

Session 6: SYNTAX

NESTING WITHIN THE PREPOSITIONAL STRUCTURE

Michael Zarechnak  
Georgetown University

This paper presents a preliminary description of an algorithmic operation to handle nested strings within the prepositional structure in Russian.

The prepositional structure is defined as consisting of a preposition with its case-determining requirement at the permitted point of entry, and of a noun or its substitute at the permitted point of exit which satisfies the requirement incurred by the preposition.

Example:     на     мосту     "on the bridge"

The flow chart shown in Appendix II illustrates in detail the variables and relations, as well as the expected results involved in the computable syntactic analysis of the prepositional structure containing one or more nested strings. The nested strings consist of either weak or strong government structure. The former usually precedes the latter. The type of nested string can be another prepositional structure, a pronominal structure, or a noun structure. The order of the nested strings is fixed at one point. A pronominal case determiner follows the first preposition. A second prepositional case determiner can only occur after the first pronominal case determiner.

This algorithmic operation has been worked out on the basis of the data extracted from 50, 000 words of continuous text in the field of organic chemistry. It has then been generalized to the degree permissible by the extracted structural cues.

The reason for the necessity of developing a special algorithm is rather simple: The analysis of the machine output disclosed failures of the existing operator covering the prepositional structures occurring in the chemical text. Here is a typical example illustrating the failure:

Position	$\frac{P_0}{P_{cd_2}}$	$\frac{P_1}{A_{cdw_5}^2}$	$\frac{P_2}{P_{cd_5}}$	$\frac{P_3}{N_{c_5}}$	$\frac{P_4}{N_{c_2}}$
----------	------------------------	---------------------------	------------------------	-----------------------	-----------------------

Session 6: SYNTAX

The codes<sup>1</sup> unloaded under the words of this prepositional structure have been produced by the computer as an input for further machine analysis:

			1122	1122
	3112			3112
	3520	3520		
5122	5122	5125	5125	<u>5125</u>

Following are the codes as they should unload:

			1122	1122
	3112			3112
	3520	3520		
5122	5122	5125	5125	<u>5122</u>

From the above example it is clear that the first prepositional requirement was not satisfied by the last noun. The technical reason for improper unloading was the fact that code 1122 was unloaded at position  $P_3$ . Whenever the latter code occurs, the prepositional operator automatically unloads the given codes as long as code 1122 continues. The structural cause for this failure was the assumption that immediate-constituency analysis would overcome the difficulties encountered within a nested structure. We tried to expand the immediate-constituency analysis by considering the transition between a pronominal weak governor and a prepositional strong governor as an example of an immediate-constituency pair. This concept is reflected in unloading code 3520 under both items at positions  $P_1$  and  $P_2$ . The immediate-constituency analysis of positions  $P_3$  and  $P_4$  failed because under the prevailing circumstances there is no semantic reason<sup>2</sup> for unloading code 1122. There is no immediate-constituency relation between positions  $P_0$  and  $P_4$ , since there is no known substitution test which can bring them together on the immediate-constituency level.

This structural situation necessitated the devising of the present algorithm for handling nested strings.

The concept of a nested structure is similar to that of the branching of a tree or a graduated series of boxes.

---

<sup>1</sup> The codes have been explained in the report on Current Research at Georgetown University given at this Symposium in Session 2.

<sup>2</sup> See examples 3 and 22 in Appendix I.

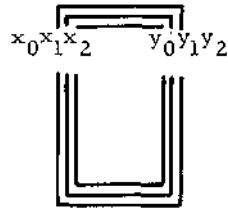


Figure 1

As seen from Figure 1 ,  $x_0$  is satisfied by the  $y$  with the highest subscript;  $x_1$  is satisfied by the next-to-the-highest  $y$ , etc. ; eventually,  $x$  with the highest subscript is satisfied by the  $y$  of the lowest subscript.

The entire GAT program consists of approximately 40 complex operators. It is assumed that the operations prior to the prepositional operator have assigned the necessary codes to distinguish the word classes. Further, they have assigned the codes for the combinations of word-class sequences on the level of apposition and agreement structures.

It should be noted that the prepositional structural analysis is only a small portion of the GAT syntagmatic analysis, which in turn is a stepping-stone toward syntactic analysis. This means that the prepositional structure is treated against the basic concept of computable language analysis as applied to the structure of a sentence. The prepositional structure itself is one of the permissible word-class combinations in the Russian sentence model.

The word classes that can open the position to be occupied by a prepositional structure are the following: noun, verb, adjectival, adverb, and conjunction. These word classes may occur either before or after the prepositional structure. (See examples in Appendix IV. )

The linguistic description of the algorithm is briefly summarized in Appendix III.

The algorithm has another application besides that of the proper unloading of the nested structure. Later it serves as an input for the transfer from Russian to English and for the rearrangement operator. In the list of data given in Appendix I, both the Russian and English equivalents are given first as they actually occur, and subsequently, followed by the letter "a", in sequences which are expected to be

Session 6: SYNTAX

produced by the computer. The expected sequences have been produced manually, but with a simple transformation they can be computed. In this transformation the operators are set in such a way that the values occurring at positions  $P_0$  through  $P_N$ , are eliminated one by one to see what effect on the prepositional structure such an elimination would have. As soon as operator  $N$  produces zeros at the transform level between positions  $P_0$  and  $P_N$ , we have all the information on which the rearrangement should be based, since  $P_0$  will be immediately followed by  $P_N$  and the remaining words will be automatically shifted in their usual sequence after  $P_N$ . On the basis of the nested structures encountered by us so far, it would appear that only one rearrangement shift is applicable.

Example:  $P_0$        $P_1$                    $P_2$      $P_3$                    $P_4$      $P_5$   
                   через   нагретую   до   150   градусов   трубку   .  
 Let:       $P_0 = a$        $P_2 = c$                    $P_4 = e$   
              $P_1 = b$        $P_3 = d$                    $P_5 = f$

Then apply the transformation as follows:

	a	b	c	d	e	f
$R_1$	∅	∅	∅	∅	∅	∅
$R_2$	a	∅	∅	∅	∅	f
$R_3$	a	f	b	c	d	e

During the next three months the above-mentioned routines will be tested on the computer against 200, 000 words of continuous Russian text. If the testing yields positive results, the operation will be further expanded to handle nested strings occurring within other types of government structures.

Session 6: SYNTAX

APPENDIX I

- 1 в окружающей пластинку среде  
1a в среде, окружающей пластинку
- 1 in the surrounding plate medium  
1a in the medium surrounding the plate
- 2 через нагретую до 160 градусов трубку  
2a через трубку, нагретую до 160 градусов
- 2 through a heated up to 160 degrees tube  
2a through a tube heated up to 160 degrees
- 3 на протекающей с большим выделением тепла реакции  
3a на реакции, протекающей с большим выделением тепла
- 3 upon a proceeding with a great isolation of warmth reaction  
3a upon a reaction proceeding with isolation of warmth
- 4 с неизменным в зависимости от природы катиона содержанием кристаллизационной воды  
4a Ø
- 4 with an indispensable depending on the nature of the cation content of crystallization water  
4a Ø
- 5 кроме приведенных в основном тексте типов  
5a кроме типов, приведенных в основном тексте
- 5 besides the given in the basic text types  
5a besides the types given in the basic text
- 6 над нагретым до 500 градусов катализатором  
6a над катализатором, нагретым до 500 градусов
- 6 over a heated up to 500 degrees catalizer  
6a over a catalizer heated up to 500 degrees
- 7 над нагретым до 140 градусов углем  
7a над углем, нагретым до 140 градусов
- 7 over heated up to 140 degrees coal  
7a over coal heated up to 140 degrees
- 9 от взвешенных в них твердых и жидких частиц  
9a от твердых и жидких частиц, взвешенных в них
- 9 because of suspended in their solid and liquid particles  
9a because of solid and liquid particles suspended in them

Session 6: SYNTAX

Appendix I (Continued)

- 10 от насыщенного на холоду раствора  
10a от раствора, насыщенного на холоду
- 10 from a saturated in the cold solution  
10a from a solution saturated in the cold
- 11 из лежащих в основе этих комплексов простых солей  
11a из простых солей, лежащих в основе этих комплексов
- 11 from the forming the base of these complexes simple salts  
11a from simple salts forming the base of these complexes
- 12 в благоприятных для реакции температурных условиях  
12a в температурных условиях, благоприятных для реакции
- 12 in favorable for the reaction temperature conditions  
12a in temperature conditions favorable for the reaction
- 13 в схематически показанной на рисунке ио 6 печи  
13a в печи, схематически показанной на рисунке ио 6
- 13 in the schematically shown in Diagram #6 furnace  
13a in the furnace shown schematically in Diagram #6
- 14 от падающей в ночь росы  
14a от росы, падающей в ночь
- 14 because of falling at night dew  
14a because of dew falling at night
- 15 на передававшейся от отца к сыну рецептуре  
15a на рецептуре, передававшейся от отца к сыну
- 15 according to passed down from father to son prescription  
15a according to a prescription passed down from father to son
- 16 от вводимых в реакцию количеств веществ  
16a от количеств веществ, вводимых в реакцию
- 16 from introduced into reaction quantities of substances  
16a from quantities of substance introduced into reaction
- 17 из описанных им методов  
17a из методов, описанных им
- 17 from described by him methods  
17a from methods described by him

Session 6: SYNTAX

Appendix I (Continued)

18	от 20 до 30 тысяч герц
18a	Ø
18	from twenty to thirty thousand cycles per second
18a	Ø
19	в подвергаемой действию звука среде
19a	в среде, подвергаемой действию звука
19	in an exposed to the action of sound medium
19a	in a medium exposed to the action of sound
20	из замешанного на жидком стекле цемента
20a	из цемента, замешанного на жидком стекле
20	from a mixed on a liquid glass base cement
20a	from cement mixed on a liquid glass base
21	со способными выделять газы веществами
21a	с веществами, способными выделять газы
21	with capable of isolating gases substances
21a	with substances capable of isolating gases
22	из получаемых на специальных машинах очень тонких стекляных нитей
22a	из очень тонких стекляных нитей, получаемых на специальных машинах
22	out of manufactured by special machines fine glass fiber
22a	out of fine glass fiber manufactured by special machines
23	из аналогичных непредельным углеводородам ненасыщен- ных силанов
23a	из ненасыщенных силанов, аналогичных непредельным угле- водородам
23	out of analogical to saturated hydrocarbons unsaturated silances
23a	out of unsaturated silances analogical to saturated hydrocarbons
24	по второму и отчасти по третьему методам
24a	Ø
24	according to the second and partially to the third method
24a	Ø

APPENDIX III

The Nesting Program

The program can be divided into four sections. The nested stretch is determined and extracted by another routine and is sent to the Nesting Program for computation.

Section I (000 series in the flow chart)

After setting the counters and other necessary housekeeping and initialization operations, the following steps are taken:

1. Inquire whether the first item in the stretch is a preposition with a case determiner. If "no", there is an error. Check for the error and restart the operation. If "yes", check whether the case determiner of the preposition is ambiguous, and if so go to the special routine for the resolution of the ambiguity. Then proceed to Section II below.

Section II (100 series in the flow chart)

1. Establish the particular case of the case determiner of the preposition.
2. Remember the case of the first preposition in one counter, and remember the cases of the following prepositions in another counter. Each time a preposition is satisfied by unloading, erase the counters.
3. Check the counters for previously stored weak government, etc., and unload on the preposition or prepositions with matching cases.

Move to the next word and proceed to Section III below.

Section III (200 series in the flow chart)

1. Check whether the following word is an adjective, adverb, conjunction, particle, or an infinitive verb. If none of these occurs in the first scanning, there is an error which is checked and the subroutine is tested again. During the subsequent scannings, however, if none of the above occurs, a further check is made to determine whether the item is a noun or another preposition.

If an adjective was found in Step 1 in this section, a further check is made to determine whether it has a strong case determiner code or a weak one. In either case it is recorded in special separate counters.



## Session 6: SYNTAX

3. If an infinitive verb was found, its determining factor is remembered in another special counter.

4. In all the above cases the routine moves to the next word (except under special circumstances where the modification is prevented by Switch No. 3) and proceeds to Section IV below.

### Section IV (300 series in the flow chart)

1. Check whether the item under consideration is a noun. If "no", set the switches for proper execution and go to the beginning of the routine (Point II in the flow chart) to check on preposition, adjective, etc. If "yes", go to Step 2 below.

2. Check whether the noun has a case determiner. If "yes", remember it in a special counter, and in either case go to Step 3.

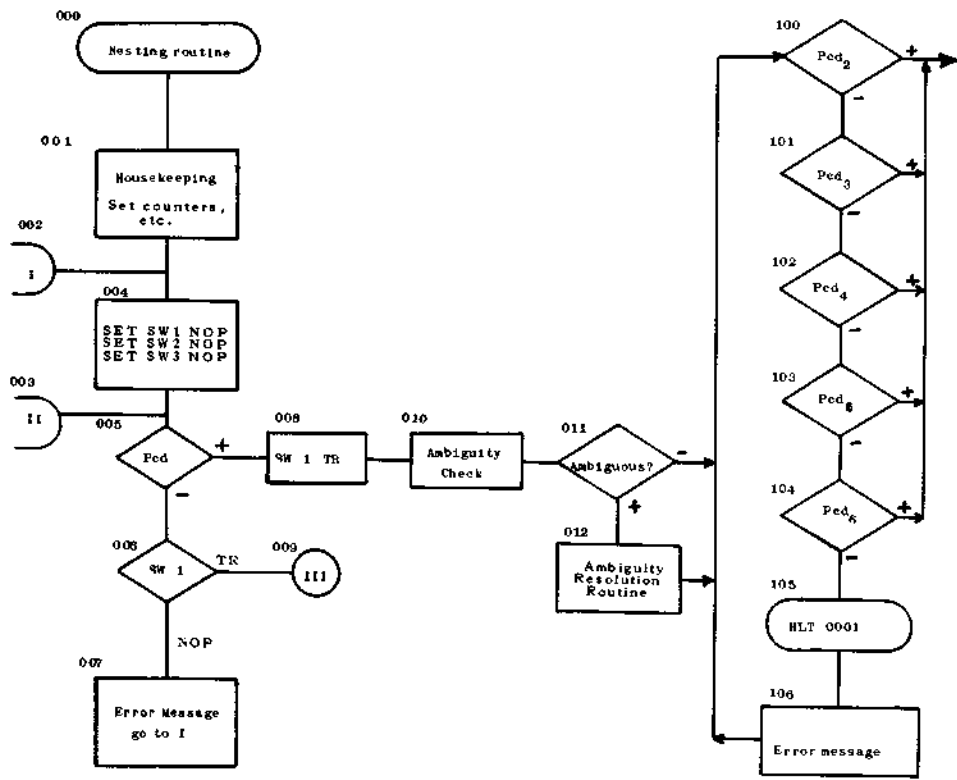
3. Does the case of this noun correspond with that of the first preposition in the stretch? If "yes", unload the code 512x ("x" being the proper case found) into the appropriate preposition and noun. Check whether there are any agreement or government-extension codes around the noun and unload the 512x code on any of such extensions found. Erase counters and go to the start to check on the next prepositional stretch (Point I in the flow chart). If the result of the first inquiry above was negative, go to Step 4.

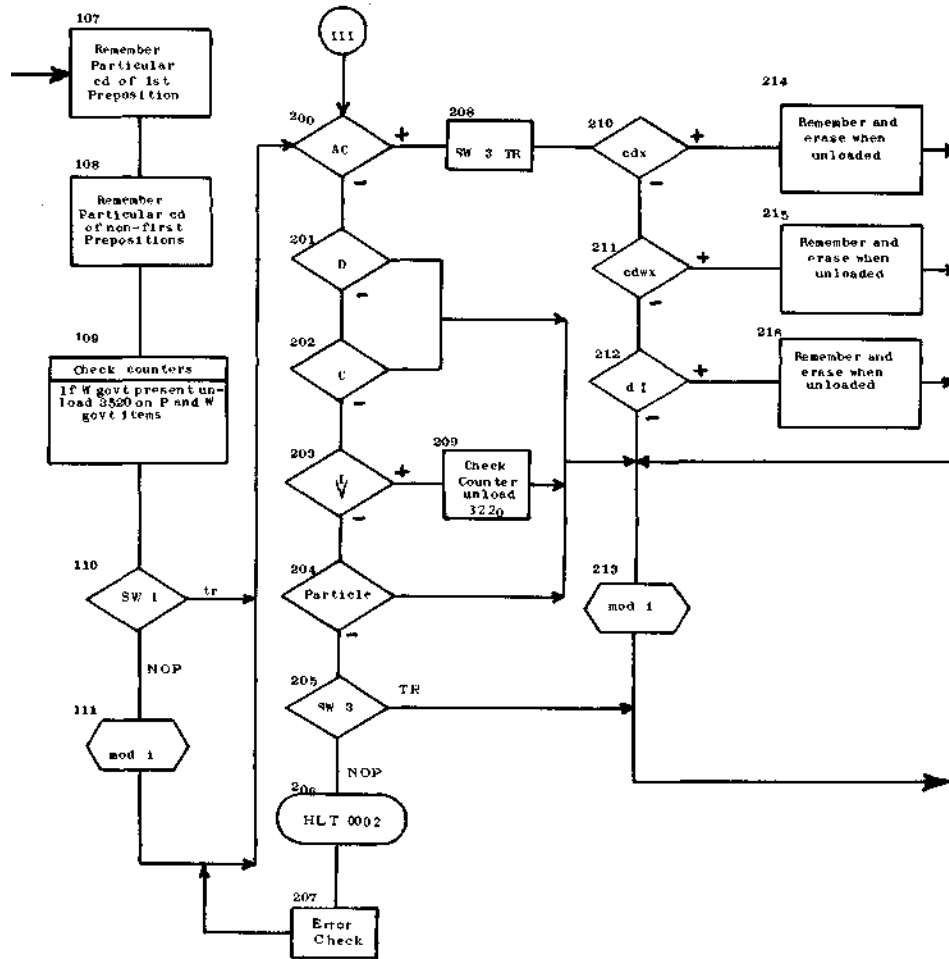
4. If an adjective with case determiner has been found and the case of the noun corresponds with the case determiner of the adjective, unload 312x on the noun and the adjective, check for extension codes and unload 312x on the extension--if any, move to the next word, and go to the beginning (Point II in the flow chart) to check the next word. If no adjective with case determiner has been found, go to Step 5.

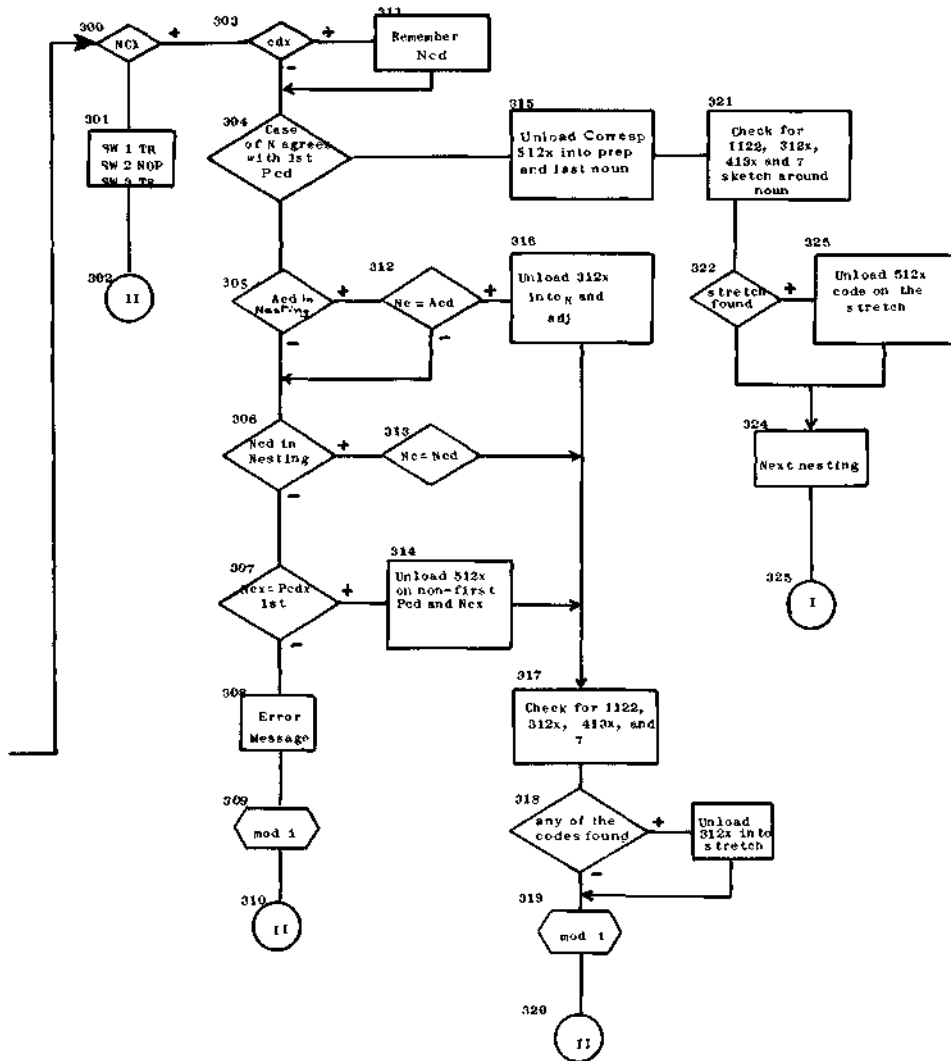
5. If a noun with case determiner has been found and the case of the noun agrees with the case determiner of the previous noun, unload and check for any extensions as in Step 4 above. If the result of this inquiry is negative, go to Step 6.

6. If the case of the noun agrees with the case determiner of the preposition (not the first one) or the prepositions, unload 512x on the appropriate noun and preposition and check for extensions as in Step 4 above. If the result of the first inquiry in this step is negative, go to Step 7.

7. If there is an error, write the message, and go to the beginning (Point II in the flow chart) to check on the next word.







APPENDIX IV

Examples:

Noun + PN

условия для возникновения	"conditions for the arising of"
развитие на земле растительности	"the development of vegetation on Earth"

Verb + PN

имея в виду	"having in mind"
накапливались в океане	"accumulated in the ocean"
к этой группе относятся	"to this group belong"

Adverb + PN

он умер далеко от друзей	"he died away from his friends"
только до известного предела	"only to a certain limit"

Adjective + PN

известных на земле.. .запасов	"reserves. . .known on Earth"
похожие по свойствам	"similar in properties"

Pronoun + PN

потребление их в виде топлива	"consumption of them in the form of fuel"
содержание его в воздухе	"content of it in the air"

Conj + PN

отметить, что в ряде случаев	"to note that in a number of cases"
если в алифатической цепочке...	"if in the aliphatic chain. . ."