

Studies in Machine Translation—8: Manual for Postediting Russian Text *

by H. P. Edmundson†, K. E. Harper, D. G. Hays, and B. J. Scott
Mathematics Division, The RAND Corporation

The present study is a practical guide to editors who refine partially machine-translated text as a basis for linguistic analysis. The post-editors' tasks are: to code preferred English equivalents, to code English structural symbols, to resolve grammatic properties, and to code syntactic connections (dependencies). A general introduction to the field of machine translation is contained in The RAND Corporation RM-2060.

1. Introduction

1.1 GENERAL

The present paper is one in a series which describes the methods now in use for research on machine translation (MT) at The RAND Corporation. Postediting follows mechanical partial translation in the research process; the editor encodes changes to yield an accurate, readable English text, and encodes the structure of each sentence in preparation for linguistic analysis. The present manual is based on studies of Russian physics and mathematics, but is presumably applicable to other textual materials, within the framework of the RAND methodology.

1.2 WORKSHEET FORMAT

The posteditor works from a text listing prepared on an IBM printer; a sample list is shown in Table 2. Each occurrence in the Russian text occupies one line of the listing; the following items of information are given for each occurrence:

- Sequence number (S), consisting of:
 - Page number (PG)
 - Line number (L)
- Occurrence number (O)
- Punctuation before the occurrence (P 1)
- Russian form of the occurrence (may be transliterated)
- Punctuation after the occurrence (P 2)
- Russian inflectional grammar code (G)
- Sentence-sequence number (S'), consisting of:
 - Sentence number (SN)
 - Occurrence number in the sentence (ON)
- Coding space, for insertion of:
 - Dependency code (DC)
 - English structural symbols (ESS)
 - Preferred English equivalent (PE)
 - Translation order (TO)

* The research herein reported was performed with the support of USAF Project RAND.

† Presently at Ramo-Wooldridge, a Division of Thompson Ramo Wooldridge, Inc. H. P. Edmundson has co-authored all revisions of this manual up to but not including the present revision.

English equivalent (E 1, E 4)
Word number
Special codes
Alternative English equivalents (E 2, E 5, E 3, E 6)

TABLE I
THE RAND PUNCTUATION CODE

Symbol Printed Before an Occurrence	Punctuation Mark
1	Start paragraph
2	Start sentence
3	Capital
/	Open parenthesis
8	Open quotation
Symbol After an Occurrence	Printed Punctuation Mark
.	Question mark
!	Exclamation point
-	Hyphen
.	Period
,	Comma
--	Dash ^a
:	Colon
;	Semicolon
/	Close parenthesis
8	Close quotation

^a This mark must sometimes be interpreted as an arithmetic symbol (minus) or as the verb *be*.

The RAND printing of punctuation adheres to generally accepted standards, within the limits set by the number of characters on an IBM printer. Original punctuation marks appear only after the Russian occurrence and are not repeated after English translation.

The Russian grammar code is fully described in The RAND Corporation MT Study 6; the posteditor should be familiar with it.

The first English equivalent (E 1) is generally an accurate translation of the Russian form; the rough

TABLE 2
POSTEDITOR'S WORKSHEET

IDENT.		SOURCE LANGUAGE				ANALYSIS						TARGET LANGUAGE						REMARKS	
NO.	PI	RUSSIAN FORM	P2	G	IN	ON	DC	ISO	PE	TO	E1	E4	W	SP	E2	E5	E3		E6
A2G01	1	ПОЛАГАЯ		61.1A	00101		12			C	ASSUMING		3869.		A+				
A2G02		7		C7	00102		1				(EQUATION)		6409.						
A2G03		СМ		21X J	00103		2				THE CM		9053.		- 21X J 6NY				
A2G04		7		C7	00104		3				(EQUATION)		6409.						
A2G05		ЧТО		32A1	00105		4			3	WHICH		9073.		943 B2A1			06 04	
A2G06		ОТВЕЧАЕТ		6.2	00106		5			A	CORRESPONDS		3427.		A. ANSWER			00 05	
A2G07		БЛИЖАЙШИМ		3T K5	00107		6			5	NEAREST		301.		B				
A2G08		ЗВЕЗДАМ		237 5	00108		7			K 3	STARS		1488.						
A2H01		ДЛЯ		81A11	00109		8				FOR		1201.						
A2H02		ОБЪЕКТА		210 3	00110		9			2	THE OBJECT		3196.		OBJECTIVE				
A2H03	/	54		C54	00111		10				(REFERENCE)		6409.						
A2H04		ПОЛУЧАЕМ		6 C5A 6H C5A	00112		11			4	WE OBTAIN		3927.		A+ ASCERTAIN				
A2H05		В		711	00113		12				AS		402.						
A2H06		КАЧЕСТВЕ		244	00114		13						1834.						
A2H07		СРЕДНЕЙ		148 W AD W	00115		14			+	AVERAGE		5210.						
A2H08		ЭФФЕКТИВНУЮ		325 8	00116		15				THE EFFECTIVE		6372.						
A2H09		ТЕМПЕРАТУРУ		23208	00117		16			-	TEMPERATURE		5430.						
A2I01		7		C7	00118		17				(EQUATION)		6409.						
A2I02		ГРАДУСОВ		216	00119		18			3	OF DEGREES		969.						
A2I03		В		930	00120		19				WHILE		402.						
A2I04		ТО			00121		20						9065.						
A2I05		ВРЕМЯ		24+ 4	00122		21						650.		+				
A2I06		КАК			00123		22						9016.						
A2I07		ДЛЯ		81A11	00124		23				FOR		1201.						
A2I08	3	СОЛНЦА		240	00125		24				THE SUN		5094.						
A2I09		НА		8E315	00126		25			3	(AT) ON -		2636.						
A2I10		ТОЖЕ		3D 6	00127		26				THE SAME		5391.		9				
A2I11		НЕ		0E	00128		27						1314.						
A2I12		ВОЛНЕ		334 6 234 6	00129		28			*	WAVE		598.						
A2I13		В		8E112	00130		29			A	IN -		402.						WAVELENGTH ON THE AVERAGE
A2I14		СРЕДНЕМ		2440	00131		30			B	THE AVERAGE		9057.		2440 34				
A3+01		7		C7	00132		31				(EQUATION)		6409.						
A3+02		ГРАДУСОВ		216	00133		32			3	OF DEGREES		969.						
A		ПОЛУЧАЕМ		6 C5A	00134		33				OBTAIN		3927.						
A		ТЕМПЕРАТУРУ		23208	00135		34				TEMPERATURE		5430.						

translation can be read by following this column down the page.

The reader will note that alternative English equivalents are sometimes printed in fields adjacent to the first English equivalent. The alternative equivalents printed to the right are sometimes preferable to the first; substitution is made by the reader or posteditor as necessary.

On the worksheet, a homograph, i.e., Russian form with two different grammar codes and corresponding English equivalents, occupies one line. The grammar-code symbol of a homograph contains ++ in the first two positions; English equivalents appear in fields E 1—E 4 and E 3—E 6, while appropriate individual grammar codes appear in fields E 2 and E 5. After the

posteditor has examined text and selected the desired English equivalent for the homographic occurrence, he replaces the original grammar-code symbol with the symbol corresponding to his choice of English equivalent.

When an idiom, i.e., a group of Russian forms translated as a group when they occur together in fixed sequence, is recognized by the computer program, the English equivalent of the idiom is printed next to the first form in the idiom. The English-equivalent fields of subsequent forms within the idiom are blank.

The special codes, printed at the right, convey information about English grammar, inflection, etc.; see The RAND Corporation MT Study 7 for details.

The coding spaces are filled in by the computer and the posteditors; Sections 2, 3 and 4 of the present Study describe their content in detail.

2. Choice of English Equivalents

2.1 SELECTION

For each occurrence in the Russian text, the posteditor selects an English equivalent. The posteditor must be guided, first, by the customary criteria for translation: accuracy and readability. However, variation for the sake of stylistic excellence is not allowed; the posteditor must expect the finished translation to be clear but dull. The order of the English equivalents on each line, from 1 to 6, is such that *E 1* is preferred more often than any other. *The posteditor must accept E 1 whenever it gives a clear, accurate translation of the original text.* The alternative equivalents are listed because they are occasionally essential for accuracy; when one of the alternative equivalents is definitely more accurate, it must be selected. The posteditor can also, when necessary, insert new alternative equivalents and recognize new idioms or homographs.

2.2 CODING

The column of the coding space marked *PE* is reserved for a one-position English-equivalent code. If the editor chooses the first English equivalent, he does not mark the space.

If he selects an alternative English equivalent, the posteditor writes 2, 3, etc., in the coding space, as appropriate, using the numbers printed as column headings on the worksheet.

To add an English equivalent, the posteditor writes an asterisk (*) in the English-equivalent coding space, and writes the new English equivalent in BLOCK CAPITAL LETTERS in the right-hand margin.

When the posteditor selects the first form of a *homograph*, he leaves the *PE* coding space blank. If he selects the English equivalent appearing in a field other than *E 1*, the number of the field from which the equivalent was chosen is inserted in the coding space.

To identify a new homograph, the posteditor writes *H* in the English equivalent coding space beside the homographic form; no further coding is required. A new English equivalent may be written in the margin if necessary.

When the posteditor accepts the first English equivalent of an idiom, he leaves the *PE* coding space blank. If the second English equivalent is desired, the posteditor writes *P* beside each word in the idiom; if the third English equivalent is desired, the posteditor writes *Q* beside each word in the idiom.

To identify a new idiom, i.e., one which is not recognized by the computer program, the editor writes *A* in the English equivalent coding space beside the first form in the idiom, *B* beside the second, *C* beside the third, and so forth. He also writes the English equivalent of the idiom in BLOCK CAPITAL LETTERS in

the right-hand margin, opposite the *first* form of the idiom. If an idiom is recognized by the computer, “—” is printed beside each form of the idiom.

If, in non-idiomatic combinations, the posteditor wishes to omit translation of an occurrence, he writes the numeral *O* in the coding space.

(Examples of all coding rules are illustrated in Table 2.)

3. English Structural Symbols

3.1 ENGLISH MORPHOLOGY AND SYNTAX

The computer program that prints the worksheet also begins the conversion of Russian structural symbolism into English, but the posteditor must complete this task. The translation, when it leaves the posteditor, must be clear and readable in construction as well as in diction; the main tools to be used are inflection and the insertion of English function words.

Inflections are rarely stored in the glossary; that is, the English equivalents stored in the glossary are usually in canonical form. For example, the singular forms of nouns, the infinitive forms of verbs (without *to*), and similar uninflected forms are usually stored. The English equivalent of a genitive Russian noun does not include, in the glossary, the preposition *of*, nor does the English equivalent of a reflexive Russian verb include the auxiliary verb *is* or *are*.

As studies of Russian-English translation progress, the computer program is improved; the work of the posteditor diminishes correspondingly. The following description of the posteditors' task assumes no modification of glossary entries by the computer. Whatever part of the work has been performed correctly by the computer is omitted by the posteditor, while any errors that the computer program has introduced are corrected by the insertion of accurate entries in the coding space. When the computer performs an inflection, it also prints a mark in the coding space.

3.2 ENGLISH INFLECTIONS

The posteditor inflects the English equivalents, as necessary for accuracy and clarity, in any of the following ways:

Nouns: plural. When a Russian noun occurs in plural number, the English equivalent is coded to show plurality.

Verbs: past tense. When a Russian verb occurs in the simple past tense, its English equivalent is coded to show past tense, except in constructions with *бы*.

Verbs: third person singular, present tense. When a Russian verb occurs in third person singular, present tense, active voice, its English equivalent is coded to show that *s* must be added.

Verbs: present participle. When a Russian present active participle occurs, or when the English equivalent of a verb must be given in progressive form (e.g., *is going*), the English equivalent is coded to show present participle inflection.

Verbs: past (passive) participle. When a Russian verb occurs in a form which must be translated into passive voice, the English equivalent is coded to show past participle inflection. This category includes most Russian reflexive constructions, passive and reflexive participles, and constructions with *obi*.

Adjectives: comparative. When a Russian adjective occurs in comparative degree, the English equivalent is coded to show comparative inflection.

Adjectives: superlative. When a Russian adjective occurs in superlative degree, the English equivalent is coded to show superlative inflection, unless the English equivalent is listed in superlative form.

Adjectives: adverb. When a Russian adjective-adverb homograph occurs as an adverb, the English equivalent is coded to show adverbial inflection.

Adjectives: comparative adverb. When a Russian adjective-adverb homograph occurs in comparative degree as an adverb, the English equivalent is coded to show comparative-adverb inflection.

3.3 ENGLISH INSERTIONS

The posteditor codes the insertion of an additional English word whenever necessary for accuracy or clarity, choosing from the following list:

Pronoun subjects: it, there, I, we, they, let us, who, one. One of these pronoun subjects is inserted whenever a Russian sentence construction includes a verb with no subject, unless context makes the omission definitely preferable in English.

Verb auxiliaries: are, was, were, do, does, did, will, will be, be, am, being, is, to, to be. The verb auxiliaries are inserted to construct passive voice, negation, past tense, future tense, or progressive form in English, as required by the construction of the Russian sentence.

Articles: a, an, the. An article is inserted in English whenever it contributes to accuracy or clarity. The articles *a* and *an* are not distinguished.

Connections: of, to, by, with, than, as, in, on. English connecting words must be inserted in the absence of a Russian equivalent, in two kinds of context situations: when the (oblique) case of a noun in Russian expresses a relationship which can best be expressed by a preposition or a conjunction in English; and when a Russian verb without a preposition requires a noun object which must be connected to the English-equivalent verb by a preposition. In either case, the connecting word must be inserted by the posteditor.

3.4 ENGLISH STRUCTURAL-SYMBOL CODE

A four-position space is included on the worksheet for coding English structural symbols. In the first position, the posteditor codes pronoun subject insertions (see Table 3). In the second position, the posteditor codes auxiliary verb insertions (see Table 4). In the third position, the posteditor codes insertions of articles and prepositions (see Table 5). In the fourth position, the posteditor codes miscellaneous inflections: verbs, participles, noun plurals, and adjective inflections (see Table 6).

TABLE 3
SYMBOLS REPRESENTING PRONOUN SUBJECT INSERTION
Position 1

Pronoun Insertion	Symbol
It	1
There	2
I	3
We	4
They	5
Let Us	6
Who	7
One	8

TABLE 4
CODE SYMBOLS FOR AUXILIARY VERB INSERTIONS
Position 2

Auxiliary Verb	Code Symbol Insertion
Are	1
Was	2
Were	3
Do	4
Does	5
Did	6
Will	7
Will Be	8
Be	9
Am	A
Being	B
Is	C
To	D
To Be	E

TABLE 5
CODE SYMBOLS REPRESENTING INSERTION OF ARTICLES
AND ENGLISH CONNECTING WORDS
Position 3

Preposition Insertion	Article Insertion		
	None	A, An	The
None		+	—
Of	1	A	J
To	2	B	K
By	3	C	L
With	4	D	M
Than	5	E	N
As	6	F	∅
In	7	G	P
On	8	H	Q
From	9	I	R

TABLE 6
CODE SYMBOLS FOR MISCELLANEOUS INFLECTIONS
Position 4

Inflection	Code Symbol
Short-form, neuter adjective/adverb performing adverbial function (<i>381</i>)	-
Noun plural	3
Positive comparative for adjectives and adverbs (modified by <i>более</i>) (addition of <i>er</i>)	4
Positive superlative for adjectives and adverbs (modified by <i>наиболее</i>) (addition of <i>est</i>)	5
Negative comparative for adjectives and adverbs (modified by <i>менее</i>) (addition of <i>er</i> or <i>less</i>)	6
Negative superlative for adjectives and adverbs (modified by <i>наименее</i>) (addition of <i>est</i>)	7
Third person singular present tense for verbs (addition of <i>s</i>)	A
Past tense for verbs (addition of <i>ed</i>)	B
Present participle verb form (addition of <i>ing</i>)	C
Past participle verb form (addition of <i>en</i>)	D

The tabulations of these codes are readily understood, with the possible exception of Table 5. Whereas the insertion of the article *a* or *an* is represented by the symbol "+", and the insertion of the preposition *of* is represented by the number *I*, when insertion of both the article and preposition are required for the same occurrence, the symbol is not + *I* but *A*. This method of symbol combination derives from the properties of IBM machines; when the letter *A* is punched into an IBM card, it is represented by two punches, "+" and *I* in a single card column.

The posteditor must be careful to distinguish the characters *G*, *C* and *6* from one another; the numeral *O* from the letter *0*; the numeral *I* from the letter *J*; the letters *U* and *V* from each other; and the numeral *5* from the letter *S*.

The line on which the codes are written must be determined in accordance with the following rules:

Verb inflections, pronoun-subject insertions, and auxiliary-verb insertions must be coded on the line on which the *verb* occurs.

Preposition insertions, article insertions, and noun plural inflections must be coded on the line on which the *noun* occurs, even though the preposition or article must actually be inserted before an adjective, for example.

Adjective or adverb inflections must be coded on the line on which the inflected word appears.

4. Structural Coding

4.1 DEPENDENCY

Sentence structure can be analyzed in many ways; one plan, which is convenient for the present research, is

based on the assumption that every occurrence in a sentence depends on some other occurrence in the same sentence (except that one occurrence in each sentence is independent). The concept of dependency is partly syntactic, partly semantic; the posteditor must have a good understanding of Russian grammar and a general familiarity with the subject matter of the scientific articles that are being analyzed if he is to do an accurate job of coding sentence structure. The posteditor must adhere, as closely as possible, to the rules laid down in this section, since the work of several posteditors is to be compared.

Syntactically, one occurrence depends on another if the inflection of the first depends on the nature of the second. Thus, it is generally said that a preposition governs the case of its noun object; hence, a noun used as the object of a preposition depends on that preposition. Semantically, one occurrence depends on another if the meaning of the first complements or modifies the meaning of the second. These definitions are related in a natural language, so it is not important to keep them distinct and to choose one or the other as a guide to postediting. Both definitions can serve as guides to the task.

The one general rule to be observed in postediting is that every occurrence must be coded as depending on one and only one other occurrence in the same sentence—an exception to this rule is made for relative clauses. One and only one occurrence in every sentence is independent. The style of Russian technical articles sometimes permits two or more independent clauses to be joined without conjunctions, so that, in effect, two sentences can be compressed into one. In such instances, the posteditor is free to establish two independent occurrences in one sentence.

4.2 RESULTANT CODING

Because usage is the factor finally determining the properties of a word, the posteditor is required to resolve grammar-code symbols appearing with Russian occurrences on the print-out sheets.

Original grammar-code symbols are those appearing in the RAND glossary with each Russian form. Individual words possess varying degrees of morphological and semantic ambiguity; further they may be capable of fulfilling a multiplicity of syntactic functions. The original grammar-code symbol is designed to reflect the intrinsic ambiguity of a given form.

Resultant grammar-code symbols are the symbols applied immediately above the original grammar code symbol on the print-out sheet after ambiguity has been resolved. Resolution is achieved mechanically whenever possible, but final responsibility for the task must rest with the posteditor. Only after examination of text is it possible to determine the unique function of a given occurrence.

Resultant grammar-code symbols presently fall into the following major categories:

(a) Resultant symbols for nouns, pronouns, adjectival pronouns (part of speech *A*), and homographs.

For example, the feminine substantive *линии* can only be imprecisely identified as a singular noun in the genitive, dative or prepositional case, or as a plural noun in either the nominative or accusative case. Assuming that examination of text has allowed the editor to determine that an occurrence of *линий* is used as a singular noun in the genitive case, the original symbol 23D is changed to 230, a precise identification of both case and number of the occurrence.

In the case of homographs, after the posteditor has examined text and selected the desired English equivalent for the entry, he replaces the ambiguous “+ +” by which the form is originally identified, with the appropriate individual grammar code.

(b) Resultant symbols for parts of speech serving as governors of substantives. Included in this category are verbs or participles acting as governors of substantives, and substantives acting as governors of other substantives. The symbols are designed to show satisfaction of a function for which the word was originally coded. Their application serves to establish complementation of the governing occurrence.

For example, assume that a verb originally coded to take objects in both the accusative and dative cases is found to be complemented only by a substantive in the accusative case. The original symbol is changed to show that possible complementation has been partially fulfilled. If, on the other hand, the occurrence is found to have both direct and indirect objects the resultant grammar-code symbol shows complete satisfaction of the verbal complementation code.

(c) English-equivalent selection-code symbols for prepositions, degree-marking adverbs and adjectives, auxiliary verbs and particles. The symbols determine the selection of one among several possible English translations and indicate the syntactic function of the occurrence. For example, a preposition serving as the head word of a simple prepositional phrase may well be translated differently and serve a different syntactic function than the same preposition serving as the head word of an idiomatic occurrence. Similarly, when *быть* acts as an independent verb, its grammar-code symbol must show it to have different properties than when it serves as an auxiliary.

(d) Resultant-code symbols for conjunctions and relative pronouns (both as governor and dependent). Conjunction grammar-code symbols are refined to identify the occurrence as coordinating or subordinating, as well as simple or compound. Certain words are capable of performing a conjunctive function alone, and they can also be used as one member of a larger conjunctive frame. Similarly, certain conjunctions may be used with a given class of governor, or their use may have no such restriction. It should be pointed out that the posteditor must apply a resultant code to establish the governor of a subordinate conjunction, while governorship of a coordinate conjunction is not indicated by a grammar-code symbol. Specific examples of resultant grammar codes appear in 4.3, and are more completely attested in the grammar code manual.

Assuredly, existing resultant codes will not suffice for every possible function of an occurrence; their number will continue to grow as more text is analyzed.

4.3 TENTATIVE DEPENDENCY RULES

The following rules are furnished as a guide; the list is not complete, since more rules will surely be added in the course of postediting large volumes of Russian text. When necessary, the posteditor deviates from these rules in order to adhere to the more general principles of completeness and syntactic-semantic consistency.

Within the structure of a phrase or clause, it is useful to distinguish the single occurrence which serves as its representative, or principal, element. This element we shall call simply the *main element*. As outlined below, for prepositional phrases, the main element is the preposition. For clauses, the main element is normally the verb or other verbal element (short-form adjective or participle). In coding the dependency of phrases and subordinate clauses, it is important to note that the relationship is indicated through the dependency of the main element of the governed structure on the most closely related element of the governing structure.

Dependency-coding rules are classified according to part of speech.

Cardinal numbers. A cardinal number is generally treated as an adjective; see below. Cardinals can also be used as nouns, e.g., *Three were chosen*. In such instances, they are assigned nominal dependency.

Ordinal numbers. An ordinal number is treated as an adjective.

Particles. Generally, a particle depends on the occurrence whose meaning it modifies or intensifies. Modifying particles (*бы,нибудь,будто*, etc.) usually depend on the preceding word, while intensifying particles (*даже,вплоть,же*, etc.) may depend on the preceding or the following element.

When the particle *пусть* appears with a finite verb or short form adjective, it is said to be the independent element.

Pronouns. In general, pronouns are treated as nouns; see below. However, relative pronouns (*который, что,какой*, etc.) have twofold functions. A relative pronoun serves as a noun within a subordinate clause, and its nominal dependency must be encoded. The same pronoun also serves to connect the subordinate clause with an element of the main clause of the sentence, and the connection must be coded as well. Relative pronouns, therefore, generate *double* dependencies.

The *first* dependency of a relative pronoun is upon the word that determines its case. For example, in the fragment . . . *которая подтверждается*, the relative pronoun is in the nominative case since it is used as the subject of the verb: '*which is confirmed.*' Again, in the fragment *у которого имеется*, the relative pronoun is object of the preposition *у*, and the prepositional phrase modifies the verb: '*for which there exists.*' The relative pronoun can even modify a noun in the subordinate clause: *сущность которого хорошо известна =*

'the substance of which is well known' The relative pronoun depends *first* upon the verb, the preposition, the noun, etc., that governs its nominal function within the subordinate clause.

The *second* dependency of a relative pronoun is upon its antecedent outside the subordinate clause. For example, in the fragment *фосфора, у которого имеется* = 'phosphorus, for which there exists', the pronoun depends on *фосфора* as antecedent. The pronoun *который* must agree with its antecedent in number and gender. The antecedent of *что*, when this word is used as a relative pronoun, is an entire clause, so agreement is irrelevant: *Кривая принимает новый вид, что указывает на ...* = 'the curve assumes a new form, which indicates . . .'. The *first* governor of *что* is the main element of the subordinate clause; the *second* governor of *что* is the main element of the independent clause. The second dependency of any relative pronoun, however, ties the subordinate clause into the sentence.

Nouns. A noun in the nominative case, serving as the subject of a sentence, can depend upon a finite verb, a shortform adjective or participle, or other predicate element. In a sentence such as *X—функция* = 'X is a function, for example, the symbol X is treated as a noun in the nominative case and is the independent element.

A noun in the genitive case, and occasionally in another oblique case, can serve as the complement of another noun. For example, *частиц* depends on *поле* in the phrase *поле неподвижных частиц* = 'field of the fixed particles'; *частиц* and *нуклонами* depend on *рассеяние* in the phrase *рассеяние частиц нуклонами* = 'diffusion of the particles by nucleons'.

Several nouns have been given grammar codes which indicate they can act as governors of other nouns. For example, *рассеяние* is coded to take complements in both genitive and instrumental cases. When genitive complementation has occurred, the symbol is changed to show complementation has been partially satisfied; when both genitive and instrumental modifiers have occurred, the complementation code is blanked out. A complete list of noun complementation types and resultant symbols appears in the grammar code manual.

A noun in an oblique case can be governed by a verb, an active or passive participle, a preposition, a shortform or comparative adjective, etc. Note that several nouns in a given sentence can be governed by the same verb; one in the nominative case, one in the accusative, etc. However, if two or more nouns are used as subjects, direct objects, etc., of the same verb, the rules of conjunction apply; see below. When the original grammar-code symbol of the noun is ambiguous, it is resolved to show the actual function of the occurrence in text (i.e., subject, object, etc.).

Adjectives. normally, a long-form adjective depends on the noun with which it agrees. It should be pointed out that several adjective/pronoun homograph forms have been formally designated as part of speech A. The grammar-code symbols of such words are converted to show the adjectival or pronominal qualities of the occurrence, as the case may be. Long-form adjective/noun

homographs are resolved as either adjectives or nouns, depending upon the function of the occurrence.

A short-form predicate adjective can serve as the independent element of the sentence. For example, in a sentence of the type *человек умен* — 'the man is wise' the adjective is independent and receives a resultant grammar-code symbol to indicate its subject-taking function.

Long-form adjectival predicates depend on the nouns which they modify.

Participles. Active and passive participles acting as noun modifiers are usually treated as adjectives. When an active reflexive participle modifies a noun, its grammar-code symbol is converted to that of a passive participle, while an active participle that follows the noun it modifies is classed as a gerund. This transformation is effected to indicate more clearly the syntactic function of such occurrences.

Short-form passive participles appearing with *быть* in modal constructions are considered to be independent. However, long-form passive participles appearing in the same construction are dependent on *быть*.

Verbs. A verb is normally the independent element of the sentence, or the main element of a dependent clause. In the latter instance, it depends secondarily on a subordinate conjunction such as *если* = 'if' or *хотя* = 'although' or on a relative adverb.

In constructions utilizing a modal (e.g., *можно, легко*) plus an infinitive, the infinitive is considered the main element in the clause and is said to govern the modal. In such constructions, and in impersonal constructions, a direct object is said to depend upon the infinitive. Thus, *условие* depends upon the infinitive in *следует определить условие* = 'it is necessary to determine the condition' as well as in *мы можем определить условие* = 'we can determine the conditions' Also, *мы* depends on *определить* rather than on *можем*.

In constructions utilizing a modal, the auxiliary infinitive *быть*, and a short form past passive participle, the modal depends upon *быть*, which depends upon the participial form—the independent element of the chain. If, however, the auxiliary is used in either past or future form (e.g., *было* or *будет*), it serves simply as a tense marker and is made to depend upon the modal.

Original grammar-code symbols of verbs are converted to resolve subject-taking and complementation functions of the occurrence. These code symbols are attested in the grammar code manual.

Prepositions. A preposition and its noun complement (together with any dependents of the noun) form a prepositional phrase; the phrase is a modifier and is similar in function to an adjective or an adverb. The preposition is said to depend on the occurrence that is modified by the phrase; this element can be a noun, verb, active or passive participle, adjective, adverb, pronoun, or cardinal number. When a prepositional phrase appears to modify an entire sentence or clause,

the preposition depends on the main element of the sentence or clause.

When the title of an article is a prepositional phrase, e.g., *О взаимодействии антипротонов с ядрами* = 'On the interaction of antiprotons with nuclei', the preposition is said to be the independent element.

The posteditor is expected to resolve the 4th and 5th position grammar code of prepositions if this has not been correctly done by machine.

Adverbs. Ordinarily, an adverb depends on a verb, adjective, or other adverb. Relative adverbs introduce dependent clauses (the clause can modify a noun, verb, etc.); the relative adverb depends first on the main element in the dependent clause, second on the proper element in the modified clause. The main element in the dependent clause is primarily independent, but secondarily it depends on the relative adverb.

Conjunction. Coordinate conjunctions, such as *и* = 'and', *или* = 'or', connect elements of the sentence that are similar in structure and identical in function. The conjunction is said to join two words, two phrases, or two clauses. In such instances, the sections joined must be developed so that the conjunction depends on the main element in the following section, and the main element in the preceding section depends on the conjunction.

In a sequence of coordinate elements (e.g., *A, B, and C*), all the elements except the last depend on the conjunction, and the conjunction depends on the last element. If there is no conjunction in the sequence, as in a series of equations, all elements except the last depend on the last element.

Such coordinating conjunctions as *либо ... либо* = 'either . . . or', *ни . . . ни* = 'neither . . . nor', and '*как . . . так и . . .*' = 'both . . . and' form idiomatic conjunctive frames, connecting semantically parallel words or phrases. The main elements, which are of similar form and identical function and follow the two units of the conjunction, must be located. Then the first unit of the conjunction and the main element of the first phrase depend on the second unit of the conjunction, which in turn depends on the main element of the second phrase. For example, in the construction, *как линия, так и точка* = 'both the line and the point', *как* and *линия* depend upon *так*, which depends on *точка*. *и* here is functionally little more than a particle depending on *так*. Similarly, in the construction *не только, X, но и Y, не* depends on *только*, which depends on *но*; *X* depends on *но*, as does the particle *и*, and *но* depends on *Y*.

Simple subordinating conjunctions, such as *хотя* = 'although', *если* = 'if', *причем* = 'whereas', introduce dependent clauses. The conjunction depends on the main element in the modified clause, and the main element in the subordinate clause depends on the conjunction.

The double conjunction *если . . . , то ...* = 'If... then' conjoins two clauses of unequal value. The main element in the dependent clause depends on *если*; *если* is made to depend on the conjunction *то*, which

in turn shows dependency on the main element in the independent clause.

Compound subordinating conjunctions consist usually of two words *так как* = 'since', *так что* = 'so that', *тогда как* = 'whereas'; or of a unit involving a prepositional phrase: *для того, чтобы* — 'in order', *в том, что* = 'in/of the fact that', *после того, как* = 'after'. Each element of the combination is said to depend on the preceding element within the combination; the first element depends on the element of the modified clause to which it is most directly related, and the main element of the subordinate clause depends on the last element of the combination.

The conjunctions *то же . . . что и* = 'the same as', and *такой же ... как и* = 'the same ... as' are idiomatic and generally indicate ellipsis of elements within the sentence structure. RAND studies have determined that the construction is used to conjoin two subjects of a single verb, two clausal modifiers or a clausal modifier and a transform of the clausal modifier used as the subject. Dependency is most conveniently established through *что—и* appears to have little syntactic significance for the construction.

Resultant grammar-code symbols identify conjunctions as coordinating or subordinating, as a single occurrence or part of an idiomatic frame, etc. Interphrasal/inter-clausal behavior of this part of speech is more fully documented in the grammar code manual.

Symbols. A symbol that is hyphenated to a noun (e.g., *х-функция*) depends on the noun. Otherwise, a symbol is treated in a manner consistent with its behavior in the sentence.

Equations. An equation can be used as a noun, as a clause, etc.; the editor determines the function of each occurrence and treats it as required by the foregoing rules.

4.4 ELLIPSIS

A common construction in Russian, especially frequent in the scientific text for which this handbook is to be used, is the conjunction of two or more phrases or clauses with omission, or ellipsis, of key words in repetition. For example, the author may write *в результате столкновения нуклона с дейтроном и дейтрона с ядрами* = 'as a result of the collision of a nucleon with a deuteron and of a deuteron with nuclei', omitting *столкновения* after the conjunction. Another example is *функции A, B нормированы на единицу объема, функция C—на единицу* = 'The functions A, B are normalized to unit volume, function C to unity'. In the latter sentence, ellipsis of *нормирована* is indicated by the dash.

The importance of the ellipsis is suggested by the fact that *на* must be referred to its governor and to its dependent for accurate translation.

The structure of a sentence containing an ellipsis is restored by the posteditor to non-elliptic form. The missing word or phrase is re-entered and dependencies are described as if it were present. Thus, in the first example above, the conjunction *и* joins two occurrences

of *столкновения* one real and one fictitious. The *real* occurrence governs *нуклона* and *с дейтроном*, while the fictitious occurrence governs *дейтрона* and *с ядрами*. In the second example, there is a conjunction of two clauses: *функции А, В нормированы на единицу объема* and *функция С нормирована на единицу*. Once the omitted element has been restored, the structure is obvious; it can be determined by the rules of Section 4.3.

4.5 CODING

The first portion of the coding space (*DC*) is used for a two-position dependency code. For all but one of the occurrences in a sentence, the posteditor indicates dependence on another occurrence in the same sentence. One occurrence in each sentence is coded as independent, except in a complex sentence or a sentence consisting of two or more complete clauses separated by commas.

Within each article, the computer assigns sequence numbers to sentences, and within each sentence, it assigns sequence numbers to occurrences. The two-digit occurrence-within-sentence number is used for dependency coding. If occurrence N_1 depends on occurrence N_2 , the posteditor writes N_2 in the coding space on line

N_1 . The posteditor writes *OO* in the coding space of the independent occurrence in each sentence.

In the case of a subordinate clause, the posteditor is required to reflect the dual dependency of both the introductory element and the verbal element in the clause that the relative introduces. He does so by writing an asterisk in the coding space for each such occurrence, and by writing two dependency symbols on the extreme right-hand margin of the sheet; the occurrence number of the first governor is written first and followed by the occurrence number of the second governor. The same plan is followed in every subordinate clause.

To restore an elliptically omitted word, the posteditor adds a line on the worksheet at the end of the sentence. Page number, line number, Russian form, Russian inflectional grammar-code symbol, Russian resultant grammar-code symbol, sentence number, occurrence number (*1E, 2E, 3E, etc.*, for several ellipses within a sentence), dependency-code symbol and word number must all be filled in. The dependency-code symbol for a restored word is the occurrence number of the word on which it would have depended if it had actually occurred. The words depending on the restored word have dependency symbols *1E (2E, 3E, etc.*, if they depend on the second, third, etc., restored words).

Received January 18, 1960