider the following example:



First the word-order class $ADJ_1 \cap CC$ is formed:

1010000001

The union sum of this class with the next word-order class is then made. The resulting class in its binary representation will be:

1110000001

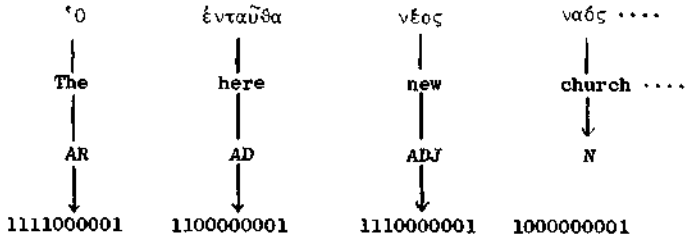
Finally, by the general method of block analysis the last word-order class $(ADJ_1 \cap CC \cup ADJ_2) \cap N$ in its binary form is:

1000000001

indicating that a noun has been encountered.

B. Occurrence of an Adverb Following an Article

In this case, as in the previous case, after the adverb is identified, the union sum with the next word-order class must be formed, at which point the process will continue according to the general method. To demonstrate this consider the following Greek example:



Forming the following logical products in the listed sequence the process of analysis yields the following results:

$$AR \cap AD = 1100000001$$

$$(AR \cap AD) \cup ADJ = 1110000001$$

$$(AR \cap AD \cup ADJ) \cap N = 1000000001$$

By this process the noun block under consideration will be isolated and identified.
In cases in which the adverb follows the adjective, the original process will suffice.

TABLE I

WORD-ORDER-CLASS BINARY TAGS

N	AD	ADJ	AR	PR	GC	PV	V	P	N
1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0	1/0
d	g	c	b	a	e		f		d

Word order class	Binary tag
01. Preposition	1111100001
02. Article in the nom., dative and acc. case	1111000001
03. Article in the genitive case	1111010001
04. Adjective in the nom., dative and acc. case	1110000001
05. Adjective in the genitive case	1110010001
06. Noun in the nom., dative and acc. case	1000000001
07. Noun in the genitive case	1000010001
08. Principal verb	0000001100
09. Auxiliary verbs and certain principal verbs which may be used as auxiliary verbs	0000000100
10. Adverb	1100000001
11. Coordinating conjunction	1010000001
12. Punctuation marks (except the comma)	1000100001
13. Participles of verbs of category (8)	1110001011
14. Participles of verbs of category (9)	1110000011
15. Participles of category (8) in the genitive case	1110011011
16. Participles of verbs of category (9) in the genitive case	1110010011

3. Verbal Block Analysis

Since the verbal block consists of one or more verbal forms, identification of the verbal tag will suffice for the isolation and identification of this block.

In some cases the auxiliary verb (to be) and principal verbs which sometimes may be used in an auxiliary sense are complemented by an adjective or by two adjectives separated by a comma, e.g.:

The sky is blue.

To consider this possibility the verb is further examined to determine whether it can be used in an auxiliary sense; and, if so, the following adjectives, when they occur, are isolated and identified as a special form of a noun block.

4. Adverbial Block Analysis

The adverbial block consists only of a single free form introducing an adverbial phrase or following the verb it qualifies (cf. IV/). Under these conditions the adverbial block can be isolated by an identification of its adverbial binary tag.

5. Coordinating Conjunction Block

Since this block consists only of the coordinating conjunction, which occurs outside of the established word blocks, it can be isolated and identified by an examination of the conjunction tag.

6. Processing of Participles

Participles are hybrid forms with the characteristics of both adjectives and verbs. Their tags, accordingly, contain all adjectival and verbal information. They may occur either within prepositional or noun blocks or outside of prepositional or noun blocks; in either case their identification is simple. Participles following nouns they qualify are in this study treated as verbal blocks (i.e. ὁ ὑπουργὸς ἀφικνούμενος εἰς τὸ ἀεροδρόμειον....the minister arriving at the airport....).

7. Processing of Punctuation Marks

All punctuation marks except the comma are represented by a special tag which is listed in Table I. The comma is not used in this process because it is used very irregularly by Greek writers. When a punctuation mark occurs, it is recognized by the process and treated as a separate word block appearing between sentences or clauses.

VI. BLOCK ANALYSIS PROGRAMMING

1. Program Description

The block-analysis method just described has been programmed for the I.B.M. 650 computer of the University of Washington Research Computer Laboratory. A program flow chart of the process is shown in Fig. 3. The list of instructions of the program appears in Appendix I.

From the flow chart in Fig. 3 it can be seen that the program functions generally as follows: the first word tag of a sentence is introduced from the start block in the initial test row, which consists of blocks 1 through 9, in order to determine the kind of block this word tag introduces. For instance, if the word related to the first word tag is a preposition, it will introduce a prepositional block, if it is an article, an adjective or a noun, it will introduce a noun block; if a verb or a participle, it will introduce a verbal block, etc. After the kind of block the word tag introduces has been established, the program stores this word tag as the first constituent of the identified block, branches off to the particular block analysis section of the flow chart, and determines the remaining constituents of the block under consideration. Finally, after the last constituent of the block has been recognized, the next word tag is introduced into the initial test row (at start) and the process continues on to isolate and identify the next word block.

At this point it should be emphasized that a word tag is introduced into the initial test row only if it is the first word tag of the test which is being analyzed or if it is the word tag which follows directly a block just processed. In other words, the initial test row is used only in order to determine the beginning and the nature of a new block.

If an invalid word tag is introduced into the initial test row, it is recognized as such and is punched out as a separate block with special identification which indicates that an invalid word tag has occurred. The process then continues by introducing the following word tag into the initial test row from START. This is accomplished by block 10 of the flow chart.

Let us specifically consider the particular block-analysis sections of the program.

A. Prepositional and Noun Block Analysis Section

This section consists of blocks 11, 2, 3, 4 and 12-37. The function of this section is to determine the constituents of the prepositional and noun blocks which are being analyzed.

Block 11 is used to form the intersection sum of the preposition with the next word-order class, while

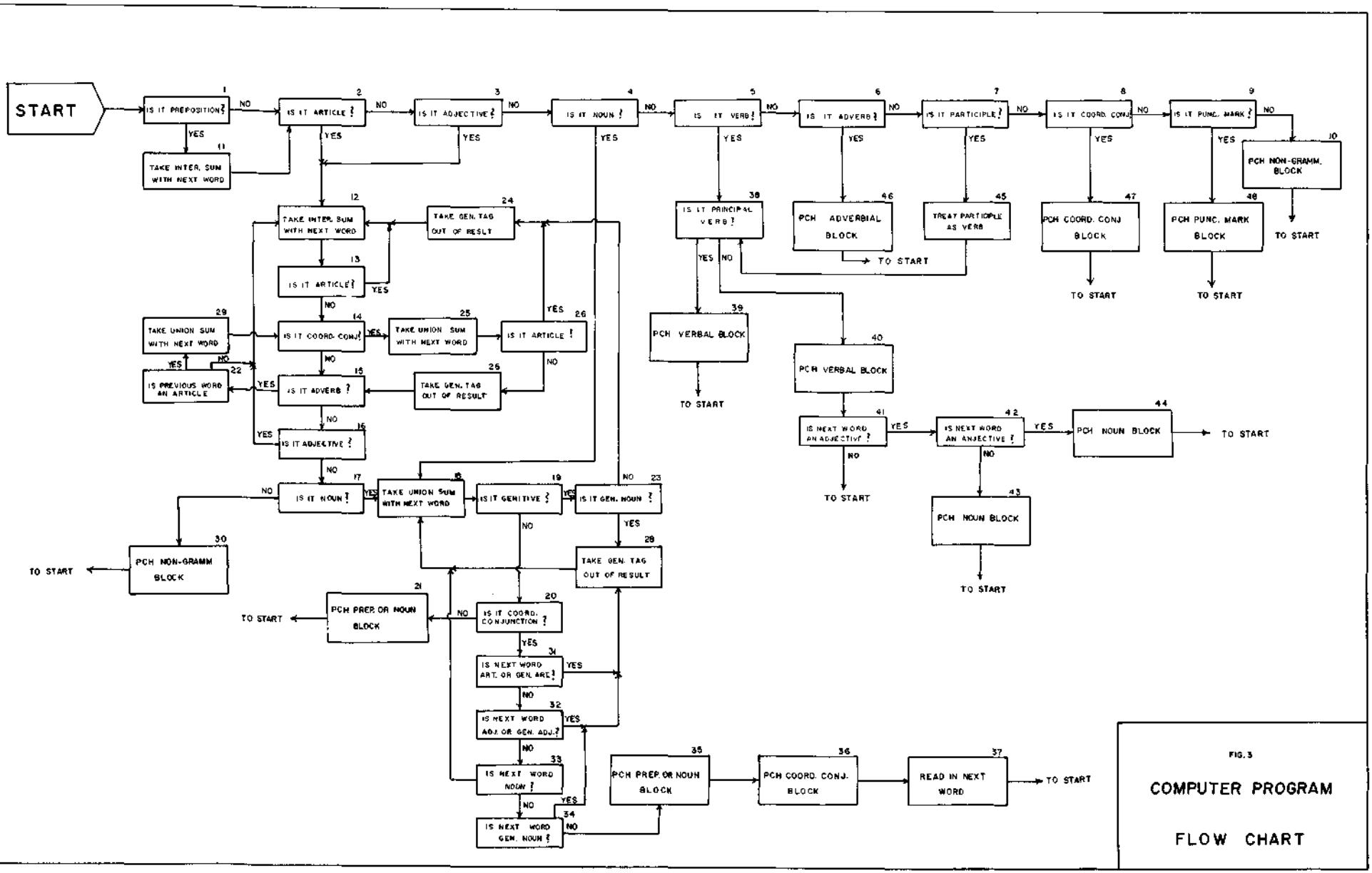


FIG. 3

COMPUTER PROGRAM

FLOW CHART

block 12 is used in the major loop to form the intersection sum of consecutive word-order classes (except the preposition word-order class), as required by the general process.

Blocks 13 through 17 are used to determine the nature of the new word-order class formed in block 12. The function of blocks 22 and 29 is to process the adverb when it occurs after an article as discussed under the special cases of prepositional and noun block structures on page 459. Blocks 25 through 27 process the coordinating conjunction when it occurs after an adjective as previously discussed on page 458. Block 30 is used to punch out an invalid word group if it occurs. The genitive case information is eliminated by any one of blocks 24, 27 or 28 when this information has been used and is of no further value to the process.

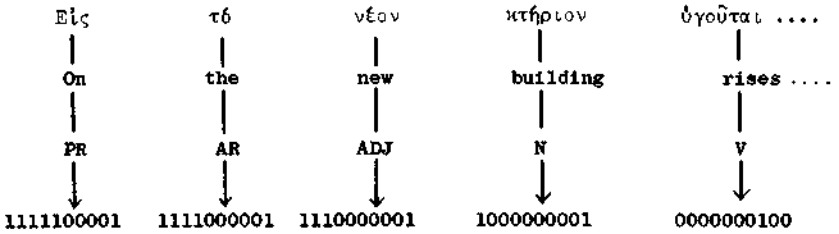
Blocks 18, 19, 20 and 31-39 are used to determine whether a noun word-order class that has occurred is the final constituent of the block. These blocks also determine whether a conjunction occurring after a noun is actually a constituent of the word block under analysis or whether it is a coordinating conjunction and therefore should be treated as a separate word block. If a coordinating conjunction occurs after a noun it is punched out as a separate block from block 36; and then the process through block 37 links back to START by introducing the word that follows the coordinating conjunction in the initial test row to determine the next word block.

The individual tags associated with the completed noun or prepositional block are punched out from blocks 21 and 35.

The function of this section of the program can be best described by the use of the following examples.

Example I

Consider the following Greek prepositional block with its word-for-word translation and the binary tags which correspond to the word-order classes of its constituents.



Initially the PR tag is fed into the initial test row to determine the kind of block this word tag introduces. Since it is a preposition it is recognized as such in block 1. The program establishes that a prepositional block exists and branches on yes to block 11. This word tag is then stored as the first constituent of the prepositional block. The intersection sum:

$$PR \cap AR = 1111000001$$

is then formed in block 11 of the flow chart and the result is identified as an article in block 2. This word tag is then stored as a constituent of the prepositional block under analysis and after the adjective tag is introduced into the loop the intersection sum:

$$(PR \cap AR) \cap ADJ = 1110000001$$

is formed in block 12. The result is subjected to the tests of blocks 13 through 16 to determine its nature. It is identified as an adjective in block 16 and stored as a constituent of the prepositional block; and the process links back to block 12. Here the intersection sum:

$$(PR \cap AR \cap ADJ) \cap N = 1000000001$$

is formed, and it in turn is subjected to the tests of blocks 13 through 17. It is identified as a noun in block 17 and stored. Next the test is made to determine whether this noun is the final constituent of the prepositional block. For this purpose the program branches to block 18 where the union sum:

$$(PR \cap AR \cap ADJ \cap N) \cup V = 1000000101$$

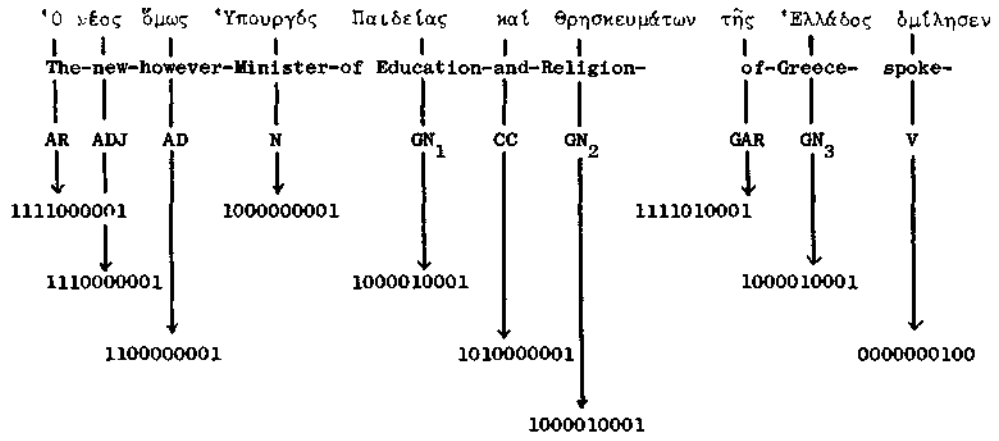
is formed. This result is examined in blocks 19 and 20 to determine whether the word which follows the noun is in the genitive case or whether it is a conjunction. Since this word is a verb, the program branches to block 21 indicating that the prepositional block:

/On the new building / . . . .

has been processed and identified. The tags of the constituents of the analyzed prepositional block are then punched out in block 21, and the process continues by introducing the tag of the word (verb in this case) which follows the prepositional block into the initial test row to determine the next block and so on.

Example II

Consider the general noun block:



Initially the tag of the AR is fed into the initial test row to determine the kind of block it introduces. It is identified in block 2 as an article and thus as the first constituent of the following noun block. The AR is stored and the program branches to block 12 where it forms the intersection sum:

$$AR \cap ADJ = 1110000001$$

which is processed through blocks 13 to 16 and finally identified as an adjective in block 16. It is stored and the program links back to block 12 to form the logical product:

$$(AR \cap ADJ) \cap AD = 1100000001$$

which is processed through the tests of blocks 13 to 15, identified as an adverb in block 15 and stored. The program then continues on to block 22 to examine the tag of the previous word (new) to determine whether it is an article. Since the previous word is an adjective, the program branches to block 12 where it forms the intersection sum:

$$(AR \cap ADJ \cap AD) \cap N = 1000000001$$

which is processed through the test of blocks 13 through 17, and the word is identified as a noun in block 17 and is stored. Next the program continues to block 18 where it forms the union sum:

$$(AR \cap ADJ \cap AD \cap N) \cup GN = 1000010001$$

which in block 19 is recognized as referring to a form occurring in the genitive case. It is stored and is identified as a genitive noun in block 23. The program extracts then the genitive case information from the resultant tag in block 28 and proceeds to block 18. Here the union sum,

$$(AR \cap ADJ \cap AD \cap N \cup \underline{GN}) \cup CC = 1010000001$$

is formed. This new tag is subjected to the tests of blocks 19 and 20, identified as a conjunction and stored in temporary storage. The program then proceeds through blocks 31 to 35 and establishes that the next word is a genitive noun and in agreement with the previous noun. Thus it is established that the coordinating conjunction as well as the following genitive noun are constituents of the noun block. These word tags are stored as constituents of the noun block. The program then proceeds through block 28 for the elimination of the genitive case information to block 18 where it forms the logical product,

$$(AR \cap ADJ \cap AD \cap N \cup \underline{GN} \cup CC \cup \underline{GN}) \cup GAR = 1000010001$$

This is identified to be a constituent of the noun block and stored. The program then branches from block 24 to block 12 where it forms the intersection sum:

$$(AR \cap ADJ \cap AD \cap N \cup \underline{GN} \cup CC \cup \underline{GN} \cup \underline{GAR}) \cap GN = 1000000001$$

which is identified as a noun in block 17 and stored. Next in block 18 the union sum:

$$(AR \cap ADJ \cap AD \cap N \cup \underline{GN} \cup CC \cup \underline{GN} \cup \underline{GAR} \cap GN) \cup V = 1000000101$$

is formed. Processing this new tag through blocks 19 and 20 establishes that the result is neither in the genitive case nor a conjunction. This means that the last noun is the final constituent of the noun block. Finally the program punches out the noun block and proceeds to investigate the next block by introducing the verb tag into the initial test row from START.

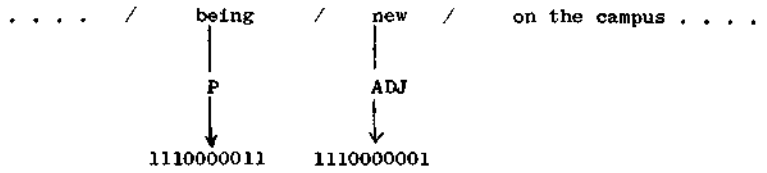
B. Verbal Block Analysis Section

The verbal block analysis section of the program consists of blocks 38 to 45 and has the following functions. It punches out the verbal block which is identified either in block 5 or in block 7. The predicate recognized in block 5 consists of a verb, while the predicate recognized in block 7 consists of a participle not found within a noun block. If the verbal block consists of a principal verb, it is punched out from block 39; while, if it is an auxiliary verb or a principal verb sometimes used in an auxiliary verb sense, it is punched out from block 40.

In block 45 the program establishes that the participle under analysis occurs outside a noun or a prepositional block and should be treated as a verb.

In block 38 the verb tag is examined to determine whether the predicate is a principal verb or an auxiliary form of a verb. If the predicate is a principal verb, the program continues to block 39 where the verbal block is punched out; and the next word tag is introduced from START into the initial test row. If the predicate is an auxiliary verb or a principal verb used in an auxiliary sense, however, the program proceeds to block 41 to determine whether the auxiliary verb is followed by one or two adjectives complementing the verb as discussed in section (V,3). This is accomplished in blocks 41 through 44. If the verb is followed by a single adjective, the adjective is punched out as a special form of the noun block from block 43; and the process then introduces the next word tag into the initial test row from START. If two adjectives occur after the verb, they are punched out as a special form of the noun block from block 44; and the program goes back to START by introducing the next word tag into the initial test row.

To demonstrate the functions of this section let us consider the following example:



After the word block preceding the participle "being" has been analyzed, the participle tag is introduced from START into the initial test row. In block 7 the 9th digit position of this tag is examined to determine whether it is a "1," i.e., whether the word is a participle. Since the word is a participle and does not occur in a noun block, the program branches to block 45, which instructs the process to treat the participle as a verb. The program then proceeds to block 38. Here the 7th digit location is examined to determine whether it is a "1" or a "0." Since there is a "0" in the 7th digit location, the verb is either an auxiliary verb or a principal verb sometimes used in an auxiliary sense; and the program branches to block 40. In block 40 the verbal block is punched out, and the program continues to block 41 to ascertain whether the next word is an adjective. Since the next word (new) is an adjective, the program stores its word tag and continues to block 42 to examine the following word tag. Since the following word (in) is not an adjective, the program branches to block 43, where it punches out the special noun block (new), and proceeds to START by introducing the unprocessed word tag into the initial test row to determine the next word block.

C. Adverbial Block Analysis Section

This section consists of the single block 46. The function of this section is to punch out the adverbial block, which consists of the single adverb recognized in block 6, and then proceed by introducing the next word tag into the initial test row to determine the following word block.

D. Coordinating Conjunction Block Analysis Section

In block 47, which is the only block of this section, the coordinating conjunction identified in block 8 is punched out as a word block; and the process continues by introducing the following word tag from START.

E. Punctuation Marks Processing Section

In this section, consisting of block 48, the punctuation mark tag is punched out as if it were a separate block. The program then continues by introducing the tag of the first word of the next phrase or sentence from START.

2. Program Input

The input of the program is prepared by selecting a portion of source-language text, determining the word-order class of every source word and assigning to it a two-digit code number which represents its word-order class. The two-digit code numbers are then punched on I.B.M. cards (40 two-digit code numbers per card)

in the order their corresponding words appear in the original text. These I.B.M. cards constitute the input of the computer program.

Each two-digit tag represents one ten-digit tag of the form discussed. The two-digit input code, which consists of the respective numbers shown in front of each word-order class in Table I, is used to simplify the preparation of the input cards for testing purposes. The program internally converts, however, this two-digit representation of each tag to its equivalent ten-digit binary tag in order that this tag may be used in the process of analysis which has been described.

3. Program Output

In the output, the ten-digit binary tags of the constituents of each word block appear in a format of one word block per card. To simplify the output it has been assumed that no word block will have more than seven words. The word tags of the constituents of each word block appear in the first seven ten-digit portions of each output card in the order of their contextual occurrence. The eighth ten-digit portion of the output card is used for identification purposes. It contains the sequential number of occurrence of the analyzed word blocks. In the event a word block is developed which is invalid for the process--any one of the low-frequency word sequences which the process does not provide for--it is punched out as a separate word block with a ten-digit special identification (2222222222) in the seventh ten-digit portion of the output card. Also in case an invalid word tag is introduced from the input into START, it is punched out as a special single-word block with the identification (2222222222) in the second ten-digit portion of the output card.

VII. TESTING AND RESULTS

The program was originally tested with selected sentences to cover all the possible variations of block structures which have been discussed. This procedure of testing was used because it is possible that the analysis of sentences selected at random from actual texts may not provide all of the low-frequency special cases for which the process was designed.

After all the loops of the program were tested by the use of the selected examples, actual texts from Greek newspapers and periodicals were used for data to determine the performance of the block analysis method. Several small articles having a total of approximately 1000 words were tested and the performance of the process proved very satisfactory. In order to measure the quality of the analysis performed by the process, however, it is desirable to establish some kind of criterion for process efficiency.

One method of defining the process efficiency is to note the number of the word blocks appearing in the text which is being analyzed in comparison to the number of the word blocks which are correctly analyzed by the process. A percent efficiency can then be defined as follows:

$$\text{Eff. (\%)} = \frac{\text{Number of blocks correctly analyzed}}{\text{Number of blocks in the analyzed text}} (100)$$

From the sentences which have been analyzed it appears that the block-analysis method has an efficiency of the order of 90 per cent. The performance of this process would seem to be satisfactory. It should be noted that no special effort has been made to include in the process any isolated exceptions to the more general structures of word blocks. The inclusion of these cases would increase the efficiency further. This could be done if the process were to be used for production purposes by analyzing large quantities of sentences and then examining the cases where the process had failed. These cases could be classified and studied. Small logical sub-routines could then be included in the general program to provide for the analysis of these special cases.

VIII. LIMITATIONS OF THE PROCESS

The general block analysis method discussed does not have any limitations. It has a satisfactory efficiency which could be improved further without prohibitive difficulty by including most of the possible low-frequency variations in block structure. This is concluded from the fact that in the study of an ever increasing sample, the variation in the block structure of a language is generally of a negative exponential form; and thus a saturation point can be reached by the analysis and study of a sufficiently large sample of sentences.

The program presented herein would be of considerable use in separating the forms of more unique variation from the general group of variations, which constitute 90 per cent of the total.

IX. POTENTIAL APPLICATION OF THE BLOCK ANALYSIS

It has been shown that word-order rearrangement of the output of a translating device can be accomplished by the following three-step procedure.

a. Block analysis of the source-language clause or sentence
b. Word rearrangement within each block
c. Block rearrangement within a clause or sentence

The method of block analysis discussed accomplishes the first step of the word rearrangement process. The next two steps might be programmed after linguistic research establishes the rules governing the equivalent structures between the source language and the target language. The final problem deals with the incorporation

of these procedures into the word-for-word translating process now under study.

A block diagram of a possible translating procedure by the use of logical processing is shown in Fig. 4. The entries in the dictionary of the word-for-word translator may have the following form:

/ Source Semantic Unit / Block Analysis - Binary Tag / Target Equivalent /

With such entries the automatic translating procedure relating to this thesis will be as follows:

Step 1. The input device extracts from the source-language text the "semantic units," counts them in the order of their occurrence and adds an occurrence sequence number which will aid in the final translation.

Step 2. The input device locates each semantic unit in the machine memory and extracts the binary tag of the semantic unit concerned.

Step 3. The binary tag is then sent to the logical processing equipment which processes and rearranges the binary tags into the target-language word order.

Step 4. The target-language equivalent or equivalents are extracted from the machine memory and substituted for the binary tags in their new order.

Step 5. The output device prints out all target equivalents of the clause or sentence concerned in the word order required by the target language.

The logical processing equipment required for the operation of this and other translation processes requires no new design methods or components. Existent basic logical devices can be intercombined to fulfill the requirements of the process described herein.

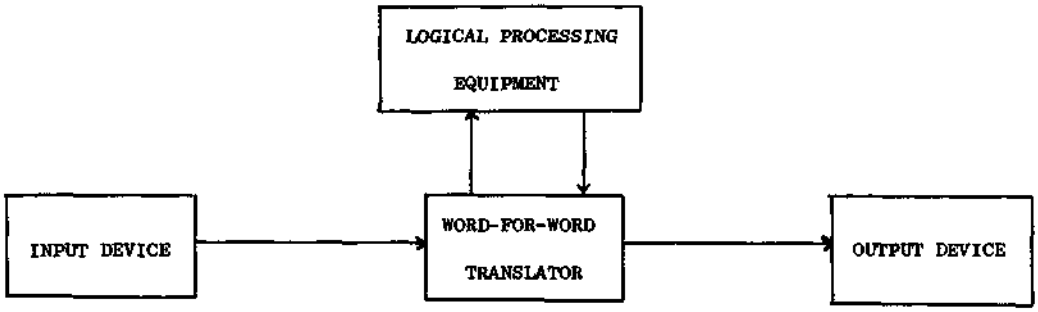


Fig. 4

X. CONCLUSIONS

This paper has been concerned with a method for block analysis of Greek word-order structures. Block analysis of word-order structures constitutes the most basic step in the solution of the problem of word rearrangement of mechanically translated material to fit the word order required by the target language.

Further, "The great advantage of the nominal (prepositional and noun) word-block concept is that the position of the noun or adjective will automatically provide syntactic information which otherwise could be obtained only by asking a great number of questions about such words in isolation."[2]

Also, with regard to the translation process "after Russian words are translated, one word at a time, into English equivalents and codes, further translation is accomplished by either operating upon word blocks (nominal or non-nominal) or entire sentences. Most of the syntactical analysis can be done by operating upon word blocks. . . ."[3]

---

[2] "Design Study for an Integrated USAF Intelligence Date Handling System," Appendix A, Ramo-Wooldridge Corporation, March 31, 1957, p.32

[3] ibid., p.62

To analyze the prepositional and the noun blocks, word-order classes are assigned to each one of the constituents of the respective word block. From these word-order classes, new classes are formed by the specified use of the intersection and union sums of logical algebra. The new word-order classes formed by the use of the intersection sums have at least a non-increasing membership, and thus it is possible to form a single-member word-order class. This single member indicates the occurrence of a word sequence ending with a noun. The formation of such a word-order class under certain unique conditions indicates the end of the analyzed word block. The occurrence of these unique conditions is determined by the use of the union sum and the evaluation thereof. By this process the prepositional and noun blocks can be isolated and identified according to the word-order class that introduces them.

Since the remaining word blocks consist of a single word, their analysis coincides with their identification.

In order that the process be mechanized, binary tags are assigned to represent each word-order class. The process has been programmed and tested on the I.B.M. 650 computer.

The performance of the process is satisfactory with an efficiency of the order of 90 per cent with possibilities of further improvement.

The immediate application of this method is the use of its results for the two remaining steps of the word rearrangement process. These steps are:

a. Word rearrangement within each block
b. Block rearrangement within each sentence or clause

The logical equipment required for the mechnization of this method does not require any new computer design methods or components. It can be designed by the intercombination of existent circuitry and components. The design and construction of such equipment would be most advantageous for further development of this and other processes. Most of all, it would allow the determination of time and component estimates which will lead to the final application of these processes to the general Mechanical Translation process.

BIBLIOGRAPHY

Birkhoff and MacLane, "A Survey of Modern Algebra," Revised edition, The Macmillan Company, New York, 1953.

Parker-Rhodes A. F., notes from a talk given at the Mechanical Translation Conference, M.I.T., October 16-20, 1957.

Ramo-Wooldridge Corporation, "Design Study for an Integrated USAF Intelligence Data Handling System," Appendix A, March 31, 1957.

Stathacopoulos, A. D., "Noun Phrase Analysis Using the IBM 650 Computer," The Trend in Engineering at the University of Washington, April, 1957.

Wall, Robert E. Jr., "The Digital Data-Processing Problem of Machine Translation of Russian to English," chapter 5.

Yngve, Victor H., "The Technical Feasibility of Translating Languages by Machine," Electrical Engineering, November, 1956.

APPENDIX I

COMPUTER PROGRAM

Appendix I includes a complete listing of the computer program used for the mechanization of the block analysis of Greek sentences.

The program, as shown in Table II, is written in standard I.B.M. 650 sequential instruction form. The block numbers refer to the individual blocks of the flow chart (Fig. 3), which the particular group of instructions represents.

TABLE II

PROGRAM INSTRUCTIONS LIST

Remarks	Flow Chart Block Number	Loc.	OP.	D.A.	I.A.
Intersection sum conversion subroutine	11,12	0250	00	0000	0011
		0251	00	0000	0001
		0200	24	0213	0201
		0201	60	0250	0202
		0202	20	0252	0203
		0203	11	0251	0204
		0204	44	0205	0213
		0205	21	0252	0206
		0206	65	0020	0207
		0207	35	0001	0208
		0208	44	0209	0210
		0209	11	0251	0210
		0210	15	8003	0211
		0211	20	0020	0212
		0212	60	0252	0203
Union sum conversion subroutine.	18,15,29	0220	24	0233	0221
		0221	60	0250	0222
		0222	20	0252	0223
		0223	11	0251	0224
		0224	44	0225	0233
		0225	21	0252	0226
		0226	65	0020	0227
		0227	35	0001	0228
		0228	44	0229	0230
		0229	15	0251	0231
		0230	15	8003	0231
		0231	20	0020	0232
		0232	60	0252	0223
Is it preposition?	1	0302	65	0520	0303
		0303	16	0001	0304
		0304	45	0305	0389
Is it article?	2	0305	65	0520	0306
		0306	16	0002	0307
		0307	45	0308	0410
Is it adjective?	3	0308	65	0520	0309
		0309	16	0004	0310
		0310	45	0311	0410
Is it noun?	4	0311	65	0520	0312
		0312	16	0006	0316
		0316	45	0317	0622
Is it verb?	5	0317	65	0520	0318
		0318	35	0007	0319
		0319	60	8002	0320
		0320	30	0009	0321
		0321	44	0359	0322
Is it adverb?	6	0322	65	0520	0323
		0323	16	0010	0324
		0324	45	0325	0404
Is it participle?	7	0325	65	0520	0326
		0326	35	0008	0327
		0327	60	8002	0328
		0328	30	0009	0329
		0329	44	0384	0330

Table II (continued)

Remarks	Flow Chart Block Number	Loc.	OP.	D.A.	I.A.
Is it coor. conjunction?	8	0330	65	0520	0331
		0331	16	0011	0332
		0332	45	0333	0404
Is it punctuation mark?	9	0333	65	0520	0334
		0334	16	0012	0335
		0335	45	0336	0404
Store invalid tag for punching out.	10	0336	69	0520	0337
		0337	24	0527	0338
		0338	69	0017	0340
Punch out subroutine.		0340	24	0528	0402
		0341	65	0018	0342
		0342	15	0019	0343
		0343	20	0534	0344
		0344	20	0018	0345
		0345	71	0527	0346
		0346	65	0021	0347
		0347	16	0019	0358
		0348	45	0349	0355
		0349	69	0022	0350
		0350	24	0527	0351
		0351	65	0023	0352
		0352	15	0350	0353
		0353	20	0350	0354
		0354	65	0024	0347
		0355	65	0350	0356
		0356	16	0027	0357
		0357	20	0350	0401
		0358	20	0024	0348
Is it principal verb?	38	0359	65	0520	0360
		0360	35	0006	0361
		0361	60	8002	0362
		0362	30	0009	0363
		0363	44	0404	0364
Punch out verbal block.	40	0364	69	0520	0365
		0365	24	0527	0366
		0366	69	0367	0400
Is next word an adjective?	41	0367	69	0369	0802
		0369	65	0520	0370
		0370	16	0004	0371
		0371	45	0372	0373
Store adjective tag and read in next word		0372	65	0520	0303
		0373	69	0520	0374
		0374	24	0527	0375
		0375	69	0377	0802
Punch out invalid tag.	10	0403	00	0000	0801
		0400	24	0401	0341
		0402	69	0403	0400
Punch out processed word block.	39,46	0404	69	0520	0405
	47,48	0405	24	0527	0402
Is next word an adjective?	42	0377	65	0520	0378
		0378	16	0004	0379
		0379	45	0380	0382
Punch out noun block	43	0380	69	0381	0400
		0381	65	0520	0303
Punch out noun block	44	0382	69	0520	0383
		0383	24	0528	0402
Is it principal verb (participle)?	38	0384	65	0520	0385
		0385	35	0006	0386
		0386	65	8002	0387
		0387	30	0009	0388
		0388	45	0404	0364
Take inter. sum of PR and next word	1	0389	69	0520	0390
		0390	24	0519	0391
		0391	69	0394	0802
		0394	65	0519	0395
		0395	20	0527	0396

Table II (continued)

Remarks	Flow Chart Block Number	Loc.	OP.	D.A.	I.A.
		0396	15	0520	0397
		0397	20	0020	0398
		0398	69	0399	0200
Examine location 0020 to determine whether it is zero or not.		0410	65	0020	0411
		0411	45	0414	0412
		0412	69	0520	0413
		0413	24	0020	0414
Store word just processed.		0414	69	0409	0049
		0409	69	0520	0415
Take intersection sum with next word.	12	0415	24	0519	0416
		0416	69	0418	0802
		0418	65	0520	0419
		0419	15	0020	0420
		0420	20	0020	0421
		0421	69	0083	0200
Is it article?	13	0083	65	0020	0084
		0084	16	0002	0085
		0085	45	0422	0414
Is it coor. conjunction?	14	0422	65	0020	0423
		0423	16	0011	0424
		0424	45	0425	0484
Is it adverb?	15	0425	65	0020	0426
		0426	16	0010	0427
		0427	45	0428	0434
Is it adjective?	16	0428	65	0020	0429
		0429	16	0004	0430
		0430	45	0431	0450
Is it noun?	17	0431	65	0020	0432
		0432	16	0006	0433
		0433	45	0451	0456
Store processed adverb.		0434	69	0435	0049
Is previous word an article?	22	0435	65	0519	0436
		0436	16	0002	0437
		0437	45	0438	0441
		0438	65	0519	0439
		0439	16	0003	0440
		0440	45	0409	0441
Take union sum of adverb with next word.	29	0441	69	0520	0442
		0442	24	0519	0443
		0443	69	0445	0802
		0445	65	0520	0446
		0446	15	0020	0447
		0447	20	0020	0448
		0448	69	0449	0220
Store and punch invalid word group with special id. tag.		0449	00	0000	0422
		0450	00	0000	0414
		0451	69	0452	0049
		0452	69	0017	0453
		0453	24	0533	0454
		0454	69	0455	0400
		0455	00	0000	0801
Store processed noun tag.		0456	69	0457	0049
Take union sum of noun with next word.	18	0457	69	0520	0458
		0458	24	0519	0459
		0459	69	0461	0802
		0461	65	0520	0462
		0462	15	0020	0463
		0463	20	0020	0464
		0464	69	0465	0220
Is it genitive?	19	0465	65	0020	0466
		0466	35	0003	0467
		0467	60	8002	0468
		0468	30	0009	0469
Is it coor. conjunction?	20	0469	44	0477	0470
		0470	65	0020	0471
		0471	16	0011	0472

Table II (continued)

Remarks	Flow Chart Block Number	Loc.	OP.	D.A.	I.A.
		0472	45	0473	0609
Punch out prep. or noun block.	21	0473	69	0474	0400
		0474	69	0022	0475
		0475	24	0020	0476
		0476	65	0520	0303
Store word just processed.		0477	69	0478	0049
		0478	69	0520	0479
Is it genitive noun?		0479	24	0519	0086
		0086	65	0519	0087
		0087	16	0007	0088
		0088	45	0480	0089
Take genitive tag out of result	28	0089	65	0020	0090
		0090	16	0023	0091
		0091	20	0020	0459
Take genitive tag out of result	24	0480	65	0020	0481
		0481	35	0005	0482
		0482	60	8002	0483
		0483	30	0009	0625
		0625	44	0626	0629
		0626	65	0020	0627
		0627	16	0023	0628
		0628	20	0020	0629
		0629	00	0000	0416
Store word tag just processed.		0484	69	0485	0049
		0485	69	0520	0486
Take union sum with next word.	25	0486	24	0519	0487
		0487	69	0489	0802
		0489	65	0520	0490
		0490	15	0020	0491
		0491	20	0020	0492
		0492	69	0493	0220
Is it article?	26	0493	65	0020	0494
		0494	16	0002	0495
		0495	45	0496	0499
		0496	65	0020	0497
		0497	16	0003	0498
		0498	45	0600	0499
		0499	00	0000	0477
Take genitive tag out of result.	27	0600	65	0020	0601
		0601	35	0005	0602
		0602	60	8002	0603
		0603	30	0009	0604
		0604	44	0605	0608
		0608	65	0020	0606
		0606	16	0023	0607
		0607	20	0020	0608
		0608	00	0000	0425
Store conj. tag for later reference.		0609	69	520	0610
		0610	24	0519	0611
Is next word article or gen. article?	31	0611	69	0612	0802
		0612	65	0520	0613
		0613	16	0002	0614
		0614	45	0615	0044
		0615	65	0520	0616
		0616	16	0003	0617
		0617	45	0618	0044
Is next word adj. or gen. adjective?	32	0618	65	0520	0619
		0619	16	0004	0620
		0620	45	0621	0044
		0621	65	0520	0092
		0092	16	0005	0093
		0093	45	0094	0044
Is next word noun?	33	0094	65	0520	0095
		0095	16	0006	0096
		0096	45	0097	0237
Is next word gen. noun?	34	0097	65	0520	0098

Table II (continued)

Remarks	Flow Chart Block Number	Loc.	OP.	D.A.	I.A.
		0098	16	0007	0099
		0099	45	0036	0028
Read in next word and store gen. noun.		0028	69	0520	0029
		0029	24	0020	0030
		0030	69	0519	0031
		0031	24	0520	0032
		0032	69	0033	0049
		0033	69	0020	0034
		0034	24	0520	0035
		0035	69	0089	0049
Punch out prep. or noun block.	35	0036	69	0037	0400
Store word unprocessed word for later reference.		0037	69	0520	0038
		0038	24	0020	0039
Punch out coor. conjunction block.	36	0039	69	0519	0040
		0040	24	0520	0041
		0041	69	0042	0049
Read in next (unprocessed) word tag.	37	0042	69	0020	0043
		0043	24	0520	0473
Store conjunction and processed word tag.		0044	69	0520	0045
		0045	24	0020	0046
		0046	69	0519	0047
		0047	24	0520	0048
		0048	69	0234	0049
		0234	69	0020	0235
		0235	24	0520	0236
		0236	69	0480	0049
Store noun tag.		0237	69	0520	0238
		0238	24	0020	0239
		0239	69	0519	0240
		0240	24	0520	0241
		0241	69	0242	0049
		0242	69	0020	0243
		0243	24	0520	0244
		0244	69	0459	0049
Store noun processed in block 23 in loc. 0520.		0622	69	0520	0623
		0623	24	0020	0456
Constants.		0001	11	1110	0001
		0002	11	1100	0001
		0003	11	1101	0001
		0004	11	1000	0001
		0005	11	1001	0001
		0006	10	0000	0001
		0007	10	0001	0001
		0008	00	0000	1100
		0009	00	0000	0100
		0010	11	0000	0001
		0011	10	1000	0001
		0012	10	0010	0001
		0013	11	1000	1011
		0014	11	1000	0011
		0015	11	1001	1011
		0016	11	1001	0011
		0017	22	2222	2222
		0018	00	0000	0000
		0019	00	0000	0001
		0020	00	0000	0000
		0021	00	0000	0009
		0022	00	0000	0000
		0023	00	0001	0000
		0024	00	0000	0000
		0025	00	0007	0000
		0026	00	0000	0000
		0027	00	0008	0000
Word tag storing subroutine.		0049	24	0072	0050
		0050	60	0027	0051
		0051	11	0023	0052

Table II (continued)

Remarks	Flow Chart Block Number	Loc.	OP.	D.A.	I.A.
		0052	44	0053	0057
		0053	21	0026	0054
		0054	60	0527	0055
		0055	44	0065	0073
		0073	60	0520	0056
		0056	21	0527	0057
		0057	66	8002	0058
		0058	20	0026	0059
		0059	69	0060	0061
		0060	60	0527	0055
		0061	24	0054	0062
		0062	69	0063	0064
		0063	21	0527	0057
		0064	24	0056	0072
		0065	65	0054	0066
		0066	15	0023	0067
		0067	20	0054	0068
		0068	65	0056	0069
		0069	15	0023	0070
		0070	20	0056	0071
		0071	60	0026	0051
Read in subroutine.		0800	70	0501	0801
		0801	69	0302	0802
		0802	24	0837	0803
		0803	60	0751	0804
		0804	11	0019	0305
		0805	44	0806	0811
		0806	21	0751	0807
		0807	65	0501	0808
		0808	35	0002	0809
		0809	20	0501	0810
		0810	65	8003	0832
		0811	60	0755	0812
		0812	11	0019	0813
		0813	44	0814	0822
		0814	21	0755	0815
		0815	60	0807	0816
		0816	10	0023	0817
		0817	21	0807	0818
		0818	60	0809	0819
		0819	10	0023	0820
		0820	21	0809	0821
		0821	60	0750	0804
		0822	69	0754	0823
		0823	24	0755	0824
		0824	65	0807	0825
		0825	16	0756	0826
		0826	20	0807	0827
		0827	65	0809	0828
		0828	16	0756	0829
		0829	20	0809	0830
		0830	70	0501	0831
		0831	60	0750	0804
		0832	35	0004	0833
		0833	69	0835	0834
		0834	22	0835	0835
		0835	69	0000	0836
		0836	24	0520	0837
		0019	00	0000	0001
		0023	00	0001	0000
		0750	00	0000	0006
		0751	00	0000	0006
		0754	00	0000	0008
		0755	00	0000	0008
		0756	00	0007	0000
Is it article? (after a preposition)	2	0399	65	0020	0075

Table II (continued)

Remarks	Flow Chart Block Number	Loc.	OP.	D.A.	I.A.
		0075	16	0002	0076
		0076	45	0077	0414
Is it adjective? (after a preposition)	3	0077	65	0020	0078
		0078	16	0004	0079
		0079	45	0080	0414
Is it noun? (after a preposition)	4	0080	65	0020	0081
		0081	16	0006	0082
		0082	45	0317	0456
Trailer card.		1952	00	0000	0800

APPENDIX II

COMPUTER INPUT AND OUTPUT SAMPLE

Appendix II includes the original text, coded input and computer output of a sample article analyzed by the use of the block analysis program. The article is one selected from the Greek newspaper, "ΠΕΛΟΠΟΝΝΗΣΟΣ" of February 5, 1957.

The computer input card contains the 40 two-digit numbers as shown in coded forms. The computer output card format is as shown.

Sample Greek Passage

Ἁρμοδίως διεφεύσθησαν αἱ πληροφορίαι ἐφημερίδος τῆς ἀντιπολιτεύσεως, ὅτι εἰς τήν ἐπιτροπήν Δικαιωμάτων τοῦ Ἀνθρώπου ἐν Στρασβούργῳ, ἐνώπιον τῆσ ὁποίας συζητεῖται ἡ Ἑλληνικὴ προσφυνὴ διὰ τὰς Βρεταννικὰς ὠμότητας εἰς τὴν Κύπρον, ἐπῆλθεν δῆθεν συμβιβασμός, μεταξὺ τῆς Ἑλληνικῆς καὶ τῆς Βρεταννικῆς ἀντιπροσωπείας.

Computer Input

10 08 02 06 07 03 07 10 01 02 06 07 03 07 01 06 01 03 07 08 01 04 06 01 02 04 06 01 02 06 08

10 06 01 03 05 11 03 05 07

Computer Output

1100000001 1

0000001100 2

1111000001 1000000001 1000010001 1111010001 1000010001 3

1100000001 4

1111100001 1111000001 1000000001 1000010001 1111010001 1000010001 5

1111100001 1000000001 6

1111100001 1111010001 1000010001 7

0000001100 8

1111100001 1110000001 1000000001 9

1111100001 1111000001 1110000001 1000000001 10

1111100001 1111000001 1000000001 11

0000001100 12

1100000001 13

1000000001 14

1111100001 1111010001 1110010001 1010000001 1111010001 1110010001 1000010001 15