RECOGNITION OF CLAUSES AND PHRASES IN MACHINE TRANSLATION OF LANGUAGES*

by

FRANZ L. ALT
and
IDA RHODES
(National Bureau of Standards Washington D.C.)

INTRODUCTION

THE process described here is intended to be used as part of the system of mechanical translation developed by Ida Rhodes of the National Bureau of Standards, as outlined e.g. in 1/: it is assumed that the reader is familiar with 1/ and in particular with the terminology used there. The present report may be considered as a more detailed exposition of one section of 1/, namely the one dealing with the establishment of a "Temporary Profile". In terms of the contemplated machine routine this is Part II-A.

The grammatical information given here is based largely on 2/, 3/ and 4/.

It was originally planned to establish the boundaries of clauses and phrases within a sentence in the course of analyzing the syntactic role of each word in the sentence (Part II-B of the contemplated routine). The idea to accomplish this part of the work in a separate part of the programme arose from a comment communicated to us by Mr. M. Sherry of the AF Cambridge Research Centre. It was enriched by numerous suggestions from Ida Rhodes, Leroy Meyers and Richard See. In particular, the author is indebted to Mr. See for the list of prepositions in Appendix II.

It is expected that the temporary profile will be established through an iterative process which scans forward and backward in alternation, perhaps many times, with relatively few commands. It seems preferable to do this separately from Part II-B because the latter is likely to have a large number of commands and fewer iterations.

The object of this part of the code is to assign to each occurrence a set of three numbers C P, b, called the clause number, phrase number, and backward signal.

Of these, C, the <u>clause number,</u> starts with 0 for each sentence and numbers clauses in the order of their <u>beginning</u> within the sentence: C=0, 1, 2,..., 7. (In hand work, often C = 1, 2, ...8). For each clause we store a <u>status symbol</u>, v, which has the following meaning: v=0, the predicate of this clause has not yet been found; v=2, the predicate has been found; v=1, a possible predicate has been found. Thus v starts with 0 in each clause anew. When a finite verb is encountered, v is set =2 for that clause. If, say, a short adjective which may also be an adverb (i.e., is in the neuter singular) is encountered while v=0, we set v=1; if subsequently a finite verb is encountered in the same clause, we set v=2 from there on. Finally, when the end of the whole sentence is reached, we check, for clauses which have not yet reached v=2 (i.e. we have found no definite predicate). If such a clause has v=1, we change it to v=2 (i.e. we accept the possible predicate as predicate); if a clause ends in v=0, there is either an error in the profile or an implied predicate. In all the foregoing we have used finite verbs and certain short adjectives as prototypes for definite predicates and possible predicates. There are, however, many others, as will be specified by the code.

The clause number, C, is =0 for the first occurrence and all subsequent ones until an indication of the beginning of a new clause (a "clause opener", or C.O.) is encountered. From there on to the next C.O., we have C=1. However, the first clause (C=0) may not yet be completed; it may be interrupted and resumed later. Therefore, at every indication of a changing C, we must decide whether to start a new clause (increase C or resume an earlier incomplete one (decrease C). In this decision we use the following

Conjecture: If clause B starts after clause A, then clause A does not resume before clause B is completed.

Stated differently, if words (or groups of words) of clauses A and B are denoted by a and b respectively, we may find the arrangement a...b...b ...a, (where B is enclosed within A), but we cannot have the arrangement a...b...a...b. (where the dots...may stand for words of A,B or other clauses).

In line with this conjecture, we propose the following

Rule: a) whenever a change in C is indicated, and in addition a clause opener follows, the new C is made higher by one than the highest C used so far in this sentence (i.e. $C_m + 1$, where C_m is the "monitoring clause number" defined below).

b) When a change in C is indicated without a new clause opener, and the most recent clause is possibly complete ($v=1$ or $v=2$), then we return to the most recent incomplete clause, i.e., we set the new C equal to the highest C previously used for which $v=0$ (i.e., equal to the quantity C_c defined below). If $v>0$ for all previous C, then we start a new clause as under (a)*.

c) When a change in C is indicated without a new clause opener, and the most recent clause is incomplete ($v=0$), then we start a new clause as under (a).

The conjecture stated above can be sharpened. The case of a clause B being "nested into" a clause A, i.e. B beginning after A begins but before A is completed, cannot occur if A is subordinate to B.** (Most likely it cannot occur when A and B are co-ordinated or unrelated, either. The only common case is that in which B, the nested clause, is subordinate to A).

In general we shall make no use of this sharper form of the conjecture since our aim is only to determine the beginning and end of each clause, not the relative roles of subordinate and super-ordinate clauses. We shall, however, need it in connection with main clauses. While the begin-ning of other clauses can usually be recognized by a clause opener (a conjunction, relative or interrogative pronoun etc.), the start of the main clause has usually no such distinctive mark. This makes it difficult to ascertain the start of the main clause in a sentence in which the main clause is preceded by one or more subordinate clauses. On the other hand, the recognition of the main clause in such a sentence is helped by the knowledge that the main clause cannot be nested into any other clause. That is to say, if the main clause is not the first clause in a sentence, then it cannot begin until all preceding clauses are completed.

This can be taken care of, without change in our general plan, by the following artifice. If a sentence begins with a subordinate clause (as indicated by a subordinate clause opener) we assign to this clause the number $C=1$, rather than $C=0$. We fictitiously assign $C=0$ to the main clause as if it started the sentence without a word. Thus, if subse-quently the end of a clause is reached according to case (b) above, and if all intervening clauses have been completed also, the new clause number chosen will be automatically $C=0$ — this playing the role of the "most recent Incomplete clause".

---

* We are thus using the existence of a predicate or possible predicate as the only criterion for completion or a clause. This may be misleading; better criteria will be found in the course of Part II, Section B of the programme where the Temporary Profile will be systematically revised.

** A possible exception is the case where B is an incidental or parenthetical clause such as "I think" or "It is well known". In such a case the bracket-ing clause may be interpreted as subordinate. It has the form of a main clause but the meaning of a subordinate one.

The phrase number, P, counts phrases within each clause. P=0 means that the current occurrence is not part of any phrase. P=1, 2, ...6 count phrases in order of their starts. The start of a phrase is indicated by the appearance of a "phrase opener", P.O., such as a preposition, participle etc. The end of a phrase is usually not determined at this stage. We assign the new phrase number to the P.O. and, in a prepositional phrase, to the next occurrence (since such a phrase has at least two words), thereafter we use the symbol P=7 (in hand work often P=x) to indicate that the occurrence belongs either to the last phrase or to no phrase at all; this is continued up to the next punctuation mark, predicate, C.O., P.O., or up to any other occurrence which cannot be part of the phrase. (For a P.O. following immediately after another P.O., P is raised a second time, so that the first P is assigned only to the P.O. itself.)

The backward indicator, b, is normally =0, and is set =1 only in certain circumstances, some of which are defined below. If b=1 it indicates that the current occurrence must be linked with an earlier one through a procedure other than the usual predictions, and this information will be utilized in Part II-B of the programme.

In addition to C, P, b for each word in a sentence, the following numbers are generated and kept up to date:

For each clause —

    the status symbol $\underline{v}$, already discussed;

    the highest phrase number used so far ("monitoring phrase number") P_m.

For the sentence as a whole —

    the highest clause number used so far ("monitoring clause number") C_m ;

    the highest number of an incomplete clause (i.e. a clause with v=0) used so far, C_c.

CLAUSES

Before establishing detailed rules for the mechanical recognition of the beginning and end of each clause and each phrase occurring in a given Russian sentence, it is advisable to obtain a bird's eye view of the kinds of clauses and phrases existing in the Russian language.

Conventional grammar distinguishes about a dozen kinds of clauses, each with its own characteristics. In enumerating these below, we do not imply that a computing machine should be programmed to draw all the distinctions among them. Such refined analysis would be quite difficult and of little use. We do think that at some future time it may be both feasible and useful for a computer to differentiate be-

tween co-ordinate and subordinate clauses and perhaps also among the three major classes of subordinate clauses. For the time being, however, our sole aim is to let the computer recognize a clause as a clause, and find its beginning and end. Our object in enumerating the types of clauses in this report is to indicate the different grammatical constructions which we may expect to encounter in Russian text, the types of clause openers, forms of the predicate, etc.

Main and Co-ordinate Clauses

The simplest periods consist of a main clause and one or more subordinate clauses. A subordinate clause may depend either on the main clause or on another subordinate clause. Furthermore, there exist co-ordinate clauses, in parallel either with the main clause or with a subordinate clause.

A co-ordinate clause in parallel with the main clause may begin with one of the conjunctions "and", "or", "nor", "but" (И, ИЛИ, НИ, А, НО), or such parallel clauses may be merely separated by a comma. A co-ordinate clause in parallel with a subordinate one begins in the same way, and in addition the clause opener of the subordinate clause may or may not be repeated. In all these respects Russian is like English, and it will suffice to give English examples:

Congress approved the bill (,) and (but) the President vetoed it.

We must hang together (,) or we shall hang separately.

United we stand, divided we fall.

All will be well if Congress passes the bill and the President
   signs it

                         or

All will be well if Congress passes the bill and if the
   President ...

The same co-ordinating conjunctions are used to join words rather than clauses:

Congress and the President favour the bill.

They may also join clauses which have the subject (or more rarely the predicate or object) in common, in which case the common part is usually not repeated:

The President signed the bill providing for ... and declared
   in his message ...

Congress passed, and the President signed, the bill providing ...
These may be considered as "elliptic clauses" in which some sentence element is missing and is understood as being repeated from another occurrence in the same period. The case of a composite subject (or

predicate) may be interpreted as an elliptic co-ordinate clause.

The case of a comma separating two co-ordinate clauses, as in the example shown above (United...) is rare in English, though more frequent in Russian. Commas are used regularly when there are more than two parallel clauses. We call these cases "listings". The last member of a listing is set off by <u>and</u> or <u>or</u> (И or ИЛИ) without a comma (in Russian):

> The committee released the bill, the Senate passed it <u>and</u>
>   the President signed it into law.

Note that in English, especially in American usage, the final <u>and</u> (<u>or</u>) in a listing is often preceded by a comma; not so in Russian.

Subordinate Clauses

With minor variations, the following grouping of subordinate clauses is accepted by many authors.
  a) Noun Clauses:
    1. Interrogative
    2. Declarative
    3. Imperative, Optative
  b) Adjectival Clauses:
    4. Relative
  c) Adverbial Clauses:
    5. Temporal
    6. Locative
    7. Conditional
    8. Causal
    9. Comparative
    10. Final
    11. Consecutive
    12. Concessive
    13. Modal

Noun clauses usually replace a noun in the nominative or accusative case, i.e. they may serve in place of a subject, predicate nominative, direct object, or apposition. They can fulfil the predictions for these sentence elements.

Adjectival clauses serve syntactically in the same role as adjective modifiers.

Adverbial clauses generally take the place of an adverb in the super-ordinate clause. "I shall return <u>soon</u> — I shall return <u>when the rain stops</u>". For the most part these clauses begin with a characteristic conjunction, by which they can be recognized. Like adverbs, they are "unpredictable" in our syntactic analysis.

1. ***Interrogative clauses*** (indirect questions) are characterised by an interrogative pronoun, interrogative adverb or other particle. He asked <u>whom</u> we would meet; ... <u>when</u> we would go; ... <u>whether</u> (if) we would return; [Double (multiple) questions:] He asked <u>whether</u> it would rain <u>or</u> snow <u>or</u> be sunny.

If, in Russian an interrogative pronoun or adverb stands at the beginning of the subordinate clause, it is immediately preceded by a comma. On the other hand, an interrogative pronoun or particle may be preceded by other words, e.g. prepositions, which are part of the same clause: They ask for <u>whom</u> the bell tolls. (When such a pronoun ("delayed clause opener") is encountered, we raise the clause number for this occurrence and for all the previous ones up to the most recent comma).

There are interrogative "adjectival pronouns" (or "pronominal adjectives") e.g. <u>what price</u> freedom? He asked <u>in which room</u> we should stay.

NB: Beware of the prejudice that interrogative clauses follow only verbs with an interrogative meaning; e.g. "I decide whom we shall see" is an interrogative clause, though no question is raised. Sometimes interrogative clauses are divided by a hairline from relative, conditional, or temporal clauses.

Syntactically, an interrogative clause frequently takes the place of a direct object. Thus, if the main clause has a transitive predicate, which predicts an accusative complement with high urgency, this prediction may be considered satisfied when an interrogative clause is encountered. An interrogative clause may, however, take the place of the subject or of a predicative nominative. In general, the interrogative clause replaces a noun in the nominative or accusative case.

To the English interrogative particle "whether" corresponds the Russian ЛИ which is placed after the first word of the clause.

2. ***Declarative clauses*** begin with ЧТО after a comma. There is therefore no ambiguity about recognizing them as clauses.

Like interrogative clauses they usually take the place of a direct object, e.g. I heard the news — I heard <u>that</u> the French government fell*. Again like interrogative clauses, they may also take the place of the subject or of a predicative nominative or apposition. "That the government fell was unexpected" (Subject). "Our hope is that the government may survive" (Predicate nominative).

There are verbs, like БОЯТЬСЯ(fear), which take a noun object in the genitive, but a clause object beginning with ЧТО like transitive verbs. In other words, these declarative clauses take the place of a complement in a case other than the accusative.

Instead of ЧТО the declarative clause may begin with ЧТОБ, ЧТОБЫ or КАК. Those beginning with КАК are close to interrogative clauses.

---

* "I heard <u>why</u> the French government fell" is an interrogative clause.

3. ***Imperative and optative clauses*** have the same form as declarative ones.

4. ***Relative or adjectival clauses*** are characterized by a relative pronoun or relative adverb. "Our father who art ..." "The place where he stands ..." Just as in interrogative clauses, if the relative pronoun or adverb stands at the beginning of the clause, it is preceded by a comma. (In this respect Russian is unlike English.) On the other hand, a relative pronoun maybe preceded by other words which are part of the same clause, e.g. a preposition: "The ground on which he stands..." (In case of such a "delayed clause opener", the clause number is treated as in the corresponding case of interrogative clauses.)

Many relative pronouns and adverbs are homonyms of interrogative pronouns and adverbs, in Russian just as in English; e.g. I wonder who came (interrog.), I saw the men who came (relative). КТО, ЧТО, КОТОРЫЙ are examples.

Relative clauses quite generally play the role of adjectives. The relative pronoun (adverb) has an antecedent in the main (or superordinate) clause, usually a noun or a demonstrative pronoun. This antecedent is modified by the relative clause, just as a noun is modified by an adjective.

5. ***Temporal clauses*** are indicated e.g. by КОГДА "when"; КАК, ЛИШЬ, КАК ТОЛЬКО, ЛИШЬ ТОЛЬКО "as soon as"; ПОКА НЕ "as long as"; ПОКУДА НЕ; ПЕРЕД ТЕМ, ЧТО "before"; ПОСЛЕ ТОГО, ЧТО "after"; ПРЕЖДЕ ТОГО, ЧТО; ДО ТОГО, ЧТО. English conjunctions: when, as, while, whenever, before, until, after, since.

6. ***Locative clauses*** begin e.g. with ГДЕ "where", "wherever", КУДА "whither" ОТКУДА "whence".*

7. ***Conditional clauses*** begin typically with ЕСЛИ "if"; also ЕСЛИ БЫ if the hypothesis is unreal; similarly ЛИШЬ, ЛИШЬ БЫ; also КОГДА БЫ or КАБЫ "if only", and others.

The predicate may be a finite verb or an infinitive.

8. ***Causal clauses*** begin with ИБО (because); in Russian this is often replaced by an antecedent-consequent pair like ПОТОМУ, ЧТО (note comma before ЧТО) and a number of similar phrases: ОТ ТОГО, ЧТО; ЗА ТО, ЧТО; ТЕМ, ЧТО; ДЛЯ ТОГО, ЧТО. Similar constructions but without causal meaning are В ТОМ, ЧТО "in that", НЕ ТО, ЧТО "not that". Intermediary between temporal and causal is РАЗ "once".

9. ***Comparative clauses*** begin with ЧЕМ "than"; also НЕЖЕЛИ, occasionally КАК. Since ЧЕМ has other meanings too, these clauses require careful handling. They are often recognizable by the previous occurrence of a comparative adjective or adverb. In other cases a ЧЕМ-clause is

---

* Note: "I shall go where you are going" — adverbial (locative) clause. "I know where you are going" — noun (interrogative) clause. "The place where you are going is near" — adjectival (relative) clause. Similar ambiguities occur with temporal clauses and others. In other words, many of the subordinating conjunctions are homonyms of interrogative and/or relative pronouns or adverbs.

followed, rather than preceded, by a comparative; in still other cases it is followed by a main (superordinate) clause beginning with ТЕМ.

However, the ЧЕМ after a comparative may not introduce a clause at all but merely a word. Also, the conjunction ЧЕМ may follow after any kind of main clause, with the meaning "rather than".

10. **Final clauses** are indicated by ЧТОБЫ, ЧТОБ "so that, in order that". The predicate is in the perfect tense or in the infinitive.

11. **Consecutive clauses** are indicated by ТАК ЧТО "such that, in such a way that". The predicate is a finitive verb or an <u>infinitive</u>.

A special difficulty lies in abbreviated infinitive constructions with consecutive meaning: ВЫ МОЛОДЫ СУДИТЬ "you are (too) young to judge".

12. **Concessive clauses** may begin with ХОТЯ or ХОТЬ, ХОТЯ И or ХОТЬ И; instead we may find НЕСМОТРЯ НА ТО ЧТО; ДАРОМ ЧТО or НУЖДЫ НЕТ ЧТО, all of which may be translated "although" or "no matter how"

13. **Modal clauses,** like the English ones beginning with КАК, КАК БУДТО, ЯКОБЫ ("аз", "as if", "as though").

<div align="center">PHRASES</div>

We consider principally three kinds of phrases: (a) prepositional; (b) adjectival, participial, gerundive, appositive; (c) incidental.

A <u>prepositional phrase</u> is unmistakably* recognizable by the fact that it starts with a preposition. The prepositions can be exhaustively listed. The phrase may or may not be enclosed by commas; in general, the longer it is, the more likely is it to have commas. As a first approximation, if a preposition is preceded by a comma, we may expect the next comma to mark the end of the prepositional phrase. There are exceptions, however. The first comma may have a different function (e.g. closing a preceding clause) and the prepositional phrase may end before the next comma:

> If he helps you today, <u>in truth</u> he must be your friend.

Or a comma may occur within the prepositional phrase:

> A true friend helps, <u>without long, detailed and searching questions,</u>
>   to the best of his ability.

The usefulness of this discussion lies more in accounting for commas than in finding the end of a phrase. Since in the commonest case — that of a phrase of three of four words not enclosed by commas — the end cannot be found with the limited information available at the time of profiling, we might as well leave undecided the end of other prepositional phrases. The rule for the profile will be that in case of doubt only the preposition and the following word are definitely in the phrase; subsequent words are left undecided up to the next clear start of a new clause. The final determination of the end of the phrase will be made in Part II-B of the machine programme, by deciding whether the subsequent words satisfy predictions made within the phrase.

---

* Note, however, that prepositions are often homonymous with adverbs.

Even this procedure may tend to close the phrase prematurely. There are prepositional phrases within prepositional phrases, and even clauses within prepositional phrases. For instance, in

    ... in the case that negotiations fail...

the subordinate (declarative) clause "that..." is part of the prepositional phrase beginning with "in".

   An <u>adjectival</u> (participial, gerundive) phrase is almost always enclosed between commas. Its principal element is an adjective (participle, gerund) which refers to a noun or nominal (antecedent) outside the phrase. The antecedent (frequently) comes before the phrase. The other elements of the phrase are dependent on the adjective (participle, gerund) and, in the subsequent Part II-B, will be predicted by it in the usual way. The phrase itself is considered unpredictable.

    The defence rests, <u>secure</u> in the expectation of victory.

    <u>Looking</u> forward to seeing you, we remain...

   If the principal element is an adjective or participle, it agrees with the antecedent in case, number and gender. To help establish this agreement in Part II-B, a "backward flag" is attached to the adjective or participle.

   Occasionally the pivotal word is an adverb rather than an adjective:

    The bill received little attention <u>relative</u> to its importance.

    The case was decided <u>similarly</u> to a precedent.

(In the first of these examples, "relative" is an adverb, though it has the form of an adjective. It modifies the adjective "little". In English it is optional, in cases like this, to use the adjectival form in place of the adverbial one, but in Russian the adverb, ОТНОСИТЕЛЬНО is used necessarily). There is no agreement of the adverb with any antecedent, and therefore no "backward flag".

   <u>Appositive</u> phrases are not recognized by phrase openers. A noun immediately following a comma is sometimes an indication of such a phrase.

   <u>Incidental</u> phrases likewise do not have phrase openers. Expressions like ПО ЭТОМУ, ОЧЕВИДНО, ТЕМ НЕ МЕНЕЕ, which frequently appear between commas, are marked "incidental" in the glossary.

CLAUSE OPENERS

   The list below contains the more frequent words by which the start of a
new clause may frequently be recognized. They are (a) the co-ordinating
conjunctions А, И, ИЛИ, НИ, НО; (b) interrogative and relative pronouns and
adjectives КОТОРЫЙ, КТО, ЧЕЙ, ЧТО, КАКОЙ, КАКОВОЙ, and their inflected
forms; (c) a large number of subordinating conjunctions. Some of the
latter are also used as interrogative or relative adverbs, e.g.ГДЕ.
   Most of these stand at the beginning of the clause, preceded by a comma,
but some can be delayed, notably the pronouns and the conjunction ЛИ.
   To some English conjunctions correspond composite expressions in Russian
which maybe considered as idioms, e.g. ОТ ТОГО, ЧТО..., literally "from
this, that...", could be treated as an idiom and translated by the single
word "because". Many more such cases occur, and it is somewhat arbitrary
how many are to be listed as idioms representing conjunctions, and how many
are to translated literally. In this connection we mention the divided
idioms, such as ТЕМ...,ЧЕМ, which we may treat as "antecedent-consequent"
pairs.
   Some of the conjunctions are also used as adverbs, so that they cannot
be taken as sure signs of the beginning of a clause: e.g. РАЗ= once (=as
soon as) or (=at one time). Also, the interrogative pronouns can occur in
main clauses, instead of as openers of subordinate clauses. In many cases
the occurrence of a comma immediately before one of these doubtful clause
openers indicates that a new clause is indeed being opened.

    А   КАКОВ НИ...БЫ
    В ТОМ, ЧТО   КАКОВОЙ   (Possibly delayed)
    ГДЕ   КАКОЙ   ("   "   )
    ДАРОМ, ЧТО   КОГДА;   КОГДА...БЫ
    ДЛЯ ТОГО, ЧТОБ   КОТОРЫЙ and inflected forms
    ДО ТОГО, КАК (or ЧТОБ)   (possibly delayed)
    ЕСЛИ; ЕСЛИ...БЫ; ЕСЛИБ   КТО and inflected forms
    ЗА ТО, ЧТО   (possibly delayed)
    И   КУДА
    ИБО   ЛИ (always delayed)
    ИЛИ   ЛИШЬ; ЛИШЬ...БЫ
    КАБЫ   ЛИШЬ ТОЛЬКО
    КАК   НЕ ТО, ЧТО
    КАК БУДТО   НЕЖЕЛИ
    КАК ТОЛЬКО
    КАКОВ

НЕСМОТРЯ НА ТО, ЧТО	С ТЕХ ПОР, КАК
НИ	СКОЛЬКО and inflected forms
НО	ТАК ЧТО
НУЖДЫ НЕТ, ЧТОБЫ	ТЕМ, ЧТО
ОТ ТОГО, ЧТО	ХОТЬ; ХОТЬ И; ХОТЯ; ХОТЯ И
ОТКУДА	ЧЕЙ and inflected forms
ПЕРЕД ТЕМ, КАК (ЧТОБЫ)	(possibly delayed)
ПОКА; ПОКА ... НЕ	ЧЕМ *(after comparative)*
ПОКУДА ... НЕ	ЧТО (interrogative or relative
ПОСКОЛЬКУ	pronoun, with inflected forms;
ПОСЛЕ ТОГО, КАК (ЧТО)	possibly delayed)
ПОТОМУ, ЧТО	ЧТО (conjunction - not declined)
ПОЧЕМУ	ЧТОБ, ЧТОБЫ
ПРЕЖДЕ ЧЕМ; ПРЕЖДЕ ТОГО,	ЯКОБЫ
ЧТОБЫ	
РАЗ	

APPENDIX II •

RUSSIAN PREPOSITIONS

БЕЗ	+ВСЛЕД	+НАКАНУНЕ
БЕЗО	ВСЛЕДСТВИЕ	+НАПЕРЕКОР
БЛИЗ	ДЛЯ	НАПОДОБИЕ
В	ДО	+НАПРОТИВ
+ВБЛИЗИ	ЗА	НАСЧЕТ
ВВИДУ	ИЗ	+НЕВПРИМЕР
+ВДОЛЬ	ИЗО	О
ВЗАМЕН	ИЗ-ЗА	ОБ
ВМЕСТО	ИЗ-ПОД	ОБО
ВНЕ	К	+ОКОЛО
+ВНИЗУ	КО	ОТ
+ВНУТРИ	КРОМЕ	ОТО
+ВНУТРЬ	о+КРУГОМ(КРУГ)	о ПЕРЕД (ПЕРЁД)
ВО	о МЕЖ(МЕЖА)	ПЕРЕ ДО
ВОЗЛЕ	МЕЖДУ	ПО
+ВОКРУГ	+МИМО	ПОВЕРХ
ВОПРЕКИ	НА	ПОД
+ВПЕРЕД И	+НАВСТРЕЧУ	+ПОДЛЕ
+ВРОДЕ	НАД	ПОДО
ВСИЛУ	о НАДО	ПО-ЗА

Note: "+" indicates that the preposition may also serve as an adverb, "o" indicates
    that the preposition is homographlc with forms other than adverbs. The diction-
    ary form of the homograph is given in parentheses when it is different from
    the preposition. The prepositions consisting of a single letter may be homo-
    graphic with letters of the Cyrillic alphabet, or even with letters of the
    Latin alphabet.

+ПОЗАДИ	ПРЕД	СВЕРХ
ПОЗАДЬ	ПРЕДО	+СВЫШЕ
ПОМИМО	+ПРЕЖДЁ	+СЗАДИ
ПО-НАД	ПРИ	СКВОЗЬ
+ПОПЕРЕК	ПРО	СО
ПОСЕРЕДИНЕ	ПРОТИВ	СПЕРЕДИ
+ПОСЛЕ	о+ПУТЕМ (ПУТЬ)	СРЕДИ
+ПОСРЕДИ	РАДИ	У
ПОСРЕДИНЕ	С	ЧЕРЕЗ
ПОСРЕДСТВОМ		

APPENDIX III

CHARACTERISTICS OF OCCURRENCES TO BE USED IN
PROFILING ("PROFILE SKELETON")

1. *Clause openers*

   co-ordinate
   subordinate
   accepting infinitive as predicate
   elliptic (subject and/or predicate of clause may be omitted)
   delayed

2. *Phrase openers*

   prepositions
   participles (long form), after comma
   gerunds
   adjectives (long) and adverbs which can be heads of adjectival or
      adverbial phrases — mainly those with case government, after
      comma
   nouns after comma

3. *Status-affecting occurrences* (bearing on choice of predicate for
      each clause)

   finitive verb
   copulative verb
   infinitive verb
   short adjectives and participles
   other predicates (МОЖНО,ВОТ,ВОН,ТАМ,ТУТ,etc.)
   words predicting an infinitive (these are noted because they prevent
      a subsequent infinitive from being Interpreted as predicate)
   dash

4. *Punctuation marks*

5. *Miscellaneous*

   antecedent-consequent pairs
   incidentals (if between commas, prevent commas from being interpreted
      as ending a clause)
   comparative adjectives and adverbs (antecedents of ЧЕМ)

<div align="center">

APPENDIX IV
EXAMPLE
</div>

   The following table shows the construction of a temporary profile. The
first column lists, in order, the words of a Russian sentence, taken from
С. Н. БЕРНЩТЕЙН, СОБРАНИЕ СОЧИНЕНИЙ, ТОМ 1 (АКАД. НАУК
СССР 1952),page 214, lines 6-9. The second column, labelled "profile
Skeleton" gives that part of the information extracted from our "temporary
choices" which is pertinent to the preparation of the temporary profile.
Based on this information, the machine assigns the numbers in the columns
headed C, v, P. The footnotes to the table explain some of the rationale
behind the rules which the machine is following. The symbol ϕ in the
Profile Skeleton indicates "no information", i.e., the grammatical Inter-
pretations ("temporary choices") of the original occurrence do not contri-
bute any information to the formation of the temporary profile.

q.	Occurrence	Profile Skeleton	C	v	P	b
1	НО	Coord. Conj.	0	0	0	0
2	ВМЕСТО	Prep.	0		1	
3	ТОГО	ϕ	0	0	1	
4	,	Cma				
5	ЧТОБЫ	C.O. [inf.pred.]	1	0	0	
6	ИСКАТЬ	inf.	1	1/2	0	
7	ТОЧНОЕ	ϕ	1		0	
8	АЛГЕБРАИЧЕСКОЕ	ϕ	1		0	
9	ВЫРАЖЕНИЕ	ϕ	1		0	
10	ДЛЯ	Prep.	1		2	
11	НАИЛУЧШЕГО	ϕ	1		2	
12	ПРИБЛИЖЕНИЯ	ϕ	1		x	
13	ПРИ	Prep.	1		3	
14	ПОМОЩИ	ϕ	1		3	
15	МНОГОЧЛЕНОВ	ϕ	1		x	
16	ДАННОЙ	partic.	1		x	
17	СТЕПЕНИ	ϕ	1	1/2	x	

q.	Occurrence	Profile Skeleton	C	v	P	b
18	,	Cma				
19	ЧТО	C.O.	2	0	0	
20	ВООБШЕ	ф	2	0	0	
21	,	Cma				
22	КАК	C.O.	3	0	0	
23	МЫ	ф	3	0	0	
24	ВИДЕЛИ	Pred.	3	2	0	
25	,	Cma				
26	НЕ	ф	2	0	0	
27	ОСУЩЕСТВИМО	partic. short neut. sing.	2	1/2	0	
28	,	Cma				
29	Я	ф	0	0	0	
30	ИЩУ	Pred.	0	2	0	
31	ВЫРАЖЕНИЕ	ф	0	0		
32	,	Cma				
33	ЯВЛЯЮЩЕЕСЯ	Partic.	0		4	1
34	ВПОЛНЕ	ф	0		x	
35	ТОЧНЫМ	ф	0		x	
36	ТОЛЬКО	ф	0		x	
37	ДЛЯ	prep.	0		5	
38	БЕСКОНЕЧНЫХ	ф	0		5	
39	СТЕПЕНЕЙ	ф	0	2	x	
40	.	.				

Notes pertaining to each occurrence   q:

   1. The first word always starts C=0, unless it is a subordinating
conjunction. In this way the first clause (which is usually the main
clause) is numbered 0. Since this word is not a predicate, the status
of Clause 0 is v=0. Since the word is not a phrase opener, we set P=0,
Indicating that this word is not part of any phrase.

   2. A preposition. We are still in clause C=0, and have not encountered
its predicate (i.e. no change in v). This word opens a (prepositional)
phrase, therefore P=l.

   3. No Information in the profile skeleton. The word following a prepo-
sition must be part of the same prepositional phrase (unless it is itself
a phrase opener), therefore still P=l.

   4. A comma, so marked in the profile skeleton. This is taken to indi-
cate the end of phrase 1. Furthermore, since phrase 1 was not preceded by
a comma, the present comma must have still another function, e.g. mark the
beginning of a new clause or phrase, or separate the members of a listing.

Which of these functions is present will be indicated by subsequent occurrences.

5. Clause opener, therefore raise C. The skeleton profile also indicates that the conjunction ЧТОБЫ sometimes takes a predicate in the infinitive.

6. Verb in the infinitive. Possibly a predicate of this clause, therefore set v=1. See also Note 40.

7 - 9. No changes in profile.

10. Preposition. Start phrase 2.

11. The word following a preposition is normally part of the same phrase, therefore P=2.

12. There is no way of determining, with the information of the profile skeleton alone, whether this word still belongs to P=2. We therefore set P=x. (The machine programme will resolve this uncertainty in Part II-B).

13. Preposition. Start phrase 3.

14. See 11.

15 - 17. See 12. Note that the participle in 16 is not preceded by a comma, therefore not considered a phrase opener.

18. See 4.

19. Clause opener starts Clause 2. (This is an oversimplified example. It omits in the first iteration the complication arising from the fact that the word ЧТО may have other functions apart from that of a conjunction. Subsequent iterations may resolve such initial ambiguities).

20. No change in the profile indicated.

21. See 4.

22. Clause opener, starts Clause 3. As far as we know, clauses 0 and 2 are still Incomplete; yet the word КАК calls for starting a new clause.

23. See 20.

24. A finite verb, definitely a predicate. Set v=2 for clause 3.

25. See 4.

26. No clause opener. Since the comma may indicate a change in C and the most recent clause is incomplete, we set C equal to the most recent number of an incomplete clause, i.e. C=2 (since v=0 for clause 2). This may be changed if a delayed clause opener, such as a relative pronoun, is encountered reasonably soon.

27. A short-form participle in the neuter singular. This may possibly be a predicate. Set v=1.

28. See 4.

29. See 28. Clause 2 is considered complete since its v=1; so is clause 1; therefore return to C=0.

30. See 24.

31. See 20.

32. See 4.

33. The participle following a comma indicates a participial phrase. This is sufficient to explain the preceding comma, and therefore the comma is not considered as ending the previous clause. We remain in clause 0 and start phrase 4. For use in Part II-B, we place a backward flag, to seek explanation in a foregoing occurrence.

34 - 36. See 12. Part II-B will determine how many occurrences following 33 belong to phrase 4.

37. Start prepositional phrase. Although in reality this is a phrase within a phrase, namely, part of the participial phrase 4, our programme does not take note of this fact.

38. See 11.

39. See 12.

40. Period, end of sentence*. Check whether all clauses have found their predicates. Clauses 1 and 2 had found "possible predicates" marked v=1; these are now changed to v=2, since no inconsistency has arisen. If there were any unsolved difficulties at this point, such as clauses without predicates or with more than one predicate, we would try to resolve them by iterating the entire profiling process.

REFERENCES

1. RHODES Ida, A new approach to the mechanical syntactic analysis of Russian, *MT,* 1960,**VI**,2,
2. FORBES, N. Russian Grammar. 2d ed., Oxford Clarendon, 1945.
3. MAZON, A. Grammaire de la langue Russe. Paris, Inst.d'Etudes Slaves 1949.
4. HOUSE, H.C. and HARMAN, S.E., Descriptive English Grammar. New York, Prentice-Hall, 1950. _____

* A separate subroutine will determine whether a period does actually indicate the end of the sentence. Moreover, if some other symbol (? or !) is used for the purpose, an appropriate signal will be stored for use in Part II-B.