

63-4-1

MARCH 5, 1963

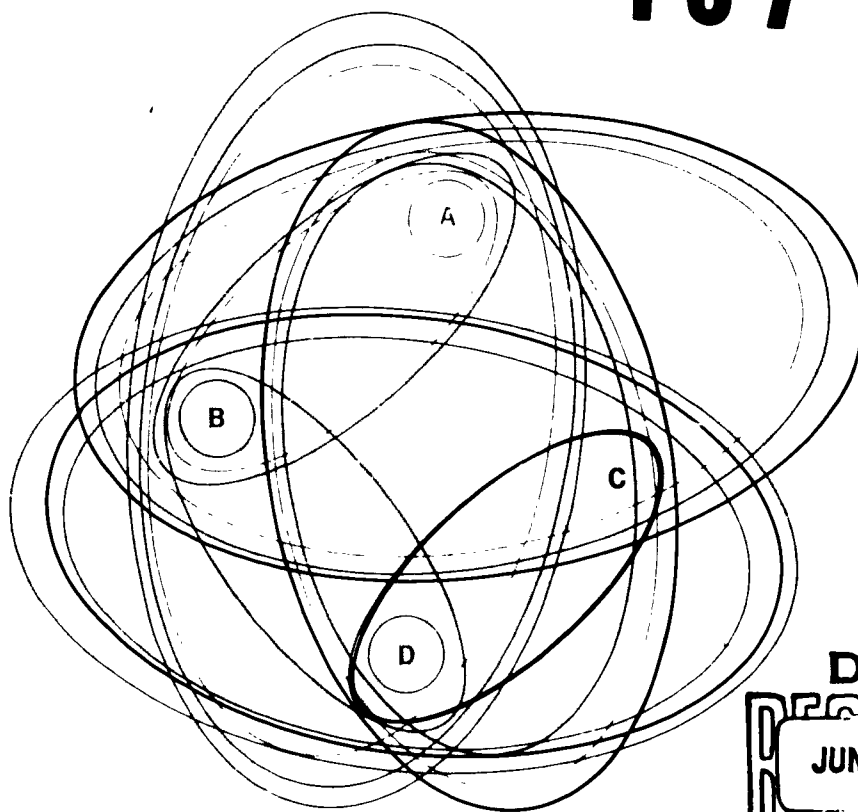
RADC-TDR-63-168

FULCRUM TECHNIQUES TO LANGUAGE ANALYSIS

TRW COMPUTER DIVISION
THOMPSON RAMO WOOLDRIDGE INC.

407 444

CATALOGED BY DDC
AS AD No. _____



DDC
RECEIVED
JUN 5 1963
TISIA B

ЗНАЧИТЕЛЬНЫЙ

(A) SIGNIFICANT (B) CONSIDERABLE (C) LARGE (D) GREAT

PATENT NOTICE: When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Qualified requestors may obtain copies of this report from the ASTIA Document Service Center, Dayton 2, Ohio. ASTIA Services for the Department of Defense contractors are available through the "Field of Interest Register" on a "need-to-know" certified by the cognizant military agency of their project or contract.

RADC-TDR-63-168

5 March 1963

FINAL REPORT
FULCRUM TECHNIQUES TO SEMANTIC ANALYSIS

THOMPSON RAMO WOOLDRIDGE INC.

TRW COMPUTER DIVISION

C169-3U5

AF30(602) -2643

Project No. 4599

Task No. 459901

Prepared for the Information Processing Laboratory,
Rome Air Development Center, Research and Technology Division
Air Force Systems Command, United States Air Force
Griffiss Air Force Base, New York

FOREWORD

The Machine Translation group at Thompson Ramo Wooldridge Inc. has been working under the sponsorship of the Intelligence Laboratory of the Rome Air Development Center, Griffiss Air Force Base, since 1959. This research continues work done under previous contracts with RADC.

During the course of the present research a concurrent contract has been in effect with the National Science Foundation. Some of the tools used in the present research were created under NSF support. In general, studies in the area of Semantics have been done under contract with RADC, while work to improve the techniques for research in MT has been done under contract to NSF.

The support of the Intelligence Laboratory and its members at the Rome Air Development Center is hereby gratefully acknowledged.

This report represents the work of

Loretta Ertel

Herbert L. Holley

Paul L. Garvin

Jules Mersel (Project Manager)

Christine A. Montgomery

George Onischenko

Gerhard Reitz

Steven B. Smith (Associate Project Manager)

ABSTRACT

A new semantic research technique has been developed and employed under this contract. This technique makes possible the construction of semantic classes of words which share the characteristic of specifying a particular translation for a given polysemantic Russian content word.

Improvements have been made in the RW program for the areas of subject recognition and clause boundary determination.

Conclusions were reached about an improved English synthesis. The latter involves reformulating Russian structures into appropriate but non-corresponding English structures.

The major hardware implication of the research suggests that associative memories may provide a simplification in the area of machine translation.

Flow charts and a sample translation are included.

PUBLICATION REVIEW

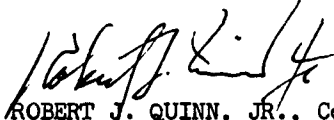
This report has been reviewed and is approved.

Approved:



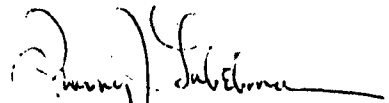
FRANK J. TOMAINI
Chief, Information Processing Laboratory
Directorate of Intelligence & Electronic Warfare

Approved:



ROBERT J. QUINN, JR., Col, USAF
Director of Intelligence & Electronic Warfare

FOR THE COMMANDER:



IRVING J. GABELMAN
Director of Advanced Studies

CONTENTS

Forward	ii
List of Personnel	iii
Abstract	iv
Publication Review Page	v
Introduction	1
Semantic Research	2
Clause Boundary Determination	21
Subject Recognition	23
Transformation Programs for Word Order Arrangement in the English Translation	24
Synthesis	26
Machine Translation of Stylistic Differences	27
Semantic Research Using the Ushakov Dictionary	30
Test Sentences for Syntax Flow Charts	31
Hardware Implications of the Research	45
Dictionary Updating	52
Sample Translation	53
Flow Charts of Determiner-Determinee Method	64
Conclusions and Recommendations	73

INTRODUCTION

The first section of the report describes the semantic research technique which was developed under this contract. This technique is dependent upon our capability for automatic sentence parsing by the Fulcrum approach.

There were two goals in this research; first to define rules for a particular type of multiple meaning problem that had heretofore been either ignored or considered impossible to solve; second, to determine what light such rules might shed on the problem of semantics in general.

Both goals were achieved; the second in particular displayed certain consistencies in the Russian semantic system which are remarkable.

In our research, semantics means the ability to choose the correct translation from among those provided in the dictionary. Our studies have shown that the other text words which determine this selection often fall into groups which themselves seem conceptually related. There has always been a hope among researchers that such semantic classes could be defined. We feel that our technique allows us to uncover the identity of these classes and, even possibly, may provide a metric for the semantic distance function.

SEMANTIC RESEARCH

The first step in our research was to select Russian text and acquire corresponding English translations of that text. The Russian text was selected from the field of biology and was taken from the Doklady Akademii Nauk. All the text was chosen from the same field because we were interested in problems of multiple meaning within a single field, rather than differentiations in translations which were field-defined.

On the opposite page we show p. 564
of Vol. 123 of the Doklady for biology.

С. А. МИЛЕЙКОВСКИЙ

**ЛУННАЯ ПЕРИОДИЧНОСТЬ НЕРЕСТА У ЛИТОРАЛЬНЫХ
И ВЕРХНЕСУБЛИТОРАЛЬНЫХ БЕСПОЗВОНОЧНЫХ
БЕЛОГО МОРЯ И ДРУГИХ МОРЕЙ**

(Представлено академиком Н. Н. Шмидта 8 VIII 1958)

Лунная периодичность размножения и нереста, свойственная многим видам литоральных беспозвоночных тропических и бореальных морей⁽¹⁾, для фауны литорали арктических морей до сих пор в литературе не отмечалась.

Наличие лунной периодичности в размножении литоральных беспозвоночных Белого моря было установлено нами путем изучения динамики численности в планктоне их пелагических личинок; достоверность же результатов обработки количественных проб морского холо- и меропланктона доказана статистически⁽²⁻¹⁰⁾.

Материал собирался в Великой Салме — проливе между п-о. Кийдо и о. Великий (Белое море, Кандалакшский залив). С 26 VI по 14 IX 1957 г. было взято 58 количественных проб планктона. Пробы брались отдельно с горизонтов 16—8 и 8—0 м сетью Джеди из газа № 43 с замыкателем. Изучалась динамика численности всех личинок донных беспозвоночных, а также (для сравнения) ряда постоянных планктеров и их личинок. Все пробы, фиксированные 4% формалином, были просчитаны тотально. Для построения графиков использовалась величина плотности данного вида в 1 м³ в слое 16—0 м, так как численность всех изучаемых видов изменялась в обоих горизонтах синхронно. Кроме того во всех графиках эмпирические неравномерные ряды (пробы брались часто, в июле через день — два, в августе — сентябре через два-четыре дня, но не совсем регулярно, главным образом из-за погоды) переведены линейной интерполяцией в равномерные — трехдневные. Это, а также взятие всех проб в одну и ту же фазу пролива (на полной воде) и на одном и том же месте В. Салмы позволило в значительной степени снять элемент случайности. Так как максимальная глубина В. Салмы около 25 м, а преобладают глубины до 20 м, то личинки донных беспозвоночных, встречавшиеся в ее планктоне, все принадлежали к литоральным и верхнесублиторальным видам.

Как показал анализ, динамике численности большей части личинок донных беспозвоночных из планктона В. Салмы присущ единый закономерный ритм, отражающий лунную периодичность нереста данных видов (рис. 1 а — е). Лунная периодичность нереста свойственна, по нашим данным, следующим беспозвоночным В. Салмы: брюхоногим *Lacuna divaricata* (O. Fabr.), *Littorina littorea* L.; *Diaphana minuta* Brown, *Philine aperta* L., *Limapontia capitata* Müll., *Eubranchius exiguus* (A. a. H.), *Tergipes despectus* Johnston, а также, вероятно, еще ряду видов из отряда голожаберных^(?); она присуща ряду двусторчатых — *Mytilus edulis* L., *Macoma baltica* L., *Mya arenaria* L.; некоторым усоногим — *Balanus balanus* L., *Verruca stroemia* O. F. Müller; ряду мшанок и иглокожих *Ophiopholis aculeata* L., *Ophiura robusta* Ayers, *Asterias rubens* L. (см. рис. 1), а также, очевидно, судя по суммарной динамике численности, и ряду видов *Polychaeta*.

The next problem was that of deciding which words to deal with. In the past, a good deal of semantic research has been devoted to problems of multiple meaning in connection with function words (such as prepositions, particles, conjunctions), while—except for a few instances of idioms—the problem of the multiple meaning of content words (i. e., nouns, verbs, etc.) was ignored. We therefore determined to fix our primary attention on content words.

On the opposite page, we give a sample of the content words investigated along with their multiple translations.

связи

- A. in the process of
- B. connection(s)
- C. on account of
- D. in view of
- E. relation
- F. in response to
- G. accordingly
- H. because of

- I. relating to
- J. owing to
- K. due to
- L. Ø
- M. relationship
- N. related
- O. thus

дальнейшем

- A. then
- B. now
- C. subsequently
- D. subsequent
- E. prolonged

- F. ultimately
- G. further
- H. later
- I. henceforth

разный

- A. various
- B. different
- C. variation in the

- D. differing
- E. varying
- F. differences in the

течение

- A. for
- B. course

- C. over the course
- D. Ø

сторона

- A. side
- B. toward
- C. direction
- D. past

- E. on the other hand
- F. on one hand
- G. laterally

случаях

- A. case
- B. occasions

- C. events
- D. incidents

Next it was necessary to find out which content words had multiple meaning problems within a single field, and occurred with sufficient frequency to allow us to make some generalizations about their behavior. For this purpose a program was written to give us a frequency count of all words occurring within a large body of text, and print out the individual Russian forms in the order of their frequency of occurrence. We thereupon decided to investigate all those content words with multiple equivalents that had one form which occurred 20 or more times, as well as certain other words of particular interest to us.

The following is a sample of the output of the word frequency program. The words are listed in order of frequency and the number to the far left indicates the position the word occupies in the list. The first group of numbers to the right of a word indicates the number of times that word occurred. The next group of numbers indicates what percentage of the total number of occurrences of all words is represented by this word and all words prior to it in the list. For example, this provides us with the information that the 77 most frequent words (out of a total of 37859 running words) represent one third of the occurrences in this text. The next group indicates how much of the total text is represented by occurrences of this word. The next to last group is the position of the word in the list multiplied by its frequency of occurrence. (This number, according to Zipf's law, should approximate 0.1. Our "Zipf" number, however, does not take into account that many numbers, in reality, occupy the same position in the list.) The last group of numbers (here zeros) has as yet no significance.

WORD FREQUENCY SORT FOR CORPORA B

Rank	Word	Number of occurrences	Fraction of text represented by this word	Word number
62	TORPOSENI=	66	.00175	10859
63	KGGDA	65	.00173	10868
64	PAVLOV	64	.00170	10872
65	USLCVNYJ	64	.00170	11042
66	GOLCVNCGO	63	.00167	11038
67	T	63	.00167	11206
68	USLCVNYE	63	.00167	11373
69	MCJNC	62	.00165	11358
70	RAZDRASITELEJ	61	.00162	11338
71	TEIENIE	59	.00157	11126
72	E	58	.00154	11092
73	IZVENENI=	58	.00154	11247
74	REBENKA	58	.00154	11401
75	ESLI	57	.00151	11357
76	PROCESSOV	57	.00151	11509
77	TEM	56	.00149	11457
78	ITI	56	.00149	11607
79	BYLA	54	.00144	11338
80	SIVGTNYX	53	.00141	11271
81	RAZDRASITPLI	53	.00141	11412
82	ITCGG	53	.00141	11553
83	AMINAZINA	52	.00138	11475
84	VSE	52	.00138	11614
85	RAZVITI=	52	.00138	11752
86	9TIX	52	.00138	11891
87	KGRPLENI=	51	.00136	11800
88	KORY	51	.00136	11936
89	ON	51	.00136	12072
90	ISSLEDOVANI=	50	.00133	11970
91	VSEX	49	.00130	11863
92	ENE	49	.00130	11993
93	LET	49	.00130	12124
94	OPYTA	49	.00130	12255
95	POLOSHTEL=NYX	48	.00128	12135
96	BEZUSLOVNCGO	47	.00125	12009

The next step was to list the contexts in which these forms occurred. In order to find them in text, another program was devised to print an alphabetical list of all the words that occurred in text along with their text locations.

The following is a sample of the output of the program to provide a concordance. The first number to the right of a word indicates the number of times that word occurred. Following this for each occurrence of the word five things are provided. The first letter (B here in every case) indicates the corpus in which the word occurred, the second letter, the article; and the third letter the page of that article. The first number indicates the line on which the word occurred and the second the position on that line (first word, second word, etc.). For example, AKTE occurred once in corpus B, article R, page A, line 12, word 2.

For each word we thereupon listed twenty or more occurrences in context, along with the professional English translation of the word and its context.

As an example of our procedure we will use the Russian word ОТНОШЕНИЕ, which translates into English as relation, regard, etc.

The following page shows a work sheet of one of our researchers, listing text locations, Russian contexts and English translations. The translations for the word under consideration are underlined.

⑥
МФА 27 3) ^{гидропротидоренные и негативные результаты} способности цитоблговой ^{показатель} ~~медвежьей~~
have contradictory or negative results as re-
gards ability of thyroid to take up

⑦
МНД 48 7) процесс их гистохимической гидропротид-
цировки в со. синтеза норадреналина отличается
- с процессом. Their histochemical differentiation
as regards noradrenalin synthesis is distinguished

⑧
МНД 5 3) отгадь тиреоидиты какой-либо
из изученных пазух в со. богатства эндогамм
медвежьей = none of investigated sinuses could be
singled out as regards abundance of glands

⑨
МКВ 38 1) ведут себя в этом со. подобно
неизменной эластике = behaved in this
respect like unmodified elastic

⑩
МСВ 21 6) можно получить синтез А и В
в со. проникновения морских инородн = if it
is possible to accept the hypothesis of A + B ... in con-
nection with the penetration of marine elements
ОТНОШЕНИЮ

⑪
ОНА 12 5) число которых по отношению к
огмоэргиям = The number of which in relation
mononuclear cells

Since translators use a proliferation of translations, many of which overlap or are synonymous (see illustration on front cover), it was necessary to determine the minimal number of translations needed to render the word in question into good English.

At the top of the following page we give all the translations of **ОТНОШЕНИЕ** used by the professional translators, and at the bottom the smaller number of translations which we found necessary for the occurrences under consideration. Instances in which the professional translator recast the entire sentence have not been included in this table.

Translations Used by Professional Translators

ratio
relationship
behavior
ration
effects
relation
respect
proportion
point of view
as regards
connection
Ø
response

Necessary Translations

respect
proportion
as regards
relation
Ø ("ly" added to translation of preceding modifier)
ratio
response

Once the preceding two steps had been completed, these contexts were "mapped" in such a way that any consistencies might be quickly brought to the attention of the researcher.

On the opposite page we give the syntactic "mapping" of the word **отношение**. The numbers down the left hand margin indicate the sentence in which the word occurred. The letters down the right hand margin refer to the translation which the professional translator used in each case. The first column on the left gives the Russian words of which this word was an object. The second column lists the prepositions of which this word was an object. The third column gives the modifiers which preceded the word in text. The fourth column lists genitive nominals which followed the word, and the next to last column shows the prepositions which were governed by the word. The final column gives the objects of the prepositions which were governed by **отношение**.

1	имеющие	несмотря на	ближайшее различное	величины для количества	к инерции	D
2	показала					D
3						D
4						F
5						G
6						F
7	имеет					D
8		в	непосредственное	числа	к эластину	F
9						F
10						F
11						A
12		в	этом	количества	к числу	B
13						E
14						C
15						C
16		в	морфологическом	аппарата сосудов способности синтеза богатства	к количеству	C
17						C
18						C
19						A
20		в	этом	проникновения	к клеткам	C
21						A
22						A
23						A

A. respect
 B. proportion
 C. as regards
 D. relation
 E. Ø ("ly" added to translation of preceding modifier)
 F. ratio
 G. response

Next it was necessary to discover which words in the environment of the word under consideration could be considered determiners for the various translations of that word.

A particular set of determiners is generally applicable within a particular syntactic context. We list here the determiners which we found in text for **отомеченне**, along with their translations and relevant syntactic contexts.

(1) в отношении	{ аппарата (apparatus) сосудов (vessels) способности (capacity) синтеза (synthesis) богатства (wealth) проникновения (penetration) }	as regards
	{ 1 г. (1 gram) }	in the proportion of

(2) отношени-	{ величины (magnitude) количества (quantity) числа (number) количества (quantity) }	к	{ первой (first) рациону (ration) числу (number) количеству (quantity) }	ratio	
	{ РНИ }	к	{ индукции (induction) }		relation
	{ лягушек (frogs) }	к	{ температуре (temperature) }		

(3) по отношению к } with respect to

(4) в этом отношении } in this respect

(5) в морфологическом отношении } morphologically

Based on the lists of determiners, we attempted to formulate general rules for the correct translation of the words which we had studied.

The group which has as its translation "ratio" displayed a certain semantic homogeneity. The one-member group 1 г. (1 gram) suggested a group represented conceptually by "amount." Other groups displayed no conceptual homogeneity and were suspected as being open-ended or residue classes.

On the opposite page we show the initial rules which were formulated for the translation of **ОТНОШЕНИЕ** on the basis of our original corpus.

(1) **В ОТНОШЕНИИ** (genitive nominal block) e.g.

A. When genitive nominal is specified amount (e.g., 1 г.)
translate as "in the proportion of"

B. Otherwise, translate as "as regards"

(2) **ОТНОШЕНИЕ** .. genitive nominal block **к** dative nominal block

A. When genitive nominal is a unit of measurement;
"ratio"

B. Otherwise, either "relation" or "ratio"; except when
rule 3A applies

(3) **ОТНОШЕНИЕ**- genitive nominal block

A. When genitive nominal is animate, "response"

B. Otherwise, "relation"

(4) **ПО ОТНОШЕНИЮ К** "with respect to" (idiom)

(5) **В ЭТОМ ОТНОШЕНИИ** "in this respect" (idiom)

It was now necessary to apply the rules which we have developed to fresh text and amend, augment or discard our previous rules.

In the case of **ОТНОШЕНИЕ** for example, an examination of additional text indicated that our rules were in the main correct. More words were added to our list of determiners for the translation "ratio." They were **ДЛИТЕЛЬНОСТИ** (length), **ВЫСОТЫ** (height), and **ДЛИНЫ** (also length). Their semantic relationship to the old set turned out to be as predicted. The expression **В ЭТОМ ОТНОШЕНИИ** (in this respect), which we had originally considered an idiom, proved to be but one of a set, which in the context **В** (modifier) **ОТНОШЕНИИ (-ИЯХ)** determines the translation as "respect". The determining modifiers found were **ДРУГИХ** (other) **ВСЕХ** (all). Here again, we note a certain semantic homogeneity, and we could extend this list on a speculative basis and, with the help of a dictionary, actually find new examples (e.g., **МНОГИХ** [many]).

The research in this area was a success. It has revealed some of the most interesting facets of the Russian semantic system to date. We feel that we have brought MT to the edge of even more important discoveries in this area.

CLAUSE BOUNDARY DETERMINATION

During the contract period a number of routines have been added to the basic syntax program to improve the quality of the translation. The most significant of these is a syntax pass which examines commas and conjunctions whose function is ambiguous (that is, conjunctions which may connect words, phrases, or clauses: for example, "and"), in order to determine whether they function as clause separators. If they are, they are marked as boundaries for subsequent syntactic and semantic searches. Two types of clause boundaries are recognized: (1) those separating independent clauses from each other (can be commas, or conjunctions, or both), and (2) those which separate dependent clauses from the clause on which they depend (commas only):

Ex.: Рибонуклеопротеиды могут иногда находиться только в клетках глубокой зоны, а элементы средней и наружной зоны полностью лишены их. = ribonucleoproteins may sometimes be found only in the cells of the deep zone, while elements of the middle and outer zones are completely devoid of them.

Ex.: Рибонуклеопротеиды, которые отсутствуют от элементов средней и наружной зон, могут находиться в клетках глубокой зоны. = ribonucleoproteins, which are not present in elements of the middle and upper zones, may be found in cells of the deep zone.

The reason for this distinction is that in the first case, syntactic and semantic searches should be discontinued at the clause boundary, whereas in the second, these searches should be interrupted at the left boundary of the dependent clause (which is not searched for syntactic and semantic codes relating to the main

clause), and should be resumed at the right boundary and continued to the end of the sentence. This makes it possible to identify syntactic and semantic elements of a single clause even when they are separated by long stretches of dependent or parenthetical material.

The clause boundary algorithm first searches the sentence for a comma or a conjunction. If the comma or conjunction is between two predicates, the immediate environment is searched to determine whether a definite decision can be made that the item is a clause boundary. (One example of this would be a comma preceded by an unambiguous conjunction, such as *уто*). If no syntactic elements which definitely establish the item as a clause boundary are present, the segment between the two predicates is searched for additional commas or ambiguous conjunctions. If none are present, the current item is marked as the clause boundary. In cases where an additional potential clause boundary is present, the environments of both items are searched to determine whether a definite decision can be made that one or both items are not clause boundaries (for example, commas or ambiguous conjunctions connecting two modifiers). If no definite decision is possible at this point, a further search for an additional comma or conjunction is made, and the same inspection of the environment is carried out on the new item. If no decisions can yet be reached, a fourth potential clause boundary is searched for and checked out. Where no definite conclusions can be made at this point, attempts to resolve the boundary of that particular clause are abandoned and processing of the next clause is begun.

After the initial search of the clause boundary algorithm, if the current comma or conjunction is not between two predicates but between a predicate and the beginning of the sentence, the segment between the comma or conjunction and the predicate is inspected to determine whether it contains a relative pronoun. If a relative pronoun is found, the current item is marked as the preceding clause boundary and a search for the following clause boundary is then initiated.

SUBJECT RECOGNITION

As clause boundaries can now be defined accurately in all but a few cases, it has been possible to make several improvements in the quality of the translation by expanding the existing subject recognition routine to handle objects and by including a routine to identify a series of subjects. The subject-object recognition routine operates within the limits of a particular clause as defined by the clause boundary pass, searching for both subject and object in order of most frequent occurrence. The search for a subject is begun in the segment preceding the predicate and if unsuccessful, continued in the segment extending from the predicate to the next clause boundary or end of sentence; in searching for an object, the order of segments searched is reversed. When a subject or object is encountered, it is marked as to syntactic function and transferred to an extended nominal blocking pass which sets the limits of the particular subject or object block.

If the predicate is plural, the subject search does not conclude when one subject is found, but transfers control to a subroutine which searches for possible additional subjects. The immediate context of the current subject is first investigated to determine whether commas or ambiguous conjunctions are present. If such coordinators are found, a search for an additional subject is made; if no commas or conjunctions are found, the search is abandoned. In the case of a singular subject and a plural predicate a "trouble" signal is stored if no additional subjects are encountered. When additional subjects are found, they are marked as subjects and processed by the subject blocking routine.

TRANSFORMATION PROGRAMS FOR WORD ORDER ARRANGEMENT IN THE ENGLISH TRANSLATION

Since the functional load of word order in inflected languages (such as Russian) is much lower than in non-inflected languages (such as English), it is clear that an adequate machine translation program must provide for the resolution of word order discrepancies between the source and target languages. In our 1961 report (Machine Translation Studies of Semantic Techniques AF 30(602)-2036), we discussed in detail existing programs for three types of word order rearrangement in the English translation. These three types were described as subject-object rearrangement, rearrangement of governing modifier packages, and rearrangement of auxiliaries and modals within predicate blocks. The first type of rearrangement is the most complex, since it involves the identification of more syntactic elements and reshuffling of large blocks of text. Furthermore, in order to operate successfully within multclause sentences, it requires a clean definition of boundaries between clauses. At the time of that report, large-scale rearrangements could be effected only on single-clause sentences, as there was no systematic way of identifying separate clauses. It was possible to rearrange the English translation of:

в среде движется пучок заряженных частиц

(In the medium moves the beam of charged of actives particles)
to read:

The beam of charged particles moves in the medium;
or to rearrange the English translation of:

пучок заряженных частиц объяснил профессор

(The beam of charged particles explained the professor)
to read:

The professor explained the beam of charged particles.

If an additional clause were added, the sentence would read:

пучок заряженных частиц объяснил профессор, а студент
не обращал внимание

The beam of charged particles explained the professor, but
the student wasn't paying attention.

Then the first clause could not be handled by the word order transformation program because there was no program for determining whether the ambiguous conjunction "a" connected words, phrases, or clauses.

Since the decisions of the clause boundary routine discussed above are now one of the information inputs to the rearrangement program, two separate searches for possible rearrangement requirements are made on the sample sentence: one on the first clause (the segment between the sentence beginning and the conjunction); the next on the second clause (from the conjunction to the period). If a requirement for rearrangement is revealed by search of the first segment, the English translation of that segment is rearranged as specified, independently of the second clause, which is not checked for a rearrangement requirement until the rearrangement requirement of the first clause has been satisfied. Thus, the resulting English translation of the sample sentence is:

The professor explained the beam of charged particles,
but the student wasn't paying attention.

SYNTHESIS

Several steps have been taken in this area to achieve something significantly beyond a modified word by word translation by a concentration on the translation algorithm (and translating transformation). With a very few exceptions in the past some Russian constructions (i. e. idioms) have been translated into corresponding English constructions. Because English and Russian display certain syntactic similarities, this policy, though often awkward, is usually comprehensible. In conjunction with the semantic research previously described, exploratory work was undertaken in this area. As a result rules for the handling of Russian constructions containing **можно** and **следует** were devised and applied to new text. The rules proved very successful and serve to render what were previously tricky and awkward constructions into idiomatic English. We felt this was ample justification for programming these rules. This activity is so recent that it is not reflected in the sample output in the appendix of this report.

MACHINE TRANSLATION OF STYLISTIC DIFFERENCES

In connection with our preparations for the machine translation of fiction we have given considerable thought to the problem of the automatic detection and correct rendition of stylistic differences. At the present stage of our research we are able to specify two areas in which we shall ultimately be able to do this. One is the stylistically correct use of vocabulary which is part of the over-all problem of multiple meaning to which the research under the present contract is addressive. This aspect of the stylistic problem is not essentially different from other problems of multiple meaning resolution. The second aspect of style is related to the stylistic function of syntactic elements. We have given consideration to the well-known fact that the position of subjects and objects with respect to the predicate have a definite function in the Russian sentence, in that there is a tendency for the initial position in the sentence, whether it is filled by subject or object, to indicate old information, while there is a tendency for the final position, whether it is filled by subject or object, to indicate new information. This stylistic function of position within the sentence can be preserved in the English translation and can contribute to producing machine translation which is not only intelligible but also stylistically more closely comparable to the original. At this stage we are exploring two possibilities for the rendition of this stylistic value of position into English.

One possibility is the use of passive translations for Russian active sentences whenever in the Russian original the subject is in the sentence-final position, indicating that it constitutes new information. In this case, the Russian subject would be rendered by the English agent object which likewise would be in the final position, and hence equally indicative of new information. Our assumption here would be that the passive English sentence is the stylistic equivalent of the Russian active sentence with inverted order.

As an example, let us take the following Russian sentence: "Какой огромный путь в своем развитии прошла наша страна за сорок два года после свержения власти капиталистов и помещиков и установления советской рабоче-крестьянской власти!" It stems from the lead article "Воспримый научный подвиг" (Unparalleled Scientific Advance) in Pravda of 27 October 1959, dealing with the Soviet moon shot and the pictures taken of the far side of the moon. In the light of our interpretation of style, we may assume that the object of this sentence, "Какой огромный путь..." (What enormous path...) is old information, in the light of the preceding discussion of the great achievements reported on, and as new information is presented the subject "наша страна" (our country) and particularly the final predicative complement "за сорок два года..." (after forty-two years...). Clearly, then, the most idiomatic translation, taking the stylistic function of the sentence portions in the Russian original into account and rendering them equivalently in the English would involve a change from the active to the passive voice, and a retention of the order in which the semantic components of the sentence appear in Russian. Thus, instead of the rearrangement which our present program aims at, namely, "Our country in forty-two years after the overthrow of the rule of the capitalists and landowners and the establishment of the workers' and peasants' rule passed what enormous path in its development," we shall aim at the following translation which preserves the original order by changing the voice of the English verb from active to passive: "What enormous path of development was passed by our country in forty-two years after the overthrow of the rule of capitalists and landowners and the establishment of the workers' and peasants' rule!"

The second possibility is to consider the function of the English indefinite article or absence of the article in the plural as an indication of new information. In this case we would give consideration to the possibility of using the indefinite article or no

article when the corresponding Russian sentence element occupies the final position. An example would be the sentence "Каждое из этих достижений--беспримерный научный подвиг!" taken from the same document. If we translate the final element of this sentence, the predicative complement "беспримерный научный подвиг," using the indefinite article in English, we shall come closest to an idiomatic translation: "Each of these achievements is an unparalleled scientific advance."

This research is as yet in the exploratory stage.

SEMANTIC RESEARCH USING THE USHAKOV DICTIONARY

The utilization of examples of usage given in the Ushakov Dictionary, which was proposed for the present research, has been attempted. Unfortunately, in the tests which we conducted, using about 100 entries, the majority of the examples given in the Ushakov do not allow multiple meaning due to different subject matter field. These differences are not amenable to the determiner-determinee treatment which we are at present in a position to implement.

TEST SENTENCES FOR SYNTAX FLOW CHARTS

In order to facilitate manual and machine checkouts of the syntax routines, a large number of test sentences (189) were created to serve as tools in this process.

The aim of these checkouts is different from that of the processing of live text for testing purposes. The processing of live text tests whether or not our syntax rules cover all significant conditions of Russian sentence structure. The purpose of our test sentences, on the other hand, is to test, not the adequacy of our rules, but whether or not the program actually carries out these rules as intended.

A test sentence was created to checkout each significant branch of every flow chart. This means that for every question asked on a flow chart, at least two sentences were created, one corresponding to the yes answer, one to the no answer. We could thus assume that whenever a failure occurred with a particular test sentence, there was a great likelihood that this failure would be due to the branch on the flow chart which this sentence was intended to test.

Since the test sentences were designed for purposes of the syntax, they were as much as possible deliberately restricted to the syntactic conditions which they are intended to test. In particular we wanted to avoid complicating the tests by introducing more than the minimum necessary number of dictionary entries. Consequently, to display the syntactic difference to maximum advantage, the same dictionary words were used over and over again in the test sentences, bringing about a somewhat unnatural impression. Since some of the test sentences were introduced to test flow chart exits marked "notice of error", these particular sentences were designed to be grammatically incorrect. An example is sentence No. 2 for testing the relative pass.

We wish to illustrate the creation of the test sentences by giving some examples of sentences pertaining to Homograph Resolution Pass HR 1. (Homographs of the type "трудно, согласно," called predicative-adverb-preposition homographs.) Test Sentence No. 1 illustrates the yes answer to the question "Is the predicative-adverb-preposition homograph immediately preceded by a preposition and immediately followed by a modifier?" Note that this sentence contains a homograph (прямо) in the required position between a preposition (от) and a modifier (рекуррентных). This test sentence is correctly translated if the homograph is rendered by an English adverb.

Test Sentence No. 14 on the other hand illustrates a yes answer to the question "Is predicative-adverb-preposition homograph immediately followed by a governed block?" and a no answer to the question "Is predicative-adverb-preposition homograph immediately preceded by a comma?" This test sentence is correctly translated if the homograph is rendered by an English predicate.

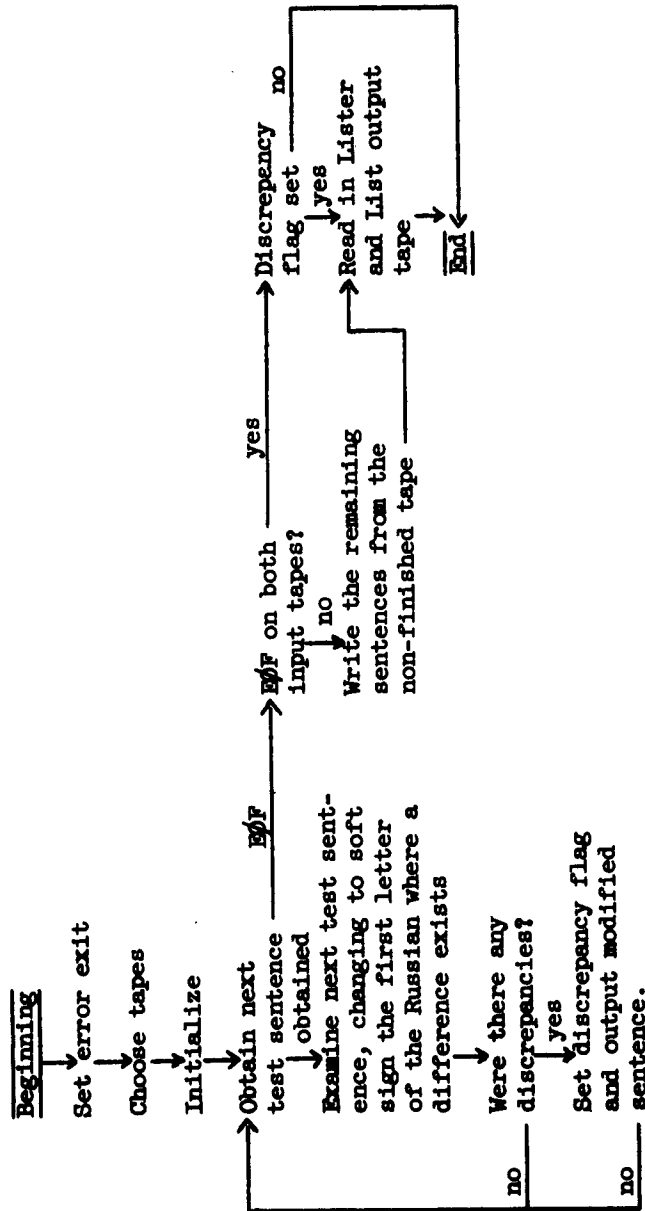
Test Sentence No. 15 differs from Test Sentence No. 14 by having a yes answer to the question "Is predicative-adverb-preposition homograph immediately preceded by a comma?" which is in turn followed by a yes answer to the question "Is governed block immediately followed by a comma?" This sentence is correctly translated if the homograph is rendered by a preposition.

Each time a change is made in the program, the test sentences are run in order to ascertain the effects of that change. The resulting vertical listing is then available for analysis. This is, however, not enough. It does not take into account the possibility that a change which has produced an improvement in one place in the program may inadvertently result in a change for the worse elsewhere.

Consequently, each time a change is made in the program, the test sentences are run, and in addition to printing them out

directly, the information tape of the current run is automatically compared to the information tape of the latest previous run of the same sentences. This comparison is conducted bit by bit. If any discrepancies are detected, a vertical listing of the full sentence is selected for output and each word which displayed a discrepancy is flagged for the benefit of the analyst.

The inventory of test sentences can be expanded when necessary. The flow chart of the test sentence comparer on the following page indicates the provisions that have been made for them.



Test Sentence Comparer

Test Sentences for Symbol Pass

1. α -частицы проникают через ядро.
2. Важные Σ проникают через ядро.
3. Важные частицы проникают через Σ .
4. Нельзя знать, потому что Σ проникает через ядро.
5. Частица Σ проникает через ядро.
6. Постоянная Σ проникает через ядро.
7. Σ проникает через ядро.
8. Падение нажима и Σ проникают через ядро.
9. Создающая Σ частица проникает через ядро.
10. Созданное Σ проникает через ядро.
11. Частица производит Σ .
12. Производя Σ , частица проникает через ядро.
13. Частица может производить Σ .
14. Частицы Σ и $\ominus$ проникают через ядро.

Test Sentences for Homograph Resolution Pass: type "трудно," "согласно"

1. Наши результаты происходят от прямо рекуррентных методов.
2. Явление произошло согласно нашим результатам.
3. Явление произошло между прочим согласно нашим результатам.
4. Производя результаты согласно нашему предположению, опыты произошли без трудностей.
5. Явление произошло ясно.
6. Явление произошло между прочим ясно.
7. Производя результаты согласно, опыты произошли без трудностей.
8. Явление было согласно с нашим предположением.
9. Явление согласно было с нашим предположением.
10. Явление будет согласно с нашим предположением.
11. Трудно предполагать такие результаты.
12. Трудно нам предполагать такие результаты.
13. Предполагать нам трудно такие результаты.
14. Такое явление согласно нашим результатам.

Test Sentences for Homograph Resolution Pass: type "трудно," "согласно"

15. Такие опыты, согласно нашему предположению, произошли без трудностей.
16. Работа, происходящая довольно трудно, совершается сегодня.
17. Важные ясно неизвестные результаты происходят без трудностей.
18. Ясно, что нельзя.
19. Ясно, для такой цели нельзя.
20. Такие опыты, ясно, происходят без трудностей.
21. Такое явление ясно, а наш опыт произойдет без трудностей.
22. Такое явление безразлично трудно.
23. Ясно известный результат происходит без трудностей.
24. Совершенствование таких результатов очень трудно.

Test Sentences for Homograph Resolution Pass: его, ее, их

1. Работа произошла после его очень важных усилий.
2. Предположения их очень важны.
3. Сотрудник сохраняет ее.

Test Sentences for Homograph Resolution Pass: type "постоянная"

1. Постоянная трудная работа иллюстрирует жизнь.
2. Жизнь иллюстрирует постоянных пять работ.
3. Следующее слово принадлежит профессору.
4. Сотрудник сказал следующее.
5. Следующее такую работу понятие принимается.
6. Постоянная принимает разные значения.
7. Наши доклады являются учеными.
8. Наши доклады написаны учеными.
9. Наша работа проведена ученым.
10. Наша работа приводит к ученым.

Test Sentences for Homograph Resolution Pass: type "постоянная"

11. Наша работа приводит к постоянной.
12. Наша работа исходит от постоянной.
13. Наша работа является постоянной.

Test Sentences for Nominal- and Prepositional-Blocking Pass

1. Вопрос открытого окна решается сегодня.
2. Вопрос открытой решетки решается сегодня.
3. Два открытых окна находятся здесь.
4. Две открытых решетки находятся здесь.
5. Открытое и закрытое окна находятся здесь.
6. Нельзя работать без открытого окна.
7. Нельзя работать без открытых окна и решетки.

Test Sentences for Inserted-Structure Pass

1. Конечно, частицы проникают через ядро.
2. Сегодня, частица прошла через ядро.
3. Частицы, конечно, проникают через ядро.
4. Частица, сегодня, прошла через ядро.
5. Вообще говоря, частицы проникают через ядро.
6. Частицы, вообще говоря, проникают через ядро.
7. Частицы, как все знают, проникают через ядро.
8. Важные, таким способом, частицы проникают через ядро.
9. Важные, по нашему предположению, частицы проникают через ядро.
10. Частицы проникают через, таким способом, ядро.
11. Частицы проникают через, по нашему предположению, ядро.
12. Сегодня, частицы прошли, конечно, через ядро.
13. Конечно, частицы прошли, сегодня, через ядро.
14. Конечно, частицы проникают, вообще говоря, через ядро.
15. Частицы, конечно, проникают через, по нашему предположению, ядро.
16. Через, по нашему предположению, ядро мюоны, конечно, проникают.

Test Sentences for Governing-Modifier Pass

1. Вращающиеся каскады идут вперед.
2. Каскады, вращающиеся вокруг ядра, идут вперед.
3. Каскады нуклеонов, вращающиеся вокруг ядра, идут вперед.
4. Вращающиеся вокруг ядра каскады идут вперед.
5. Вращающиеся вокруг ядра каскады нуклеонов идут вперед.
6. Вращающиеся вокруг ядра быстрые каскады идут вперед.
7. Вращающиеся вокруг ядра быстрые каскады нуклеонов идут вперед.

"который" pass—test sentences

1. Каскады, которые идут вперед, очень велики.
2. Каскады которые идут вперед очень велики.
3. Каскады, которые вперед, очень велики.

Test Sentences for Main Syntax Pass—Key To Number Code

X. 0. 0. 0.	Predicate
0. X. 0. 0.	Subject
0. 0. X. 0.	Object
0. 0. 0. X.	Eliminated candidates for subject or object
1. 0. 0. 0.	Predicative plural non-past
2. 0. 0. 0.	Predicative followed by бы
3. 0. 0. 0.	Predicative plural past
4. 0. 0. 0.	Predicative singular non-past
5. 0. 0. 0.	Predicative singular past feminine
6. 0. 0. 0.	Predicative singular past masculine
7. 0. 0. 0.	Predicative singular past neuter
8. 0. 0. 0.	Predicative plural followed by быть
9. 0. 0. 0.	Predicative plural followed by быть followed by second predicative.

Test Sentences for Main Syntax Pass—Key to Number Code

10. 0. 0. 0. Predicative is **был** and is followed by second predicative
 11. 0. 0. 0. Predicative is **будет** and is followed by second predicative
 12. 0. 0. 0. Predicative is a comparative.
 13. 0. 0. 0. There is no predicative, but **есть** followed by infinitive.
 14. 0. 0. 0. There is no predicative, but **есть** **он** followed by infinitive.
 15. 0. 0. 0. Predicative is followed by infinitive
 16. 0. 0. 0. No predicative, but gerund
 17. 0. 0. 0. No predicative, but gerund followed by infinitive

-
0. 1. 0. 0. Unambiguous nominative plural
 0. 2. 0. 0. Genitive singular/nominative-accusative plural ambiguity
 reduced by unambiguous modifier
 0. 3. 0. 0. Genitive singular/nominative-accusative plural ambiguity
 reduced by nominative numeral
 0. 4. 0. 0. Potential nominative singular, any gender
 0. 5. 0. 0. Potential nominative singular, feminine
 0. 6. 0. 0. Potential nominative singular, masculine
 0. 7. 0. 0. Potential nominative singular, neuter
 0. 8. 0. 0. No subject, but predicative can be impersonal
 0. 9. 0. 0. No subject, predicative can not be impersonal—rearrange-
 ment required

-
0. 0. 1. 0. Predicate governs accusative, nominative/accusative
 potential object
 0. 0. 2. 0. Predicate governs genitive, genitive potential object
 0. 0. 3. 0. Predicate governs accusative plural potential object
 0. 0. 4. 0. Predicate governs accusative, genitive singular/nominative-
 accusative plural potential object, ambiguity resolved by
 numeral
 0. 0. 5. 0 Predicate governing accusative contains negative particle,
 nominative/accusative potential object

0.0.6.0.	Predicate governing accusative contains negative particle, genitive potential object
0.0.7.0.	Predicate governs instrumental, potential instrumental object
0.0.8.0.	Predicate governs dative, potential dative object
0.0.9.0.	Predicate governs preposition, prepositional block present as potential object

- 40

Test Sentences for Main Syntax Pass—Key to Number Code

- | | |
|-----------|---|
| 0.0.0.13. | Instrumental candidate for object eliminated because of preceding other nominal |
| 0.0.0.14. | Dative candidate for object eliminated because of preceding preposition |
| 0.0.0.15. | Dative candidate for object eliminated because of preceding other nominal |

Number followed by "м ": nominal block includes modifiers

Test Sentences for Main Syntax Pass

- | | |
|-----------|--|
| 1.1.1. | Товарищи увеличивают союз. |
| 1.1м.1м. | Наши товарищи увеличивают социалистический союз. |
| 1.1.1.1. | Через районы товарищи увеличивают союз. |
| 1.1.1.1м. | Через большие районы товарищи увеличивают союз. |
| 1.1.1.2. | Через сопротивления товарищи увеличивают союз. |
| 1.1.1.2м. | Через большие сопротивления товарищи увеличивают союз. |
| 1.1.1.3. | В смысле такого сопротивления товарищи увеличивают союз. |
| 1.1.1.8. | Товарищи увеличивают через район союз. |
| 1.1.1.9. | Товарищи увеличивают в смысле такого сопротивления союз. |
| 1.1.2. | Товарищи достигают усилий. |
| 1.1.2.10. | Товарищи достигают без трудностей усилий. |
| 1.1.2.11. | Товарищи достигают постановкой трудностей усилий. |
| 1.1.3. | Товарищи увеличивают потери. |
| 1.1.3.8. | Товарищи увеличивают через район потери. |
| 1.1.3.9. | Товарищи увеличивают в смысле такого сопротивления потери. |
| 1.1.4. | Товарищи увеличивают две потери. |

Test Sentences for Main Syntax Pass

- 1.1.5. Товарищи не увеличивают союз.
1.1.6. Товарищи не увеличивают союза.
1.1.7. Товарищи служат техниками.
1.1.7.12. Товарищи служат с поправкой техниками.
1.1.7.13. Товарищи служат без управления состоянием техниками.
1.1.8. Товарищи отвечают сотрудникам.
1.1.8.14. Товарищи отвечают по составу сотрудникам.
1.1.8.15. Товарищи отвечают без подтверждения сотрудникам составу.
1.1.9. Товарищи отвечают на такой вопрос.
1.2.1. Такие сопротивления увеличивают союз.
1.2.1.1. Через районы такие сопротивления увеличивают союз.
1.2.1.2. Через сопротивления такие партии увеличивают союз.
1.2.1.3. В смысле такого сопротивления такие партии увеличивают союз.
1.3.1. Два сопротивления увеличивают союз.
1.3.1.1. Через районы два сопротивления увеличивают союз.
1.3.1.2. Через сопротивления две партии увеличивают союз.
1.3.1.3. В смысле такого сопротивления две партии увеличивают союз.
1.9. Союз увеличивают товарищи.
2.1.1. Товарищи сравнивали бы метод.
2.1м.1м. Нани товарищи сравнивали бы научный метод.
2.1.1.1. Через примеры товарищи сравнивали бы метод.
2.1.1.2. Через разложения товарищи сравнивали бы метод.
2.1.1.3. В смысле такого разложения товарищи сравнивали бы метод.
2.2.1. Такие сопротивления провели бы метод.
2.2.1.1. Через примеры такие сопротивления провели бы метод.
2.2.1.2. Через сопротивления такие примеры провели бы метод.
2.2.1.3. В смысле такого сопротивления такие примеры провели бы метод.

Test Sentences for Main Syntax Pass

- 2.3.1. Два сопротивления провели бы метод.
2.3.1.1. Через примеры два сопротивления провели бы метод.
2.3.1.2. Через сопротивления две партии провели бы метод.
2.3.1.3. В смысле такого сопротивления две партии провели бы метод.
2.9. Метод провели бы товарищи.
3.0.1. Они сравнивали метод.
4.4.1. Пример показывает метод.
4.4.1.4. Через пример мера показывает метод.
4.4.1.4м. Через такой пример мера показывает метод.
4.8. Ему удастся.
4.9. Пользу показывает метод.
5.5.1. Мера показала метод.
5.5.1.5. Через медь мера показала метод.
5.5.1.5м. Через простую медь мера показала метод.
5.9. Пользу показала мера.
6.6.1. Союз показал метод.
6.6.1.6. Через пример союз показал метод.
6.6.1.6м. Через такой пример союз показал метод.
6.9. Пользу показал метод.
7.7.1. Сравнение показало метод.
7.7.1.7. Через понятие сравнение показало метод.
7.7.1.7м. Через такое понятие сравнение показало метод.
7.8. Ему удалось.
7.9. Пользу показало сравнение.
8.1. Товарищи могли быть.
8.9. Могли быть товарищи.
9.1. Товарищи могли быть показаны.
9.9. Могли быть показаны товарищи.
10.1. Товарищи были показаны.
10.9. Были показаны товарищи.
11.1. Товарищи будут показаны.

Test Sentences for Main Syntax Pass

- | | |
|---------|----------------------------------|
| 11.9. | Будут показаны товарищи. |
| 12.1.2. | Товарищи быстрее союза. |
| 12.4.2. | Мера быстрее метода. |
| 12.9. | Союза быстрее товарищи. |
| 13.0.1. | Если увеличивать союз. |
| 14.0.1. | Если бы увеличивать союз. |
| 15.1.1. | Товарищи могут увеличивать союз. |
| 15.8.1. | Ему удастся увеличивать союз. |
| 15.9. | Союз могут увеличивать товарищи. |
| 16.0.1. | Создавая союз. |
| 16.0.1. | Задавался создавать союз. |

Test Sentences for Cleanup Pass CP-1

- | | |
|----|--|
| 1. | Важные товарищи идут. |
| 2. | Очень важные товарищи идут. |
| 3. | Не товарищи идут. |
| 4. | Важные товарищи наших сотрудников идут. |
| 5. | Очень важные товарищи наших сотрудников идут. |
| 6. | Не товарищи наших сотрудников идут. |
| 7. | Не важные товарищи идут. |
| 8. | Не важные товарищи наших сотрудников идут. |
| 9. | Товарищи важных профессоров наших сотрудников
идут. |

Test Sentences for Cleanup Pass CP-2

- | | |
|----|---|
| 1. | Нас видеть нельзя. |
| 2. | Простую медь видеть нельзя. |
| 3. | Простую медь такого рода видеть нельзя. |
| 4. | Такого рода простую медь видеть нельзя. |
| 5. | Очень простую медь видеть нельзя. |
| 6. | Полезную медь видеть нельзя. |

HARDWARE IMPLICATIONS OF THE RESEARCH

Machine translation research at the RW Division has been implemented on general purpose computers. We recognize such hardware requirements as print readers, multi-font printers, and composing machines.

We have recognized the need for larger memories. This recommendation, however, has been the general observance—the larger the memory the easier the solution.

In the areas of dictionary search, we have been able to program our way around the fact that the general purpose computers that we have been using for high-speed memories are too small to contain our lexicon.

Both programming foresight and cleverness were required for the general solution of our ordinary dictionary lookup. However, if one wishes that a far larger memory had been available, it becomes possible to ask what new attributes other than larger size are desired of this memory.

What strikes one as the flow charts are examined, is that the need exists for a totally different type of computer. Such a computer would have an associative memory.

The reasons which first appear for an associative memory are those based upon the facts that the data is non-numeric. Such instructions as:

find the word in the dictionary

or

find the prefix of the word

or

compare the end of the word with a possible list of suffixes

immediately present themselves. Their uses are real; our experience in machine translation, however, has led up to the point where those portions of the program which require the ability to recognize alpha-numerical configurations require a small portion of the total running time.

Attention is directed to the figures on the following pages. These charts are taken from the 1960 RW machine translation report. They illustrate a small portion of the syntactic program. Furthermore, these charts represent the thinking of a linguist and had not as yet been seriously modified by a programmer's restatement based on efficient usage of a traditional general purpose computer.

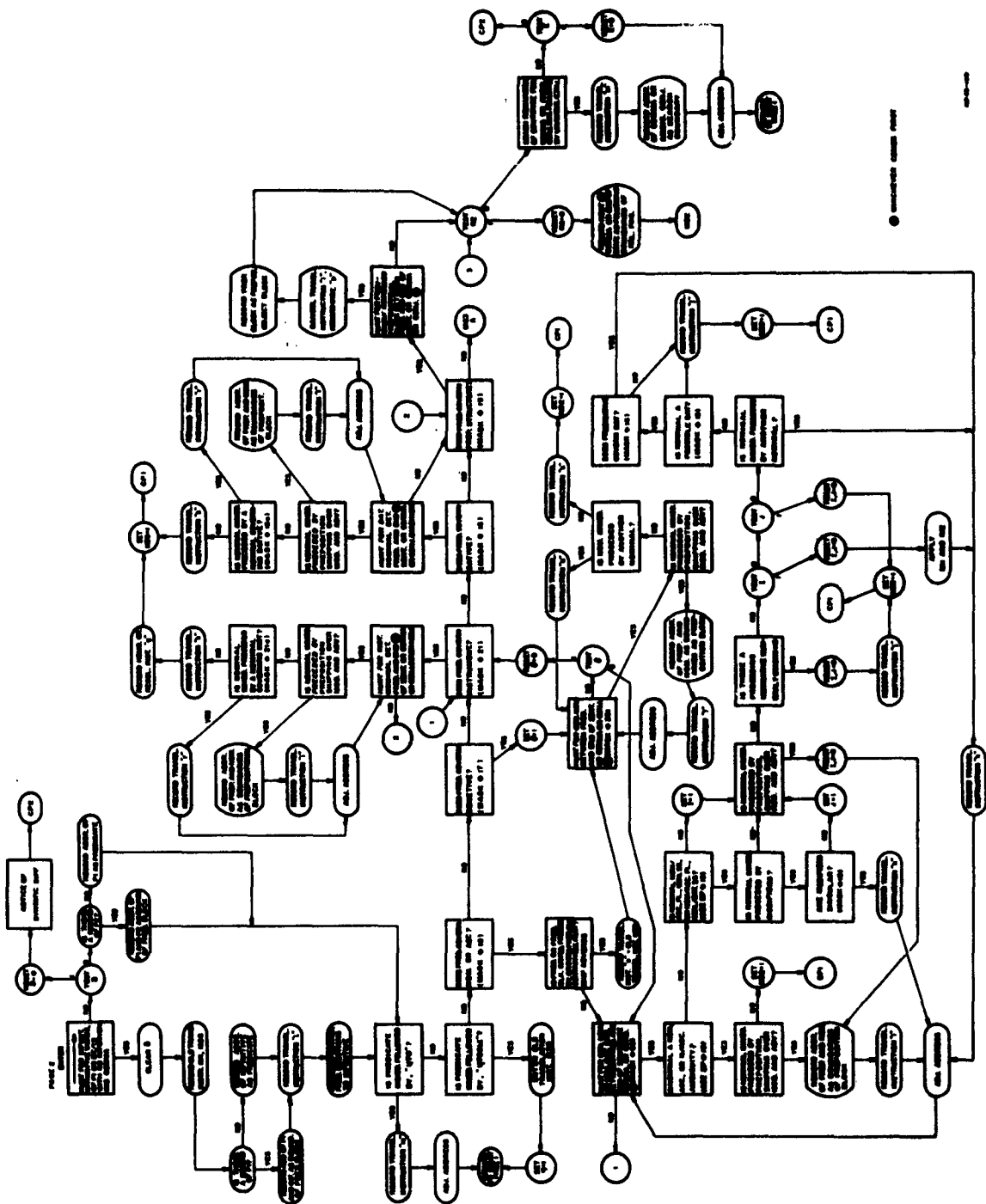
Though there is an occasional reference to actual word forms in these figures, most references are to grammatic attributes which would have been found only after a successful dictionary lookup.

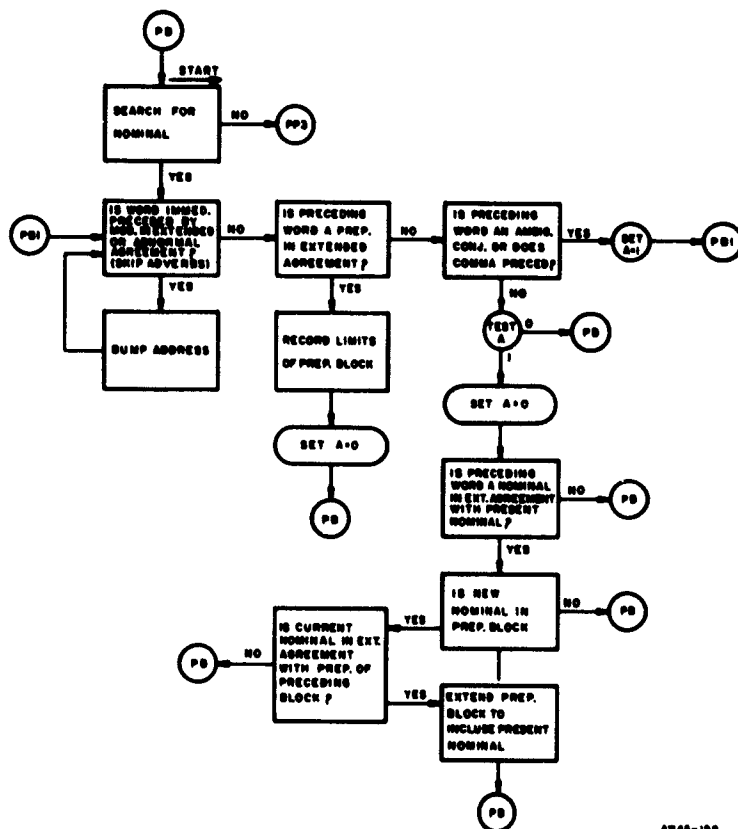
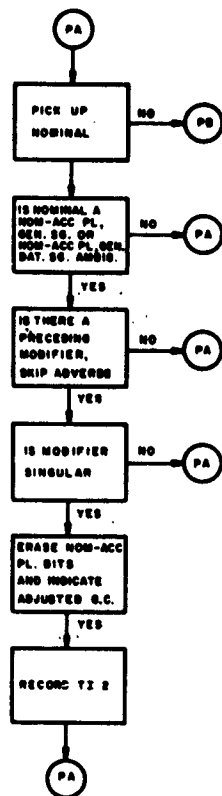
In these charts, a large number of the boxes begin with such words as "search," "hunt," or by phrases such as "is nominal immediately preceded by preposition skipping over modifiers and adverbs." Here we have evidence of the need for a search type memory past the stage where the problem has consumed alphanumeric data.

Among the important aspects revealed by the flow charts is the need for a search to be made within a region whose boundaries are defined syntactically. This searching within a non-graphically defined area is one of the dominant features of language data processing; this is made evident as we examine different portions of our bilingual programs.

Syntactic Analysis

It is generally recognized by all research centers that automatic translation cannot be done without performing detailed syntactic analyses of the text. In syntactic analyses searches are made on other than lexical units. These searches, made upon grammar codes and resultant syntactic codes, have as part of their nature a conjunction of conditions within variably defined syntactic boundaries. Up to now, the prominent methods of analysis have necessarily reverted to standard computer techniques. Yet, as indicated in the following figures, this portion of the analysis, at least in the Fulcrum Technique, is heavily loaded with "search" and "hunt" procedures. In this area of the translation procedure, one would expect to find heavily used associative memory type algorithms and a different word size and





memory module size indicated from that of the earlier portions of the program.

Semantics

The portion of all machine translation programs which have been most impervious to solution is that represented by correctly erasing all but one of the possible English translations for a given foreign word. Some groups have even been dismayed enough to indicate that human intervention will always have to be required at this point. Our own research has led us to the use of techniques which with hindsight would appear to have called for a cascaded search. Among the problem types are homograph resolution, idiom identification, word combination recognition, and determiner-determinee paring.

Idiom recognition requires the knowledge that two or more words, when closely connected syntactically, possibly even contiguous, have a quite different meaning and possibly a different grammatic aspect than would be expected from the sum of the parts of the idiom. Here techniques that have been used include the dictionary labelling of each member of the idiom as being a potential idiom and then, when a conjunction of possible members occurs, searching in a special idiom table. Even the necessary identification of the idiom potentiality requires search procedures. The interaction between idiom identification and associative memory computer configurations should be studied.

Homograph Resolution

Many words share the same spelling. A dramatic example in English are the three different meanings of the spelling "fast." As a verb, it means to not eat; as an adjective, it contains the two opposite meanings of tied tight and rapid. Resolution of this type of ambiguity is often made by searching in the neighborhood of the word for indications of the word class of the homograph. Such resolution would be more easily effectuated with a search memory.

Word Combinations

Word combinations represent an idiom type whose configuration is both lexical and syntactic; that is, the correct translation of a primary word or of associated prepositions is determined by the existence of the primary word and a particular type of syntactic environment. This environment can be represented, for example, by a word in the same clause with governed genitive and dative dependencies, but with an explicitly missing accusative object. Since indications of the limits of this environment are only vaguely fixed, search is again called for.

Determiner-Determinee

More than any other type of semantic resolution described thus far, determiner-determinee resolution requires associative memory techniques as can be seen in the flow charts for the method. Here the resolution problem is represented by the existence within wide syntactic boundaries of two or more semantically connected words. The "hunt" initially consists of ascertaining whether a determinee exists. Since determinees are capable of having different determiners, the next aspect of the search directs one to "hunting" first the identification of the possible determiners, next the establishment of their coexistence and then, finally, verification of whether they exist in proper syntactic relationship to the determinee. Determiner-determinee resolution has been awkward to program on a standard computer. Its resolution should be far more powerful when associative memories are used.

Output

A consumer of automatic translation thinks of a printed page as being the output. This is only partially true. In truth, the printed translation is the fruit of the process. The necessary evolutionary improvements receive their sustenance from the research output. Since nature of handling the research output almost invariably calls

for the identification of particular translation configurations in a large body of text, this most necessary area of automatic translation would seem to benefit greatly from any hardware development which eases the burdens of search. Here one would study the proper way of emitting the research output so that it would be best married to the coming hardware realities.

Dictionary Lookup

The problem of dictionary lookup in RW machine translation has two major aspects. Most Russian words found in the text that we process actually exist in our dictionary. A sizeable percentage of the words, however, are represented in the dictionary with a different ending. By stripping off the possible suffixes of text words which are not found directly by lookup and by then searching on stems, the semantic and morphological aspects of these words can be derived. The use of a search type memory in this portion of the problem is obvious.

DICTIONARY UPDATING

Our initial machine glossary has been updated in eight separate runs. This added 6,909 new forms to the dictionary, which now contains 23,713 forms. Since our stem-affixing procedures allow us to grammar-code new forms from different forms of the same stem found in the dictionary, this will give us a coverage of about 50,000 Russian forms.

SAMPLE TRANSLATION

UNPARALLELED SCIENTIFIC FEAT

PHOTOGRAPHS INVISIBLE FROM GROUND OF THE PART OF MOON E PLACED TODAY IN ** TRUTH.**. MATERIALS
PHOTOGRAPHS EARTH BATCH FEATURED
CONCERNING THE NEW REMARKABLE VICTORY OF SOVIET CONQUERORS OF THE COSMIC SPACE ARE PUBLISHED. PHOTOGRAPHIC IMAGES
ABOUT OF

OF THE INVERSE SIDE OF MOON, INVISIBLE FROM OUR PLANET WERE OBTAINED FIRST IN HISTORY OF THE MARKING
OPPOSITE DIRECTION INTO
AS A RESULT OF HEROIC STRESSES OF SOVIET SCIENTIFIC DESIGNERS, ENGINEERS, TECHNICIANS AND OPERATE, HAVING CREATED
LEARNED,

AUTOMATIC INTEPLANETARY STATION.

ORBIT BY THE AUTOMATIC INTEPLANETARY STATION OF MOON AND GROUND S. THE PHOTOGRAPHY OF THE INVERSE
EARTHS OPPOSITE
SIDE OF MOON FROM FIGHTS THIS COSMIC LABORATORY FROM DISTANCE IN (NTS) OF THOUSANDS OF KILOWETERS. THE
DIRECTION INTO
TRANSFER OF IMAGES ON GROUND FROM DEPTSH OF THE COSMOS. EACH OF THESE ACHIEVEMENTS THE UNPARALLELED
EARTH OF ATTAINMENTS

PAGE 002

SCIENTIFIC FEAT. AND THE FACT THAT, HE WAS ACHIEVED BY SOVIET PEOPLE, THE PHENOMENON DEEP NATURAL IS. HE
IT ACCOMPLISHED REGULAR ARE IT

WITNESSES CONCERNING UNSURPASSED SUCCESSES OF THE MATHEMATICS, MECHANICS, PHYSICISTS, RADIO ELECTRONICS,
TESTIFIES ABOUT MATHETICIANS
INDICATES OF

CHEMISTRIES, ACHIEVEMENTS OF SOVIET METALLURGISTS, ASTRONOMERS AND REPRESENTATIVES OF OTHER SCIENCES, OUR
ATTAINMENTS

ENGINEERS, TECHNICIANS AND OPERATE, WHICH EMBODIED THEORETICAL CALCULATIONS INTO THE RAFT OF TECHNICAL
WHO FLOAT

CONSTRUCTIONS, ARE INTO OUR TIME BY THE CROWN OF ACHIEVEMENTS OF THE HUMAN GENIU. HE IS THE EXPRESSION
BEING PERIOD IT
APPEARING ATTAINMENTS

OF THE HIGH INDUSTRIAL POWER OF OUR COUNTRY, HAVING BEEN ABLE TO CREATE THE EQUIPMENT, APPARATUS AND MATERIALS,
NECESSARY FOR THE REALIZATION OF THESE VAST PROJECTS. HE SIGNIFIES ITSELF UNQUESTIONABLE ADVANTAGES OF THE
EXECUTION IT OURSELVES

SOVIET SOCIALIST OPERATION, GRAPHICLY DEMONSTRATING THE ALL WORLD, ON THAT IS ABLE THE CREATIVE CREATIVE
SYSTEM WHOLE CAPABLE CONSTRUCTIVE
FORMATION GIFTED
ORDER CLEVER

LABOR OF THE PEOPLE, FREE FROM SHACKLES OF THE CAPITALISTIC SUPPRESSION.

CONRADE N. S. KHRUSHCHEV, WRITED ** BECAUSE SOVIET PEOPLE, FIRST IN THE WORLD SUCCESSFULLY SOLVED
WHY PEACE PERMITTED

SO THE TRULY GREATEST PROBLEM OF THE PARCEL ROCKET ON MOON, IS, AND INDEED THIS THE PROBLEM WAS WITH MUCH
 PACKAGE ARE THIS IS MANY

UNKNOWN. THIS VICTORY BECAME POSSIBLE AS A RESULT OF THE FACT THAT, THE SAME SOVIET PEOPLE BY ONES HANDS, BY
 BEGAN ITS MY OUR HIS

ONES HEROIC LABOR INTO THE SHORTEST HISTORICAL TIME WERE ABLE TO SOLVE THE GREATEST SOCIAL PROBLEM IS THEY
 ITS DIFFICULTY EARLIER DATE TERM ARE

56
 CONSTRUCTED THE SOCIALIST SOCIETY AND CONFIDENTLY BUILD COMMUNISM.**
 BUILT

EPOCH-MAKING SUCCESSES OF OUR SCIENTIFIC EXPERTS AND OPERATE THE GREAT VICTORY OF ALL THE SOVIET
 ADVANCEMENTS LEARNED, THE WHOLE

PEOPLE, BUILDING COMMUNISM UNDER THE DIRECTION OF THE LENINIST PARTY. THIS THE OUTSTANDING INVESTMENT IN
 CONTRIBUTION

THE DEVELOPMENT WORLD OF SCIENCE. PROGRESSIVE PEOPLE OF ALL THE WORLD ENTHUSIASTICALLY APPLAUD SOVIET
 GROWTH THE WHOLE PEACE

CONQUEAORS OF THE COSMOS, CONGRATULATING BY THEM WITH REMARKABLE SUCCESSES.

FAMOU, THAT NEW UNSURPASSED VICTORYS OF SOVIET SCIENCE AND TECHNICIANS ARE COMPLETED THE DAY BEFORE THE

TURNING POST	STATION	ITEM	PLANK
1	100	1	1
2	200	2	2
3	300	3	3
4	400	4	4
5	500	5	5
6	600	6	6
7	700	7	7
8	800	8	8
9	900	9	9
10	1000	10	10
11	1100	11	11
12	1200	12	12
13	1300	13	13
14	1400	14	14
15	1500	15	15
16	1600	16	16
17	1700	17	17
18	1800	18	18
19	1900	19	19
20	2000	20	20
21	2100	21	21
22	2200	22	22
23	2300	23	23
24	2400	24	24
25	2500	25	25
26	2600	26	26
27	2700	27	27
28	2800	28	28
29	2900	29	29
30	3000	30	30
31	3100	31	31
32	3200	32	32
33	3300	33	33
34	3400	34	34
35	3500	35	35
36	3600	36	36
37	3700	37	37
38	3800	38	38
39	3900	39	39
40	4000	40	40
41	4100	41	41
42	4200	42	42
43	4300	43	43
44	4400	44	44
45	4500	45	45
46	4600	46	46
47	4700	47	47
48	4800	48	48
49	4900	49	49
50	5000	50	50
51	5100	51	51
52	5200	52	52
53	5300	53	53
54	5400	54	54
55	5500	55	55
56	5600	56	56
57	5700	57	57
58	5800	58	58
59	5900	59	59
60	6000	60	60
61	6100	61	61
62	6200	62	62
63	6300	63	63
64	6400	64	64
65	6500	65	65
66	6600	66	66
67	6700	67	67
68	6800	68	68
69	6900	69	69
70	7000	70	70
71	7100	71	71
72	7200	72	72
73	7300	73	73
74	7400	74	74
75	7500	75	75
76	7600	76	76
77	7700	77	77
78	7800	78	78
79	7900	79	79
80	8000	80	80
81	8100	81	81
82	8200	82	82
83	8300	83	83
84	8400	84	84
85	8500	85	85
86	8600	86	86
87	8700	87	87
88	8800	88	88
89	8900	89	89
90	9000	90	90
91	9100	91	91
92	9200	92	92
93	9300	93	93
94	9400	94	94
95	9500	95	95
96	9600	96	96
97	9700	97	97
98	9800	98	98
99	9900	99	99
100	1000		

TIME BRIGHT SHOW, THAT THE GREAT OCTOBER SOCIALIST REVOLUTION, TO PUT AN END WITH EXPLOITATIONS OF THE MAN
INTENSE INDICATE FROM USES
TESTIFY
READ

BY THE MAN, OPENED BEFORE PEOPLES THE PRAVILNYJ PATH OF THE DEVELOPMENT + I THE PATH THE BILLION THE MAN
DISCOVERED IN FRONT GROWTH IS+

**0C000000 0U000000CJ0C000000C00000000000000000 AT PRESENT FROM ABOVE OF THE BUILDING OF SOCIALISM AND ALSO
PRESENTLY PROJECT**

COMMUNISM, LEADING TO THE UNPRECEDENTED FLOURISHING OF THE ECONOMY, CULTURE AND THE WELFARE OF PEOPLE, THE WELL-BEING AND PROSPERITY ALSO

PATH, IN WHICH, FOLLOW
ON WITH RESPECT TO

OUR COUNTRY FOR FORTY TWO YEAR AFTER THE OVERTHROW OF THE POWER OF CAPITALISTS AND LANDOWNERS AND DURING PER BEYOND

DETERMINATIONS OF THE SOVIET WORKERS AND PEASANTS POWER WENT ANY ENORMOUS PATH IN ONES DEVELOPMENT. FROM THE WHICH ITS HIS THEIR OF SOME THEIR

COUNTRY BACKWARD AND RUINED SHE WAS CONVERTED INTO THE ALL-POWERFUL POWER OF THE WORLD, SUCCESSES OF WHICH IN
IT PEACE ADVANCEMENTS
REGION OF THE ECONOMY, SCIENCES AND CULTURES STARTLE ALL THE MANKIND. THE FLIGHT THIRD SOVIET COSMIC ROCKET WAS
THE WHOLE
WERED BY THE STARTLING EXPRESSION OF THIS POWER, ECLIPSED ITSELF ALL THE FEATS, PERFECTED TO THIS IN THE
THEMSELVES THE WHOLE BEFORE
OURSELVES

NAME OF SCIENCE AND THE PROGRESS.

THE MULTISTAGE ROCKET CREATED BY SOVIET PEOPLE SENT THE INTERPLANETARY STATION ON THE BEFOREHAND
CALCULATED ORBIT AND DIRECTED BY IT INTO THE ORBIT OF MOON. THE STATION GUIDED FROM GROUND WENT NEAR MOON,
HER ALSO CONTROLLED EARTH
HAVING LOOKED ON ITS OPPOSITE SIDE, WHICH NEVER STILL VIDEL THE HUMAN EYE. THE SPECIAL SYSTEM OF THE
HER DIRECTION YET
ORIENTATION LED THE STATION INTO SUCH POSITION, AT WHICH THE LUNAR DISK HIT INTO THE FIELD OF THE VISION OF
ENTERED SIGHT EYESIGHT VIEW

LENS. DURING FORTY MINUTES WITH DIFFERENT EXPOSURES, THAT RDER NOT TO ERR, THE SOVIET COSMIC PHOTOCORRESPONDENT
DISTINCT THAT
LED THE RECORDING OF THAT SIDE OF MOON, CONCERNING WHICH WE UNTIL NOW NOTHING KNOWN. THE AUTOMATIC
VEL DIRECTION ABOUT OF

PAGE 006

PHOTOGRAPHIC APPARATUS DEVELOPED AND DRICD THE FILM, STAMPING INVALUABLE FRAMES. THEN IN THE COMMAND
ON WITH RESPECT TO

FROM GROUP'D THE SPECIAL SYSTEM TRANSMITTED OBTAINED IMAGES THE LUNAR SURFACE. HAVING COVERED COLOSSAL SPACES INTO
EARTH

HUNDREDS OF THOUSANDS OF KILOMETERS, THESE IMAGES REACHED GROUND S, WERE RECEIVED.
EARTHS

THE SOVIET SCIENCE AND THE TECHNOLOGY OF OUR DAYS IS ABLE TO PERFORM SO THAT. PROOFS OF ITS
TECHNICIAN HERE HER

50 SUCCESSES, ITS SUPERIORITY IN MOST IMPORTANT BRANCHES OF HUMAN KNOWLEDGE ARE STILL NECESSARY ANY. WHICH
HER YET SOME WHAT

MOON FROM IMMEMORIAL TIMES ATTRACTED THE ATTENTION OF THE MAN. STILL INTO ANCIENT TIMES
PERIODS YET PERIODS

PHILOSOPHERS EXPRESSED THE CORRECT IDEA, THAT MOON THE INDEPENDENT CELESTIAL BODY, THE PRACTICALLY SPHERICAL
FORM, CAN BE, SIMILAR TO GROUND. THE NEW PERIOD IN THE STUDY OF OUR NATURAL SATELLITE BEGAN IN (NTS) THE YEAR,
EARTH INTO

WHEN GALILEO DIRECTED ON MOON ONES FIRST PRIMITIVE TELESCOPE. HE DISCOVERED ON MOON OF THE PLAIN AND
IN WHICH IT
MY OUR

MOUNTAINS. FROM THIS MOMENT IT BEGAN THE CREATION OF NEW SCIENCE IS THE SELENOGRAPHY, OCCUPIED BY THE STUDY
MOMENTUM ARE ENGAGED IN CONCERNED

OF FORMATIONS AT SURFACE OF MOON.
PRODUCTIONS

FOR THREE WITH THE HALF CENTURYS, ELAPSED AFTER FIRST TELESCOPIC OBSERVATIONS, THE SELENOGRAPHY WAS
DURING PAST
PER
BEYOND

MADE ENORMOUS SUCCESSES. THESE ACHIEVEMENTS ARE TODAY CROWN BY THE UNPARALLELED SCIENTIFIC VICTORY, WONNED
DID ADVANCEMENTS ATTAINMENTS

SOVIET INVESTIGATORS OF THE COSMOS. THE NEW EPOCH IN THE STUDY OF CELESTIAL BODIES WAS BEGANNED.
RESEARCHERS

THE ALREADY PRELIMINARY ACQUAINTANCE WITH FIRST PHOTOGRAPHS OF THE INVERSE SIDE OF MOON PERMITS
EVEN OPPOSITE DIRECTION

BY SOVIET SCIENTIST TO MAKE IMPORTANT CONCLUSIONS CONCERNING PECULIARITIES OF ITS SURFACE. THE FURTHER
SCIENTIFIC CO DERIVATIONS ABOUT PARTICULARITIES HER BELOW FOLLOWING

STUDY OF DETAILS OF THE SURFACE OF THE INVERSE SIDE OF MOON WIDENS COGNITIONS OF THE MAN
INVESTIGATION COMPONENTS OPPOSITE DIRECTION BROADEN KNOWLEDGES
CONCERNING THE DEVELOPMENT OF PLANETS. BUT ALREADY AND NOW SELENOTRAFI- WAS ENRICHED BY NEW NAMES, APPROVED
ABOUT EVEN ASSERTED
OF

PAGE 008

BY THE COMMITTEE OF SCIENCES OF THE USSR. THEY STAMPED VICTORIES OF SOVIET SCIENCE, WHICH BROUGHTED CONTRIBUTED CARRIED

THE INVALUABLE INVESTMENT IN WORLD SCIENCE.
CONTRIBUTION

THE GREAT CRATER SEA IS CALLED BY THE SEA OF MOSCOW IS BY THE NAME OF THE CAPITAL OF OUR NATIVE LAND,
LARGE ARE

FIRST IN THE WORLD OF THE SOCIALIST GOVERNMENT, HAVING CONSTRUCTED THE PATH TO THE CONQUEST OF THE COSMOS. INTO
PEACE

THE SEA OF MOSCOW IS THE 2 HAVING REDDENED ASTRONAUTS. WHO CAN DOUBT THE FACT THAT, BY THE FIRST COSMIC BODY,
ARE

WHICH VISIT IN THE FUTURE CONQUERORS OF THE COSMOS, WILL MOON. CRATERS CIOLKOVSKIJ, BY SOLIO- CURIE IMMORTALIZE
COMING
NEXT

ITSELF EXISTS GREAT SCIENTISTS. THEIR LABORS HAD THE ENORMOUS VALUE FOR THE DEVELOPMENT OF SCIENCE,
THEMSELVES LEARNED MEANING GROWTH
OURSELVES

AND ALSO THOSE ITS BRANCHES, WHICH IED AT PRESENT TO THE BEGINNING OF THE CONQUEST BY THE MAN OF THE
HER WHO PRESENTLY ACHIEVEMENT
CIRCUMSOLAR SPACE. THE MOUNTAIN MOUNTAIN RANGE SOVIET WILL ALWAYS TO REMIND THE MANKIND OF FEATS OF SOVIET
PUT ONE IN MIND

PEOPLE, CREATING THE COMMUNIST SOCIETY. THE SEA OF DREAM THIS INTO THE HONOR FIRST SOVIET COSMIC ROCKET,
THIS IS

WHICH BECAME BY THE FIRST ARTIFICIAL PLANET. WHETHER THIS THE NAME CONCERNING THE FACT THAT, DOES NOT SAY
BEGAN OR THIS IS ABOUT SPEAKS

SOVIET PEOPLE CONVERT AGE-LONG DREAMS OF THE MANKIND INTO THE REALITY, THAT WE ENTERED INTO THE EPOCH OF THE
ACTUALITY

UNPRECEDENTED FLOURISHING OF SCIENCE AND TECHNICIANS, ECONOMYS AND CULTURES, WHEN TRULY TALES BECOME BY THE
BEGIN IN WHICH

FACT.

AS CALCULATED SCIENTISTS, THE HALF A YEAR AND EXISTS AUTOMATIC INTEPLANETARY STATION WILL THEN BURN UP
HON SCIENTIFIC

IN DENSE LAYERS OF THE ATMOSPHERE. BUT THE UNPARALLELED SCIENTIFIC FEAT OF SOVIET PEOPLE, REALIZ DARING FLIGHTS
OF COSMIC SHIPS INTO MOON AND INTO THE ORBIT OF MOON, WILL ALWAYS TO SHINE UNFAD BY THE LANDMARK IN HISTORY
ALSO INTO

OF THE STUDY OF NEVER-ENDING DEPTSH OF UNIVERSE.
INVESTIGATION

SOVIET SCIENTISTS, DESIGNERS, ENGINEERS, TECHNICIANS, WORKING, BECAME FAMOUS OUR NATIVE LAND FIRST IN
SCIENTIFIC INTO

HISTORY BY THE COSMIC FLIGHT INTO MOON, BY THE MAN OF THE COSMIC SPACE YES PROPER FROM- RYVIM THE NEW ERA IN THE
CONQUEST.

TODAY, IN THE DAY, WHEN ARE PUBLISHED FIRST RESULTS OF THE UNPARALLELED SCIENTIFIC EXPERIMENT, INTO IN WHICH

ZPKONIVJEGOS= BY NEW VICTORIES OF SOVIET SCIENCE AND TECHNICIANS, PROCEEDS WITH THE PAPER THE ALTERNATE SESSION ALSO
WORK NEXT

OF SUPREME SOVIET ELECTED REPRESENTATIVES OF THE PEOPLE DISCUSS MOST IMPORTANT QUESTIONS, CONNECTED WITH THE FROM COUPLED

BELOW
FURTHER DEVELOPMENT OF THE COMMUNIST BUILDING, IS POSSIBLE NOT TO DOUBT, THAT THE SESSION WILL TAKE SOLUTION, ASSUME DECISION PROJECT

WHICH GUARANTEE NEW LARGE-SCALE
WHO ENSURING LARGE-SCALE
OUTSTANDING

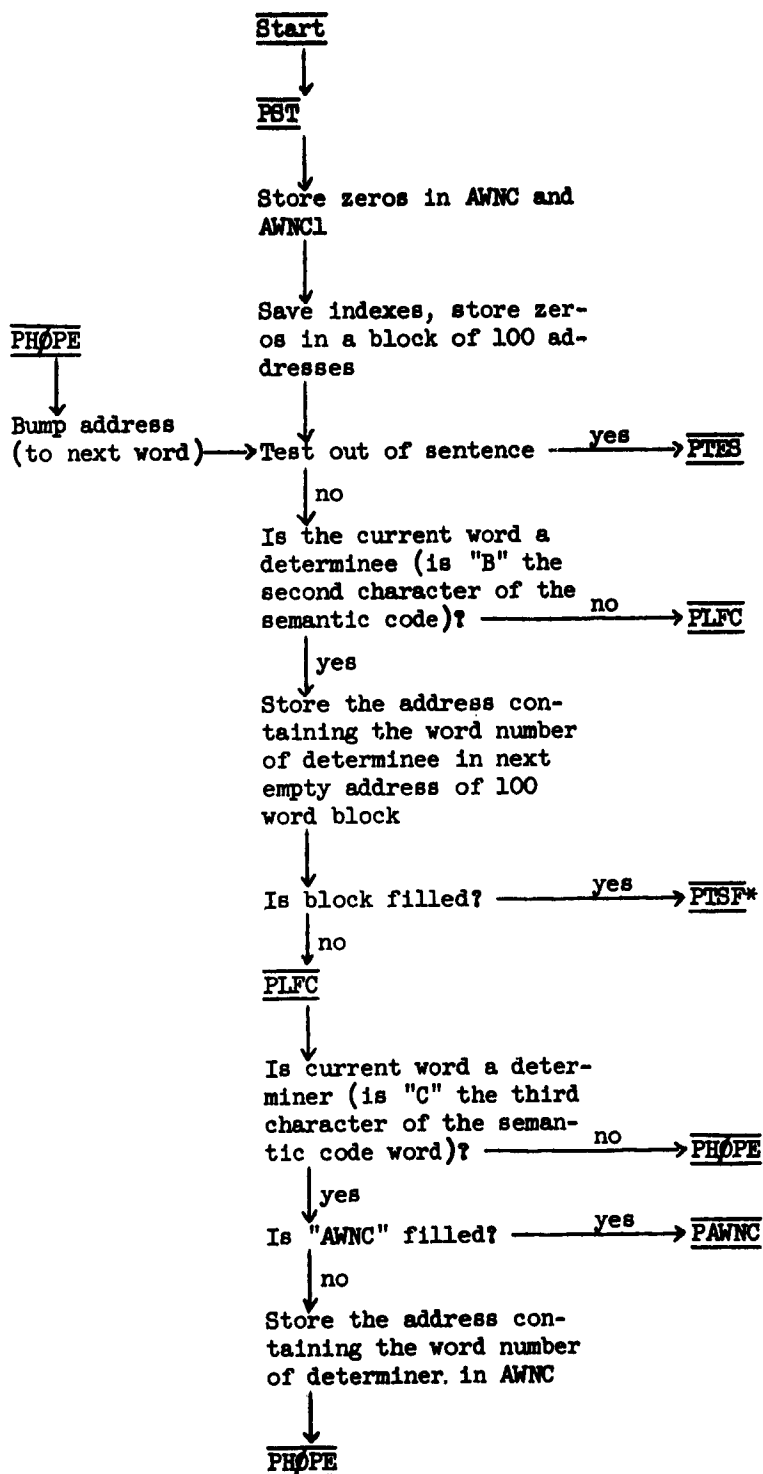
ECONOMY, SCIENCE, TECHNICIANS, CULTURE OF OUR COUNTRY. 00000000 00000000000000000000000000 00000000

YES PROPER'S OUR OWN SOVIET PRAVITEL • SGVO.
NATIVE

THE COMMUNIST PARTY OF SOVIET UNION THE GREAT INSPIRING AND LEADING FORCE OF THE SOVIET PEOPLE IN THE DIRECTING GUIDING

STRUGGLE FOR THE CONSTRUCTION OF COMMUNISM YES PROSPERS.

**FLOW CHARTS OF
DETERMINER-DETERMINEE METHOD**



PAWNC

↓
Store the address containing the word number of determiner in AWNCL

↓
PHOPE

PTES

↓
Is first address of 100 word block containing addresses of word numbers of determinees filled? ————— no —————→ ENT*

↓ yes
Is AWNCL filled? ————— no —————→ ENT*

↓ yes
Are the contents of first address of 100 word block and AWNCL equal? ————— no —————→ PBS

↓ yes
Is the second address of the 100 word block filled? ————— yes —————→ PBS

↓ no
Is AWNCL filled? ————— yes —————→ PBS

↓ no
ENT*

PBS

↓
Pick up next entry in 100 word block and pick up word number referred to there

↓
Is the word number less than the word number of the first entry in the determinee table? ————— yes —————→ PNIT*

↓ no
Is the word number the same as the first entry in the determinee table? ————— yes —————→ PLPM

↓ no
Is the word number more than last entry in the determinee table? ————— yes —————→ PNIT*

↓ no
Is the word number the same as the last entry in the determinee table? ————— yes —————→ PRPM

↓ no
PCLN

PCLN

↓
Add together the addresses of the top and bottom parameters of the table and divide by 2 (gives mid-point of table)

↓
Is the contents of this mid-point of the table more than the word number of the current determinee? ————— yes

→ PLHT

→ Make midpoint of the table the new bottom parameter

↓
PCLN

no
Is the contents of the midpoint equal to the word number of the current determinee? ————— yes

→ PMPM

no
Made the midpoint of the table the new top parameter

↓
Do the top parameter and the bottom parameter of the table differ by 1? ————— yes

→ PNIT*

no

PCLN

PLPM

↓
Save address of match

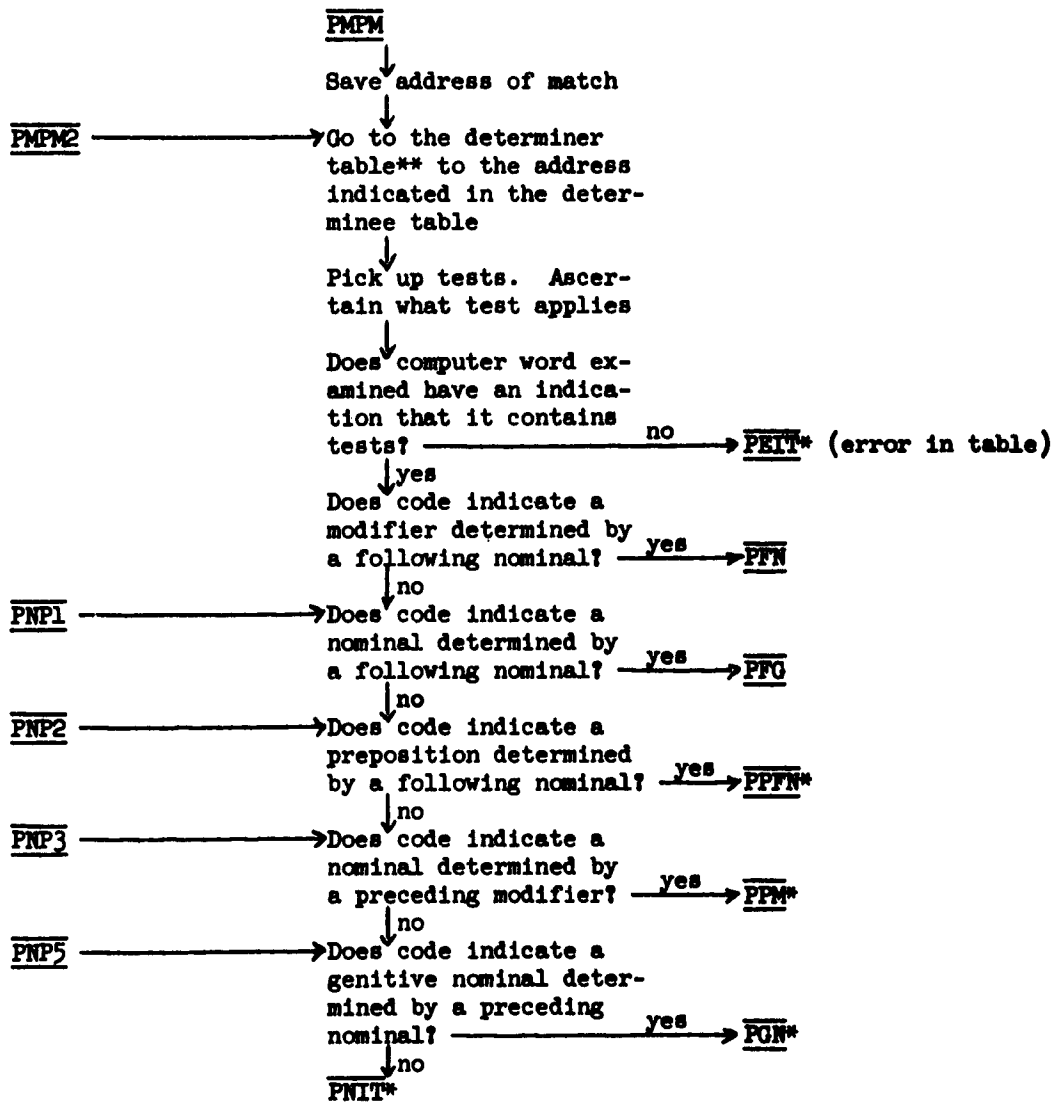
↓
PMPM+2

PRPM

↓
Save address of match

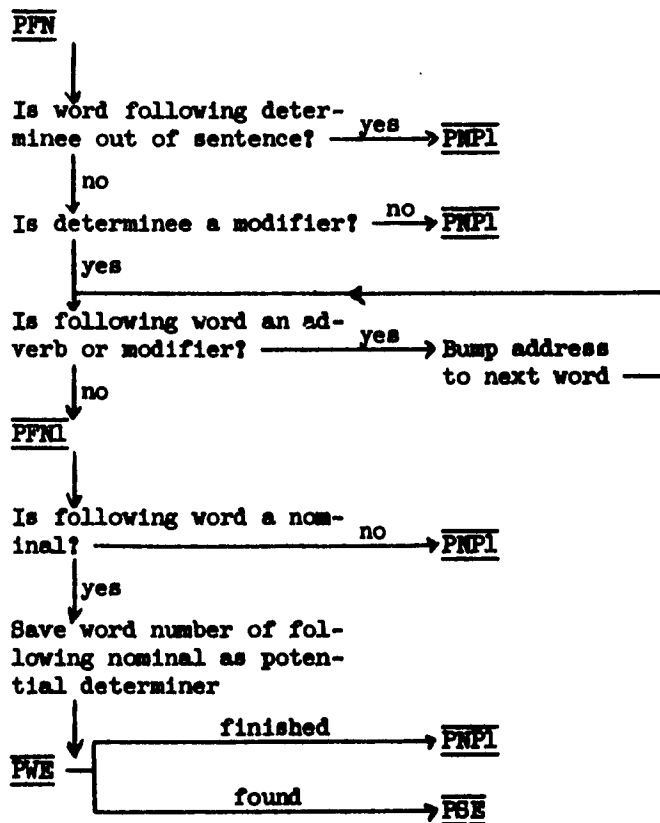
↓
PMPM+2

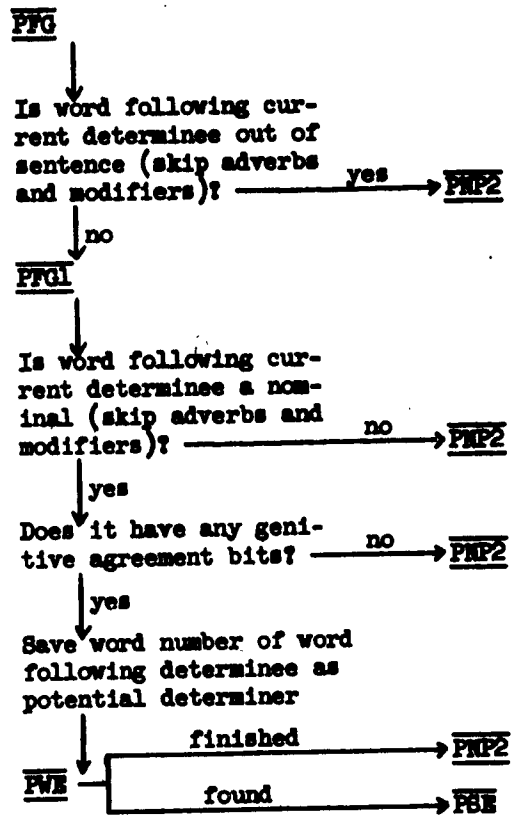
* Not yet written

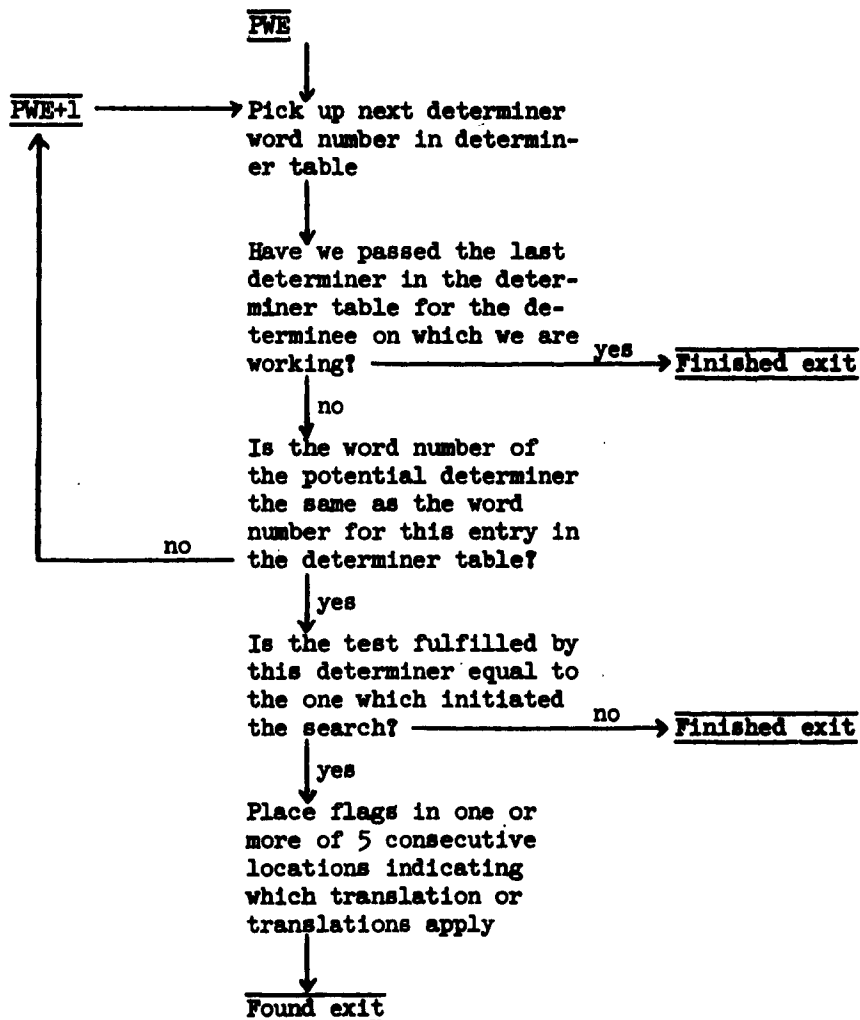


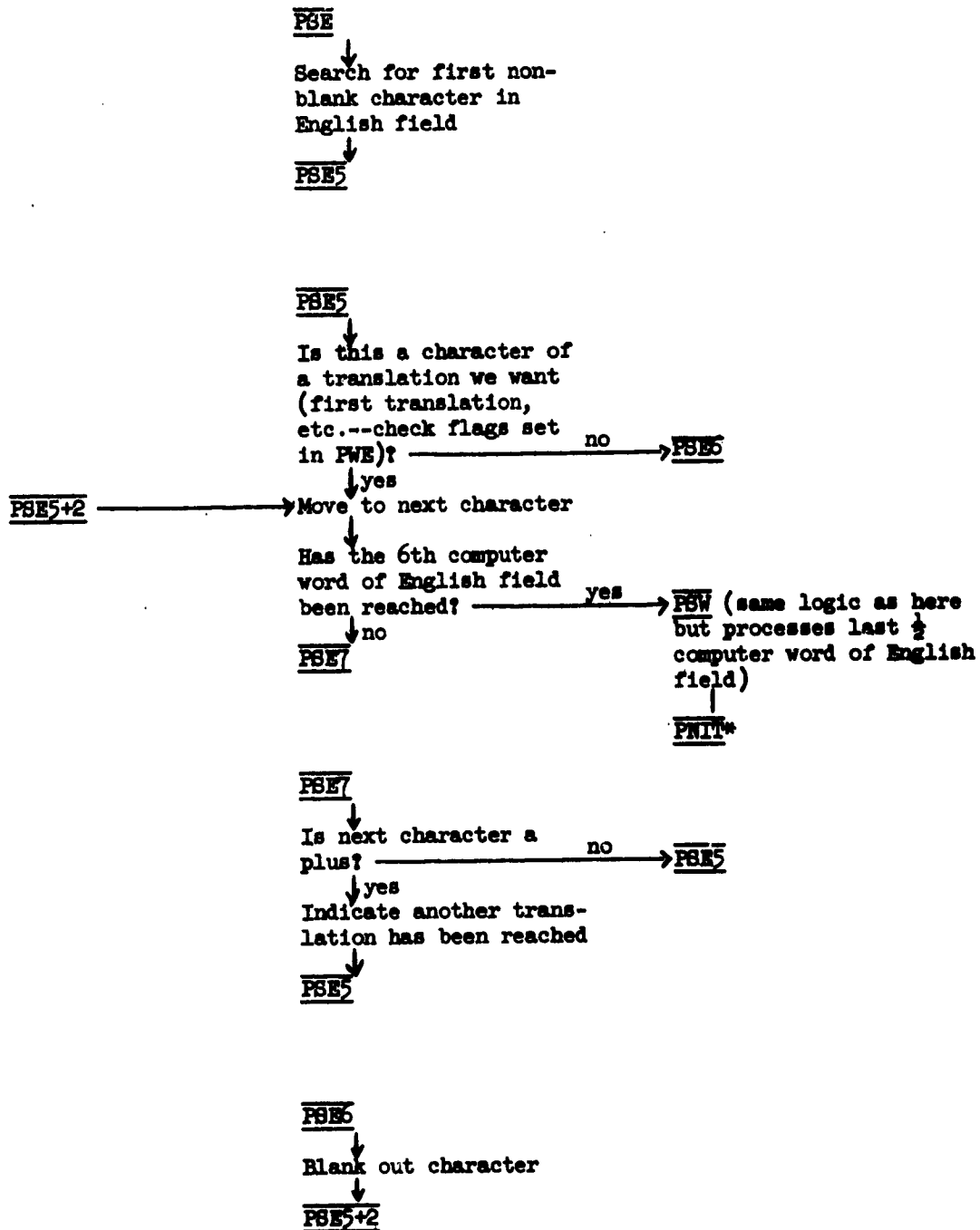
* Not yet written

**See page 5 for the format of the determiner table









CONCLUSIONS AND RECOMMENDATIONS

Semantics

The determiner-determinee method investigated under this contract yields results for resolving some classes of English equivalent ambiguity. The research procedure used pursuing this study will be fruitful in providing rules for choosing semantically appropriate English equivalents for a large list of Russian words. It is recommended that this technique be further refined and applied to a significant body of fresh text.

Syntax

The ability to rearrange translations from Russian word order into idiomatic English word order is dependent upon the correct determination of clause boundaries. Research done under the present contract in the identification of subject and object packages in conjunction with other clause boundary indicators allowed an improved rearrangement procedure. It is recommended that continued study be applied to the interrelated problems of the resolution of clause boundaries, the identification of the functional attributes of major syntactic units, and relations between clauses within the sentence.

Diagnostic Procedure

The growing complexity of the machine translation programs and the recognition of the fact that they are constantly being changed for experimental purposes have highlighted the need for diagnostic procedures. Systematic diagnostic procedures similar to those developed for computer checkout can be applied to machine translation programs. Such a procedure can help in isolating the causes of deterioration of previously trusted areas when a change is applied to a supposedly unrelated portion of the program. A computer program was developed under this contract to perform this diagnostic function.

Computer Configuration

Work in the semantic area of this contract highlighted the desirability of using a search type memory throughout the machine translation process.