

## The Morphological Abstraction of Russian Verbs

by Milos Pacak\*, assisted by Antonina Boldyreff, Institute of Languages and Linguistics, Georgetown University

1. *The purpose of this paper is the establishment of classes of verbals according to the morphemic alternations of base-form finals;*
2. *Verbals which are subject to morphemic alternation are treated as single entries instead of as multiple entries;*
3. *The patterns of compatibility between a given set of compound suffixes and a class of verbal bases are designed to be suitable whether used as input for translation from Russian or as output during translation to Russian;*
4. *The proposed procedure is flexible; it can be modified or added to without any change in the logical structure;*
5. *This procedure can be applied to other Slavic languages as well.*

### Preface

This report is a continuation of an earlier study\* of Russian morphology as prescribed by the demands of machine translation.

There are three main reasons why it has been found necessary to handle the morphology of Russian verbs in a separate paper.

1. The idea of using infix operations for the recognition of participle forms has, for programming reasons, been temporarily abandoned.

2. The high frequency of verb-base alternations has led to the conclusion that some procedure should be worked out which would make it possible to list as single entries those verb bases which are subject to alternations (see Appendix VII), and to decrease ambiguity.

The establishment of distribution classes of Russian verb-base alternants in terms of sets of paradigmatic suffixes should demonstrate the usefulness of the suggested procedure. The listing of pertinent distribution classes is given in Appendix IV; therefore it has not been found necessary to describe them in further detail in the report itself.

3. The morphological procedures described can be used as well for input as for output.

### General Description

A previous paper described how to handle verb items, and how to identify participle forms by using infix operations.

It was stated that verb bases which were subject to morphemic alternations must be listed in the dictionary as multiple entries.

The purpose of the present study is to describe the analysis of verb morphemic alternations in terms of machine translation and of information retrieval.

\* This research was supported in part by a grant from the National Science Foundation, Washington 25, D. C. The author of this paper wishes to express his gratitude to Dr. William A. Austin and Mr. Philip H. Smith, Jr., for their suggestions concerning this paper. @1959, Georgetown University.

The frequency of verbs which undergo the process of morphemic alternation is relatively high. Therefore it seems practical to develop a procedure which would permit handling this type of verb base as single entries instead of entering two or more bases. In other words, the number of dictionary entries will be reduced.

The second aim is to establish specific classes of verb bases: their matching is bound to a *limited set* of suffixes. The mutual exclusiveness of certain types of bases with certain suffixes will result in a decrease in the number of possible ambiguities.

A base form as used here is either a simple root or a stem, depending on the type of verb involved.

A base-forming vowel, which may be zero, is assigned either to the root or to suffixes indicating infinitive, past tense, or gerund.

These two criteria of assigning the connection vowel in different ways can be justified in terms of machine translation only. The main purpose is to list a minimum number of entries with maximum combinatory possibilities. Morphemic alternations are described *only when base-form finals are involved*. In case of noncontiguous changes two or more bases must be listed.

The transliteration system used was developed by the GAT group at Georgetown University (See Appendix I.)

### Distributional Classes of Verbal Alternants

The patterns of morphemic alternations as listed in Appendix II and IV are modified according to the given set of suffixes.

Thirty-eight different patterns of morphemic alternants have been established and coded.

They fall into three major classes:

1. 1-1 alternation (24 patterns)
2. 1-2 alternations (12 patterns)
3. 1-3 alternations (2 patterns)

### Alternation Code

The four-digit code which has been used for coding different patterns of alternations is alphabetic, because

this type of code is felt to be mnemonic and easier to use.

The first digit indicates the part of speech: 2 here designates a verb form. The digits in the second, third, and fourth positions indicate the type of alternation, or alternant 2.

Example: The verb PISAT6 'write' will be entered in the dictionary thus: PIS- 2W. The W code shows that the final S (alternant 1) of the entered base for alternates with W (alternant 2). If an input form, say PIWET, is matched in the dictionary and finds no stem PIW-, the program checks for W as the only possible alternant to S. This type belongs to the group of 1-1 alternations.

An example of 1-2 alternation is the verb RISOVAT6 'draw'. It will be listed in the dictionary as RISU2OV. The one-position final U alternates with the final two-position OV.

The patterns of alternations are listed and coded in Appendix II.

### Patterns of Alternations—Base Form

The patterns of base-form alternations—as described below—are classified in terms of their positional value.

The introduction of zero functioning as alternant 1 makes it possible to treat the types which Jakobson describes as "deeper truncation" as follows:

Verbs of the type GASNUT6 will be listed as Ø-N alternation type: GAS-2N. The extension of the base by connecting the zero alternant will result in the following suffix operations:

GAS Ø Ø; LA; LO; LI.

GAS N U; EW6; ET; EM; ETE; UT.

The *positional* value of the zero alternant (alternant 1) and of N (alternant 2) is equal, but their function in the paradigm is different.

The second type, JIT6 'live', is treated similarly (Ø-V alternation). The dictionary will contain JI- 2V, and the following suffix operations will be possible:

Ji Ø T6; L; LA; LO; LI.

Ji V U; EW6; ET; EM; ETE; UT.

Verbs which are subject to concomitant changes (before dropped A in the stem the group OV is regularly replaced by U—cf. RISOVAT6) are handled as 1-2 alternants.

The base is entered with the form which ends in U, and with alternant code 2OV. This code indicates the function of OV as alternant 2 to the base final U (alternant 1). Thus, RISOVAT6 will be listed in the dictionary as RISU-2OV, and the following suffix operations will be possible.

RISU —H; EW6; ET; EM; ETE; HT; 4.

RISOV—AT6; AL; ALA; ALO; ALL

In the same category fall 1-2 alternation types U-EV (JEVAT6) and H-EV (PLEVAT6), in which the group EV is replaced by U or H.

Types in which O is inserted before the base-final consonant are listed as V-OV, N-ON, and B-OB6 alternation patterns.

An example of V-OV; the dictionary form: POZV-

POZV —AT6; AL; ALA; ALO; ALI.

POZOV—U; EW6; ET; EM; ETE; UT; 4.

An example of N-ON alternation; dictionary form: DOGN-

DOGN —AT6; AL; ALA; ALO; ALI.

DOGON—H; IW6; IT; IM; ITE; 4T; 4.

An example of B-OB alternation; dictionary form: RAZB-

RAZB —IT6; IL; ILA; ILO; ILI.

RAZOB6—H; EW6; ET; EM; ETE; HT.

The pattern R-ER includes two types of alternations: one is the type BRAT6 'take', where E is inserted before the final R; the other is type TERET6 'rub', where E is dropped before the final R. Examples:

BR —AT6; AL; ALA; ALO; ALI.

BER—U; EW6; ET; EM; ETE; UT; 4.

TR —U; EW6; ET; EM; ETE; UT.

TER—ET6; 0; LA; LO; LI.

The reason why both types are classified as R-ER alternation is purely mechanical. Alternant 1 (base-final of the entered dictionary base) is always one-positional, for reasons of consistency and simplicity of search. Otherwise the type TERET6 must be listed as ER-R alternation (2-1 alternation type), which would contradict the proposed basic concept.

Bases with O final (O in monosyllabic stems and zero in non-syllabic stems) are coded as Y-O (MYT6) and 1-6 (PIT6):

MY—20

MY—T6; L; LA; LO; LI.

MO—H; EW6; ET; EM; ETE; HT; 4.

PI —26 'drink'

PI —T6; LA; LO; LI; L.

P6 —H; EW6; ET; EM; ETE; HT.

Non-syllabic bases with A final are listed as A-N and A-M alternants:

JA —2N 'mow'

JA —T6; L; LA; LO; LI.

JN—U; EW6; ET; EM; ETE; UT.

JA —2M 'squeeze'

JA —T6; L; LA; LO; LI.

JM—U; EW6; ET; EM; ETE; UT.

The semantic ambiguity of verbs mentioned above is, at least for non-past forms, solved by the alternant code (N = mow; M = squeeze).

Verbs of the type KLAST6 'put', GRESTI 'dig', PLESTI 'knit' ("convergence of final consonants in closed full stems in S before the infinitive desinence"—Jakobson) are listed as Ø-D, Ø-B, and Ø-T alternations. Consider the examples:

KLA —2D.  
 KLAØ—ST6; L; LA; LO; LI.  
 KLAD—U; EW6; ET; EM; ETE; UT; 4.  
 GRE —2B.  
 GREØ—STL  
 GREB—U; EW6; ET; EM; ETE; UT; Ø; LA; LO;  
 LI; 4.  
 PLE —2T.  
 PLEØ—STI; L; LA; LO; LI.  
 PLET—U; EW6; ET; EM; ETE; UT; 4.

Verbs of the type NESTI ‘carry’ are treated as zero alternation type, and are coded 2000F. They are entered as single bases (see Appendix III).

NES—2000F.  
 NES—TI; U; EW6; ET; EM; ETE; UT; Ø; LA;  
 LO; LI; 4.

Types with soft final consonant which preserve their softness throughout the paradigm with the exception of the first person singular, non-past, are coded in the following way:

Type T—C:	XOT —2C	(XOTET6)
Type K—C:	VLEK —2C	(VLEC6)
Type S—W:	NOS —2W	(NOSIT6)
Type G—J:	BEG —2J	(BEGAT6)
Type D—J:	VOD —2J	(VODIT6)
Type Z—J:	VOZ —2J	(VOZIT6)

As for the suffix operations, the reader is referred to Appendix VI.

Alternation types ST—5 (PUSTIT6) and SK—5 (ISKAT6) are coded as 2ST and 2SK alternations, for the reasons explained above: the starting point of alternation operations is *always* and *only* the one-position final of the listed base.

Verbs of the type STAVIT6, LHBIT6, GRAFIT6 can be included in the category of Ø—L alternation. Example:

LHB —2L.  
 LHB —IT6; IW6; IT; IM; ITE; 4T; IL; ILA; ILO;  
 ILI; 4.  
 LHBL—H.

Types with hard final consonant in the base, when followed by A, exhibit the following alternations:

Type K—C:	PLAK—2C	(PLAKAT6).
Type S—W:	PIS —2W	(PISAT6).
Type Z—J:	V4Z —2J	(V4ZAT6).

These types of alternations were mentioned above. The reason they are repeated is because of the different function of alternants with regard to the matching possibilities within the given set of suffixes.

Alternation type K—C includes four different types of conjugation subclasses in terms of the “matching” value of alternant 1 (K) and alternant 2 (C).

Alternant 1 (K) within the same type of alternation, has four different values when compared to the list of suffixes:

1. U; UT; Ø; LA; LO; LI (VLEC6).
2. TI; U; UT; Ø; LA; LO; LI (VLEKTI).
3. AT6; AL; ALA; ALO; ALI (PLAKAT6).
4. U; UT; LA; LO; LI (TOLOC6).

Note: The forms TOLOC6 and TOLOK will be listed as full forms, not subject to morphological analysis.

The same fundamental concept of conjugation subclasses has been applied to alternation pattern Ø—D, Ø—N, G—J, S—W, Z—J, D—J, T—5, T—C, R—ER, (see Appendix IV).

Types with base final in U are listed as two different patterns:

1. If the base prefinal is a vowel then this type is treated as zero alternation. Example: POM4N—2000E.

POM4N—UT6; U; EW6; ET; EM; ETE; UT; UL;  
 ULA; ULO; ULI.

2. If the base prefinal is a consonant it exhibits Ø—N alternation pattern with a different set of suffixes for the past tense (i.e. zero suffix in masculine past tense). Example: GAS—2N.

GASØ — Ø; LA; LO; LI.  
 GASN —UT6; U; EW6; ET; EM; ETE; UT.

Types with inserted E in the infinitive within a non-syllabic base (JEC6) are entered in two forms: JEC6 and JEG are entered as full forms, and the base JG— as alternation type 2J.

JG—U; UT; LA; LO; LI.  
 JJ —EW6; ET; EM; ETE.

Verbs classified by Jakobson as exceptions are entered as single-base forms with the proper alternation code (see Appendix IV). Examples:

XOTET6	‘want’	XOT	—2C
BEJAT6	‘run’	BEG	—2J
KLAST6	‘put’	KLA	—2D
MERET6		MER	—2ER
SPAT6	‘sleep’	SP	—2L
KLEVETAT6		KLEVET	—25
BRAT6	‘take’	BR	—2ER
EXAT6	‘ride’	EX	—2D
GNAT6	‘drive’	GN	—2ON
STLAT6		STL	—2EL

Two base-forms are required for types such as POSLAT6 ‘send’ and MOLOT6 ‘grind’; prefinal S alternates with W and prefinal O alternates with E in the examples given. Therefore for reasons given above two bases are necessary.

All forms of anomalous verbs (EST6 ‘eat’, ITTI ‘go’, etc.) will be listed in full.

The matrix of alternations shows the possible combinations of alternants 1 and 2 (see Appendix VIII).

#### Search for Verb Alternants and Suffix Operations

The suffixes which are listed in Appendix V include:

1. Non-terminal (prefinal) suffixes (e.g.: L);
2. Free (final) suffixes (Ø, A, O, I);
3. Compound (non-terminal suffixes plus free suffixes: LA).

For simplicity, the term suffix will be used indiscriminately for all the above three types of suffixes.

The suffixes are divided into three groups, according to length. The total number of suffixes belonging to the first group (one-letter suffixes) is 9; the second group (two-letter suffixes) contains 20; and the third (three-letter) 26. All operational verb suffixes are listed in Appendix V.

The output value of listed verb suffixes equals the recognition of non-past and past tense, present gerund, number, gender, and person.

The aspect of Russian verbs (perfective and imperfective) will be expressed by codes: X for imperfective and Z for perfective.

If an analyzed verb carries the code X then the output value of non-past suffixes will equal present tense (T2). The output value of the same suffixes will be changed to T3 (future tense) if the verb base carries Z.

Participle bases will be listed together with corresponding participle markers (N, NN, M, T, H5, U5, VW), as extended verb bases. They will be coded in the same way as adjectives, and with an additional code, indicating their participle function.

#### SEARCH FOR VERB ALTERNANTS

When a verb base has been identified by a previous lookup operation the dichotomy search is performed on two levels:

*Level A.* Search for zero-alternant type. Is the verb base 2000X (where X represents A, B, C, D, or E)? In other words, the program checks whether the base belongs to the zero-alternant type. If it does, the suffix operation goes into effect and suffixes are matched with the zero-alternant type.

*Level B.* Search for alternant 1 or 2. If the identified base carries an alternant code, the program checks for the base-final. If the stored base-final (alternant 1) is identical with the input base-final, the suffix operation continues.

If the compared bases are not identical, the program checks for alternant 2. Example: Input item is PISAT6 'write'. Dictionary form is PIS—2W. The dictionary stem matches with the first three letters of the input item, and the AT6 operation goes into effect.

The input item is PIWET. No base PIW- is found. The program checks for the only possible alternant of W, and locates S. The ET suffix operation proceeds.

#### SUFFIX OPERATIONS

There are two different approaches to performing suffix operations. They are both described here.

*Approach A.* Each listed suffix (see Appendix V) is compared with each matchable type of verb base (zero alternant type) and with alternant 1 or 2. Example: The 4T operation. If the verb base is coded 2000B or

alternant type Ø1 or D1 or Z1 or S1 or T1 or ON2 or L2 or ST2:

store: (N2•V1•P3•T2).

All pertinent suffix operations are listed in Appendix VI.

*Approach B.* Three patterns of similarity and dissimilarity of functional alternants of verb bases have been established, in terms of the set of suffixes they can take:

1. Base-finals of the listed bases (alternant 1) Ø, G, A, Y, I, X, U, H, R, Z, S, 4, K.
2. Base-finals functioning as (alternant 2); i.e., they occur only as alternants with the base-final 1: C, M, O, 6, W, EL, OV, IM, SK, ST, EV, ON, ER, OV, OB6, VA, IM, OJM.
3. Base-finals of the listed bases (not exhibiting base alternants 1 or 2 but followed by different sets of suffixes; they may function as alternant 1 or 2: B, N, E, D, T, V, L, 5, J.

The different types of alternant bases are listed in Appendix II and IV.

Twenty-four distinct types of suffix operations are called for, according to the positional value of listed alternants 1 or 2. By establishing the matching value of alternants 1 and 2 we proceed to the following operations:

*Operation I:* If Y1 or T1 or 41 or VA2, then: T6, LA, LO, LI, L, 4.

*Operation II:* If X1 or V1 or L1 or J1 or EV2 or SK2, then: AT6, AL, ALA, ALO, ALL

*Operation III:* If U1 or H1 or E2 or O2 or 62 or EL2 or OB62, then: H, EW6, ET, EM, ETE, HT, 4.

*Operation IV:* If N2 or T2 or 51 or 52 or M2 or W2 or IM2 or OZM2, or OJM2 or IM2, then: U, EW6, ET, EM, ETE, UT, 4.

*Operation V:* If R1 or V2 or OV2, then; U, EW6, ET, EM, ETE, UT, 4, A, AT6, AL, ALA, ALO, ALI

*Operation VI:* If B1, then: IT6, IL, ILA, ILO, ILI.

*Operation VII:* If B2, then: U, EW6, ET, EM, ETE, UT, Ø, LA, LO, LI.

*Operation VIII:* If G1, then: U, UT, Ø, LA, LO, LI, AT6, AL, ALA, ALO, ALI

*Operation IX:* If N1, then: 4T6, 4L, 4LA, 4LO, 4LI, AT6, AL, ALA, ALO, ALI.

*Operation X:* If S1, then: AT6, AL, ALA, ALO, ALI, IT6, IW6, IT, IM, ITE, 4T, IL, ILA, ILO, ILI.

*Operation XI:* If Z1, then: IT6, IW6, IT, IM, ITE, 4T, ILA, ILO, ILI, AT6, AL, ALA, ALO, ALI.

*Operation XIII:* If D1, then: ET6, IT6, IW6, IT, IM, ITE, IL, ILA, ILO, ILI.

*Operation XIII:* If D2, then: U, EW6, ET, EM, ETE, UT, 4, IM, IW6.

*Operation XIV:* If C2, then: U, EW6, ET, EM, ETE, UT, IW6, IT, IM, ITE, 6, A.

*Operation XV:* If T1, then: IT6, IW6, IT, IM, ITE, 4T, IL, ILA, ILO, ILI, AT6, AL, ALA, ALO, ALI, ET6, EL, ELA, ELO, ELI.

*Operation XVI:* If L2, then: H, EW6, ET, EM, ETE, 4T, 4.

*Operation XVII:* If J2, then: U, EW6, ET, EM, ETE, UT, IW6, IT, IM, ITE.

*Operation XVIII:* If Ø1, then: STI, ST6, T6, IW6, IT, IM, ITE, 4T, ET6, EW6, EM, ETE, HT, EL, ELA, ELO, ELI, IL, ILA, ILO, ILI, L, LA, LO, LI, Ø.

*Operation XIX:* If ER2, then: ET6, Ø, LA, LO, LI, U, EW6, ET, EM, ETE, UT, 4.

*Operation XX:* If ON2, then: H, IW6, IT, IM, ITE, 4T.

*Operation XXI:* If ST2, then: IT6, IW6, IT, IM, ITE, 4T, IL, ILA, ILO, ILI, IV, 4.

*Operation XXII:* If Z1, then: 4T6, 4L, 4LA, 4LO, 4LI.

*Operation XXIII:* If E1, then: T6, ST6, L, LA, LO, LI.

*Operation XXIV:* If A1, then: T6, LA, LO, LI, 4, H, EW6, ET, EM, ETE, HT.

The imperative suffixes have been temporarily omitted because their frequency in scientific text is not high. The most productive alternant type is LØ1, because it has consonantal and non-consonantal function. The less productive alternants are A1, Y1, E1, 41, and Z1, which can be matched with only a limited set of suffixes representing infinitive and past tense.

For pre-programming purposes the COMIT method, developed by V. H. Yngve could be used for the operations mentioned above. If we assign the value of constituents to verb bases and to the corresponding suffixes, the search for match conditions between each of the constituents can be formulated in terms of COMIT and carried out by the computer. The working out of these formulations should not be too difficult, because the various steps in the search routine are adequately described in the COMIT procedure.

#### Output Value of Suffixes

The output value of suffixes is a logical product of dichotomy operations as described above.

The principle of substitution has been used in the way described in an earlier paper. The symbols used below have the following interpretation:

233	Present passive participle
G1	Masculine gender
G2	Feminine gender
G4	Neuter gender
N1	Singular number
N2	Plural number
V1	Active voice
V2	Passive voice
T1	Past tense
F1	Long form (of adjective or participle)
F2	Short form
T2	Non-past tense
T3	Future tense
P1	First person
P2	Second person
P3	Third person
21	Infinitive
24	Present gerund
2X	Imperfective verbs
2Z	Perfective verbs.

These symbols can be replaced by any numerical or non-numerical code if desired.

*Output (21) [infinitive]:*  
If IT6, AT6, STI, TI, UT6, 4T6, C6, 6.

*Output (N1•T2•V1•P1):*  
If U or H, and 2X.

*Output (N1•T3•V1•P1):*  
If U, H, and 2Z.

*Output (N1•T2•V1•P2):*  
If EW6, IW6, and 2X.

*Output (N1•T3•V1•P2):*  
If EW6, IW6, and 2Z.

*Output (N1•T2•V1•P3):*  
If ET, IT, and 2X.

*Output (N1•T3•V1•P3):*  
If ET, IT, and 2Z.

*Output (N2•T2•V1•P1) •(233•G1•N1•F2):*  
If EM, IM, and 2X.

*Output (N2•T3•V1•P1):*  
If EM, IM, and 2Z.

*Output (24):*  
If A, 4, A4, 44, and 2X.

*Output (N2•T2•V1•P2):*  
If ETE, ITE, and 2X.

*Output (N2•T3•V1•P2):*  
If ETE, ITE, and 2Z.

*Output (N2•T2•V1•P3):*  
If UT, HT, AT, 4T, and 2X.

*Output (N2•T3•V1•P3):*  
If UT, HT, AT, 4T, and 2Z.

*Output (N1•G1•T1•V1):*  
If Ø, L, IL, AL, EL, 4L, and 2X or 2Z.

*Output (N1•G2•T1•V1):*

If LA, ILA, ALA, 4LA, ELA, ULA, and 2X or 2Z.

*Output (N1•G4•T1•V1):*

If LO, ILO, ALO, 4LO, ELO, ULO, and 2X or 2Z.

*Output (N2•G7•T1•V1)*

If LI, ILI, ALI, 4LI, ELI, ULI, and 2X or 2Z.

The output value of  $\emptyset$  suffix is the same as for suffixes L, IL, AL, 4L, and #1. In fact it functions as a final (free) suffix if matched with the corresponding type of verb-base.

The output value of Russian verb suffixes may be considered as a logical synthesis product in English translation.

### Classification and Prediction

The morphological scheme of Russian verbs could be described in terms of a theory of classification and prediction as follows:

The theory of Tanimoto is based on three assumptions:

1. Which objects are to be considered;
2. What attributes are pertinent;
3. Whether a particular object does or does not possess a specific attribute of the set of pertinent attributes.

All the objects with which we are concerned must be distinct kinds of objects, and all the attributes must be distinct too."

By applying this theory to morphological analysis of Russian verbs we could classify the verb bases as "objects" and the suffixes as pertinent "attributes". "If we consider 'B' as a finite set of 'n' objects [distinctly coded verb bases] and 'a' as a particular attribute [any suffix] possessed by some elements of 'B', then the definition of the probability 'p' that an element of 'B' [any verb base] chosen at random will possess the attribute 'a' [e.g., zero suffix] will be:

$$p = \frac{N(aB)}{N(B)} = \frac{6}{46} = 1.30$$

where  $N(aB)$  is the number of elements 'B' [number of verb bases which can be matched with suffix  $\emptyset$ ] which possess the attribute 'a' [ $\emptyset$  suffix] and  $N(B)$  is the total number of elements in 'B' [number of coded verb bases]."

In this way it would be possible to establish the probabilities of occurrence of listed suffixes in a random text. By knowing approximately the probability of occurrence of suffixes (attributes) with respect to types of verb bases, the suffixes could be stored in terms of the probability of occurrence. This new frequency order could mean a substantial saving in machine time in the lookup operations.

"If we know the finite set of attributes [suffixes] associated with the finite set of objects 'n' [types of verb bases] we can define the matrix as  $R = m \times n = 2530$ ,

in which 1 holds if some object possesses the attribute 'a' and  $\emptyset$  if it does not possess the attribute 'a'."

In other words 1 expresses the permissible matching of a given verb base (object) with a given suffix or suffixes (attributes) and  $\emptyset$  if the matching of a given verb base and a given suffix or suffixes is not permissible.

On the basis of the matrix mentioned above it would be possible to prepare two matrices of similarity.

"Matrix S ( $n \times n$ ) is the matrix of the similarity coefficients of the object B [verb base] and with regard to the set of attributes A [suffixes], and matrix Z ( $m \times m$ ) which is the matrix of the similarity coefficients of attributes A [suffixes] with respect to the set of objects B[verb bases]".

By establishing the matrices of similarity we could proceed to the theorem of prediction in terms of information theory as formulated by Tanimoto. The application of this theorem could prove very useful—mainly for purposes of information retrieval.

### Conclusions

1. The proposed procedure is flexible. It is possible to add new patterns of alterations or to modify the existing patterns without any change in the logical structure.

2. The size of the dictionary will be reduced, since only one base will be required for what are today different dictionary verb stems. The proposed system should at the same time reduce the possibility of ambiguous or wrong morphological analysis.

3. In general, the system which has been developed for Russian verbs can be applied to other Slavic languages as well. It will be of greater value for Czech and Polish because of the high frequency of morphemic alternations in these languages.

The establishment of patterns of similarity and dissimilarity on the comparative level will have the following features:

- a. Patterns of similarity will be of considerable importance for developing a more compact multi-Slavic-English dictionary.
- b. Patterns of dissimilarity might be used as recognition cues for information retrieval: some unique patterns of dissimilarity will indicate membership in a specific language. For example: the alternation R-R is the signal for Czech *only*.

4. The analytic scheme described is applicable to *input* and *output*. If the given verb is an input item it is analyzed according to the operations described above. The same operations can be used for synthesis of output items with small modifications of the suffix operations. These modifications will consist in coding the established conjugation subclasses of listed alternation types, and in formulating the required suffix operations.

5. It seems quite possible that patterns of similarity and dissimilarity could be extended to spoken languages, by establishing the phonemic and morphemic patterns for languages under consideration.

## References

1. CARLSEN, I. M. and EDWARDS, M. J.: A numericon of Russian inflections, University of British Columbia, 1955.
2. CHERRY, HALLE, AND JAKOBSON: Toward the logical description of languages in their phonemic aspect, *Language*, 1953. Vol 29. 34-46
3. DANES, F.: Intonace a veta ve spisovné češtině [Intonation and the Sentence in Standard Czech], Prague, 1958.
4. JAKOBSON, R.: Russian conjugation, *Word*, 1948, No. 3.
5. JOSSELSOHN, HARRY: Russian word count, 1952.
6. KOPECKY, L. and HAVRANEK, B.: Velky rusko-český slovník [Large Russian-Czech Dictionary], Prague, 1953.
7. LEE, C. N.: Verb transfer and synthesis, Georgetown University Occasional Papers on Machine Translation, No 18, 1959.
8. LO CATTO, E.: Grammatica della lingua russa, Firenze, 1950.
9. PACAK, M.: Scheme of Russian morphology in terms of mechanical translation, Georgetown University Seminar Paper 74, 1958.
10. POTAPOVA, N. F.: Russian, Moscow, 1955.
11. SALEMME, A. J.: Keypunch instruction manual, Georgetown University Occasional Papers on Machine Translation, No 2, 1959.
12. TANIMOTO, T. T.: An elementary mathematical theory of classification and prediction, IBM, 1958.
13. YNGVE, V. H.: A programming language for mechanical translation, *Mechanical Translation*, Vol. 5, No 1, pp 25-41, July 1958.

## Appendix I TRANSLITERATION SYSTEM

A A	E E	K K	R P	Q Ц	Y Ы
B Б	J Ж	L Л	S C	C Ч	6 Ъ
V B	Z З	M M	T T	W Ш	3 Э
G Г	I И	N H	U Y	5 Щ	H Ю
D Д	1 Ё	O O	F Ф	7 Ь	4 Я
		P П	X X		

## Appendix II

### ALTERNATION CODE

#### 1 to 1 Alternation Patterns

Type of Alternation	Code
Ø B	2B
Ø D	2D
Ø T	2T
Ø L	2L
Ø N	2N
Ø V	2V
G J	2J
N M	2M
A N	2N
Y O	2O
I 6	26
I E	2E
E O	2O
S W	2W
Z J	2J
D J	2J
4 N	2N
X D	2D
K C	2C
T 5	25
T C	2C
A M	2M
X W	2W
E T	2T

continued next page

## Appendix III

### CONJUGATION TYPES WITHOUT ALTERNATION

#### 2000A

1. CITA: (T6; H; EW6; ET; EM; ETE; HT; L; LA; LO; LI; 4)
2. BURE: (T6)
3. GUL4: (T6)

#### 2000B

1. GOVOR: (IT6; H; IW6; IT; IM; ITE; 4T; IL; ILA; ILI; ILO; 4)
2. VEL: (ET6)

#### 2000C

UC: (IT6; U; IW6; IT; IM; ITE; AT; IL; ILA; ILO; ILI; A)

#### 2000D

SOS: (AT6; U; EW6; ET; EM; ETE; UT; AL; ALA; ALO; ALI; 4)

#### 2000E

POM4N: (UT6; U; EW6; ET; EM; ETE; UT; UL; ULA; ULO; ULI; 4)

#### 2000F

1. TR4S: (TI; U; EW6; ET; EM; ETE; UT; 0; LA; LO; LI; 4)
2. RASTER: (ET6; 0; LA; LO; LI)  
RAZOTR: (U; EW6; ET; EM; ETE; UT)
3. RAST: (I; U; EW6; ET; EM; ETE; UT; 4)  
ROS: (0; LA; LO; LI)

#### 2000G

STO: (4T6; H; IW6; IT; IM; ITE; 4T; 4L; 4LA; 4LO; 4LI; 4)

#### 2000H

DERJ: (AT6; U; IW6; IT; IM; ITE; AT; A; AL; ALA; ALO; ALI)

## Appendix II continued

## 1 to 2 Alternation Patterns

Type of Alternation	Code
V OV	2OV
L EL	2EL
N 1M	21M
N IM	2IM
5 SK	2SK
5 ST	2ST
U OV	2OV
H EV	2EV
N ON	2ON
R ER	2ER
U EV	2EV
A VA	2VA

## 1 to 3 Alternation Patterns

Type of Alternation	Code
J OJM	2OJM
B OB6	2OB6

## Appendix IV

## DISTRIBUTION CLASSES OF VERB-BASE ALTERNANTS

	Ø B
GRE	Ø: (STI) B: (U; EW6; ET; EM; ETE; UT; 0; LA; LO; LI; 4)
	Ø D
KLA	Ø: (ST6; L; LA; LO; LI) D: (U; EW6; ET; EM; ETE; UT; 4) PAST6; PR4ST6
VE	Ø: (STI; L; LA; LO; LI) D: (U; EW6; ET; EM; ETE; UT; 4) BLHSTI
DA	Ø: (T6; L; LA; LO; LI; M; W6; ST) D: (IM; UT; ITE)
	Ø T
PLE	Ø: (STI; L; LA; LO; LI) T: (U; EW6; ET; EM; ETE; UT; 4) QVESTI
	Ø L
LHB	Ø: (IT6; IW6; IT; IM; ITE; 4T; IL; ILA; ILO; ILI; 4) L: (H) LOVIT6; KUPIT6
DREM	Ø: (AT6; AL; ALA; ALO; ALI) L: (H; EW6; ET; EM; ETE; 4T; 4)
SP	Ø: (AT6; AL; ALA; ALO; ALI; IW6; IT; IM; ITE; 4T) L: (H)
TERP	Ø: (ET6; EL; ELA; ELO; ELI; IW6; IT; IM; ITE; 4T; 4) L: (H)
STAV	Ø: (IT6; IW6; IT; IM; ITE; 4T; IL; ILA; ILO; ILI; 4) L: (H)
	Ø N
STA	Ø: (T6; L; LA; LO; LI) N: (U; EW6; ET; EM; ETE; UT) VSTAT6; STYT6
NAC	Ø: (AT6; AL; ALA; ALO; ALI) N: (U; EW6; ET; EM; ETE; UT)
ODE	Ø: (T6; L; LA; LO; LI) N: (U; EW6; ET; EM; ETE; UT)
KL4	Ø: (ST6; L; LA; LO; LI) N: (U; EW6; ET; EM; ETE; UT; 4)
GAS	Ø: (Ø; LA; LO; LI; 4) N: (UT6; U; EW6; ET; EM; ETE; UT)
	Ø V
JI	Ø: (T6; L; LA; LO; LI) V: (U; EW6; ET; EM; ETE; UT; 4) PLYT6; SLYT6
DA	Ø: (H; EW6; ET; EM; ETE; HT) V: (AT6; AL; ALA; ALO; ALI; A4) UZNAVAT6; VSTAVAT6
	G J
MO	G: (U; UT; Ø; LA; LO; LI) J: (EW6; ET; EM; ETE) JEC6; LEC6; BEREC6
BE	G: (U; UT) J: (AT6; IW6; IT; IM; ITE; AL; ALA; ALO; ALI) STEREC6; STRIC6

continued next page



*Appendix IV continued*

N M	PRI	N: (4T6; 4L; 4LA; 4LO; 4LI) M: (U; EW6; ET; EM; ETE; UT)
A N	J	A: (T6; L; LA; LO; LI) N: (U; EW6; ET; EM; ETE; UT; 4)
Y O	M	Y: (T6; L; LA; LO; LI) O: (H; EW6; ET; EM; ETE; HT; 4)
I 6	P	I: (T6; L; LA; LO; LI) 6: (H; EW6; ET; EM; ETE; HT) BIT6; VIT6; LIT6
I E	BR	I: (T6; L; LA; LO; LI) E: (H; EW6; ET; EM; ETE; HT; 4)
E O	P	E: (T6; L; LA; LO; LI) O: (H; EW6; ET; EM; ETE; HT; 4)
S W	PI	S: (AT6; AL; ALA; ALO; ALI) W: (U; EW6; ET6; EM; ETE; UT; A) CESAT6
	NO	S: (IT6; IW6; IT; IM; ITE; 4T; IL; ILA; ILO; ILI; 4) W: (U) PROSIT6; GASIT6
Z J	VO	Z: (IT6; IW6; IT; IM; ITE; 4T; IL; ILA; ILO; ILI; 4) J: (U) GROZIT6
	V4	Z: (AT6; AL; ALA; ALO; ALI) J: (U; EW6; ET; EM; ETE; UT) MAZAT6
D J	VO	D: (IT6; IW6; IT; IM; ITE; 4T; IL; ILA; ILO; ILI; 4) J: (U) XODIT6
	VI	D: (ET6; IW6; IT; IM; ITE; 4T; EL; ELA; ELO; ELI; 4) J: (U)
	GLO	D: (AT6; AL; ALA; ALO; ALI; A4) J: (U; EW6; ET; EM; ETE; UT)
4 N	PROM	4: (T6; L; LA; LO; LI) N: (U; EW6; ET; EM; ETE; UT; 4) M4T6; RASP4T6
X D	PRIE	X: (AT6; AL; ALA; ALO; ALI) D: (U; EW6; ET; EM; ETE; UT; 4)
K C	VLE	K: (U; UT; 0; LA; LO; LI) C: (6; EW6; ET; EM; ETE; A) PEC6; SEC6; TEC6; TOLOC6
	PLA	K: (AT6; AL; ALA; ALO; ALI) C: (U; EW6; ET; EM; ETE; UT; A)
T 5	POGLO	T: (IT6 IW6; IT; IM; ITE; 4T; IL; ILA; ILO; ILI; 4) 5: (U)

*continued next page*

*Appendix IV continued*

	KLEVE	T: (AT6; AL; ALA; ALO; ALI) 5: (U; EW6; ET; EM; ETE; UT; A)
T C	XO	T: (ET6; EL; ELA; ELO; ELI; IM; ITE; 4T; 4) C: (U; EW6; ET)
	PR4	T: (AT6; AL; ALA; ALO; ALI) C: (U; IW6; IT; IM; ITE; UT; A) WEPTAT6
	VER	T: (ET6; IW6; IT; IM; ITE; 4T; EL; ELA; ELO; ELI; 4) C: (U)
	WU	T: IT6; IW6; IT; IM; ITE; 4T; IL; ILA; ILO; ILI; 4) C: (U)
A M	J	A: (T6; L; LA; LO; LI) M: (U; EW6; ET; EM; ETE; UT) JAT6
X W	BRE	X: (AT6; AL; ALA; ALO; ALI; A4) W: (U; EW6; ET; EM; ETE; UT) BREXAT6; PAXAT6
E T	UC	E: (ST6; L; LA; LO; LI) T: (U; EW6; ET; EM; ETE; UT; 4)
V OV	POZ	V: (AT6; AL; ALA; ALO; ALI) OV: (U; EW6; ET; EM; ETE; UT; 4)
L EL	ST	L: (AT6; AL; ALA; ALO; ALI) EL: (H; EW6; ET; EM; ETE; HT; 4)
N 1M	PO	N: (4T6; 4L; 4LA; 4LO; 4LI) 1M: (U; EW6; ET; EM; ETE; UT) PON4T6; NAN4T6; ZAN4T6
N 1M	S	N: (4T6; 4L; 4LA; 4LO; 4LI) 1M: (U; EW6; ET; EM; ETE; UT)
5 SK	I	5: (U; EW6; ET; EM; ETE; UT; A) SK: (AT6; AL; ALA; ALO; ALI) ISKAT6
5 ST	PU	5: (U) ST: (IW6; IT; IM; ITE; 4T; IL; ILA; ILO; ILI; IT6; 4)
U OV	RIS	U: (H; EW6; ET; EM; ETE; HT; 4) OV: (AT6; AL; ALA; ALO; ALI)
H EV	PL	H: (H; EW6; ET; EM; ETE; HT; 4) EV: (AT6; AL; ALA; ALO; ALI)
N ON	DOG	N: (AT6; AL; ALA; ALO; ALI) ON: (H; IW6; IT; IM; ITE; 4T)
R ER	T	R: (U; EW6; ET; EM; ETE; UT) ER: (ET6; 0; LA; LO; LI) TERET6; MERET6

*continued next page*

Appendix IV continued

B	R: (AT6; AL; ALA; ALO; ALI) ER: (U; EW6; ET; EM; ETE; UT; 4)
U EV	J U: (H; EW6; ET; EM; ETE; HT; 4) EV: (AT6; AL; ALA; ALO; ALI) JEVAT6
A AVA	SOZD A: (H; EW6; ET; EM; ETE; HT) AVA: (T6; L; LA; LO; LI; 4)
J OJM	OT J: (AT6; AL; ALA; ALO; ALI) OJM: (U; EW6; ET; EM; ETE; UT)
B OB6	RAZ B: (IT6; IL; ILA; ILO; ILI) OB6: (H; EW6; ET; EM; ETE; HT; 4)

Appendix V

LIST OF SUFFIXES

One Letter Suffixes

Ø	L	4
A	I	6
H	U	M

Two Letter Suffixes

AL	IL	TI
AT	IM	T6
A4	IT	UT
EL	LA	UL
EM	LI	4L
ET	LO	4T
HT	ST	W6

Three Letter Suffixes

ALA	EW6	UT6
ALI	ILA	ULA
ALO	ILI	ULI
AT6	ILO	ULO
ELA	ITE	4LA
ELI	IT6	4LI
ELO	IW6	4LO
ETE	STI	4T6
ET6	ST6	

Appendix VI

SUFFIX OPERATIONS

ST	Ø1;
W6	Ø1;
M	Ø1;
I	2000F;
0	2000F; B2; Ø1; K1; G1; ER2;
H	2000B; 2000A; L2; O2; E2; 62; U1; ON2; H1; E12; OB62; Ø1; 2000G; A1;
U	2000C; 2000D; 2000E; 2000F; B2; T2; D2; N2; M2; V2; W2; 52; J2; G1; C2; K1; R1; ER2; 51; OV2; 1M2; 2000H; OJM2; IM2;
6	C2;
4	2000A; 2000B; 2000D; 2000F; L2; B2; T2; D2; Ø1; N2; V2; D1; J2; S1; O2; E2; T1; U1; ER2; H1; OV2; E12; 2000G; Z1; ST2; ØB62; VA2; ON2;
A	2000C; N2; J2; C2; 52; 51; 2000H;
4L	N1; Z1; 2000G;
4LA;	
4LO;	Same as 4L
4LI;	
ITE	Same as IM
IM	Same as IW6 and IT plus D2
IT	2000B; 2000C; Ø1; T1; D1; Z1; J2; S1; C2; ON2; ST2; 2000H;
IW6	2000B; 01; T1; D1; Z1; J2; C2; ON2; ST2; 2000C; S1; 2000G; 2000H;
4T6	N1; 2000G;
AL	2000D; Ø1; X1; V2; T1; D1; Z1; G1; S1; K1; Ev2; Ov2; R1; N1; EV2; SK2; V1; L1; 2000H;
ALA;	
ALO;	Same As AL
ALI;	
UL	2000E;
ULA;	
ULO;	Same as UL
ULI;	
EL	Ø; T1; D1;
ELA;	
ELO;	Same as EL
ELI;	
IL	2000B; 2000C; Ø1; T1; D1; Z1; S1; T1; ST2; B1;
ILA;	
ILO;	Same as IL
ILI;	
AT	2000C; 2000H;
L	2000A; Ø1; 41; E1; I1; Y1; A1; VA2;
HT	2000A; O2; E2; 62; U1; H1; EL2; OB62; Ø1; A1;
UT	2000D; 2000E; 2000F; B2; T2; D2; N2; M2 V2; W2; 52; J2; G1; C2; K1; C1; R1; ER2; 51; Ov2; 1M2; OJM2; IM2;
ETE	Same as EW6, ET, EM
EM	Same as EW6 and ET

continued next page

Appendix VII  
STATISTICAL RECORD OF  
OCCURRENCES

Type of Alternation	Number of Occurrences
2000A	325
2000B	38
2000C	28
2000D	10
2000E	5
2000F	5
2000G	5
2000H	4
C-K	4
D-J	41
E-T	1
G-J	1
I-Ø	6
N-M	2
S-W	18
T-C	11
T-5	8

Type of Alternation	Number of Occurrences
Ø -D	19
Ø -L	34
Ø -N	2
Ø -T	4
Ø -V	2
B-OB6	1
J-OJM	1
Y-O	3
A-AVA	6
N-IM	1
N-ON	3
R-ER	7
U-OV	111
V-OV	2
5-SK	2
5-ST	5
Z-J	18

This record is based on examination of approximately 100,000 Russian words, in text dealing with organic chemistry and metallurgy.

Appendix VI continued

ET	2000A; 2000D; 2000E; 2000F; L2; B2; T2; D2; N2; M2; Ø1; V2; W2; 52; J2; O2; E2; C2; 62; U1; R1; ER2; 51; OV2; EL2; IM2; OB62; H1; OJM2; A1; IM2;
EW6	2000A; 2000D; 2000E; 2000F; L2; B2; T2; D2; N2; M2; Ø1; V2; W2; 52; J2; O2; E2; C2; 62; U1; R1; ER2; H1; 51; OV2; EL2; 1M2; OB62; OJM2; A1; IM2;
TI	2000F; K1;
ET6	2000B; Ø1; D1; T1; ER2; 2000F;
STI	Ø1;
ST6	Ø1; E1;
T6	2000A; Ø1; 41; E1; I1; Y1; A1; VA2;
UT6	2000E; N2;
AT6	2000D; Ø1; X1; V2; T1; D1; Z1; S1; K1; EV2; R1; N1; OV2; SK2; V1; L1; 2000H; J1;
LI	Same as L
LO	Same as L
LA	Same as L
IT6	2000B; 2000C; Ø1; T1; D1; Z1; S1; T1; ST2; B1;
A4	2000E; D2; V2; D1; X1; 2000B; N2;
4T	2000B; Ø1; D1; Z1; S1; T1; ON2; L2; ST2; 2000G;

Appendix VIII  
(Matrix of Alternations)

Alternants "1"	Alternants "2"																					
	Ø	C	A	Y	I	X	U	H	R	S	Z	B	N	E	D	T	V	L	5	4	K	J
Ø	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Y	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
X	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
U	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
R	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Z	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
N	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
E	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+
D	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
T	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
K	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
J	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-