

# Morphological Classification in the National Bureau of Standards Mechanical Translation System

LEROY F. MEYERS

*The Ohio State University,\* Columbus, Ohio*

**Abstract.** In linguistic work, morphology is the study of the rules by which words undergo changes in form. Morphological classification is the systematic organization of these rules. A detailed account of the morphological classification for Russian is given as used in the National Bureau of Standards project on the mechanical translation of Russian into English. The account includes the heuristic reasoning from which the classification system evolved, as well as complete classification tables, a description of the method of their construction, and some examples of their use.

## 1. Introduction

In an earlier publication [7] describing the work done by the mechanical translation group at the National Bureau of Standards, I. Rhodes pointed out the tremendous difficulty involved in mechanizing the process of translation from one natural language, the *source* language, into another, the *target* language. (In our work, the source language was Russian, and the target language, English.) A properly operating translating routine must be able to deal with ambiguities, often interdependent, arising from the morphology, syntax and semantics of the source language.

The present report describes how the morphological features of a Russian source word are determined mechanically. This portion of the NBS translating routine has been completely coded and has been operating successfully for several years on the IBM 7094 computer and its predecessors (7090 and 704).

Russian words may be classified for morphological purposes into two principal types:

- (a) *uninflectable*: prepositions, conjunctions, particles, interjections, certain adverbs; and
- (b) *inflectable*: verbs, nouns, pronouns, numerals, adjectives, participles, other adverbs.

In Russian, inflectable words have many forms, mostly differing in only the last few letters. These forms differ in only certain easily established ways in their translations and grammatical uses. For example, the imperfective verb *читать* and its

This report is one of a series describing the present state of the National Bureau of Standards mechanical translation project. The report is devoted entirely to the morphological classification scheme used in the NBS glossary. Subsequent reports will describe the glossary arrangement and various details of the translation program.

Most of the classification scheme presented here was developed in 1961 and is based on tentative schemes tried during the period 1959-1961. There has been much interaction between translation program requirements and classification scheme ideas.

This work was sponsored by the US Army Electronics Research and Development Laboratory, Fort Monmouth, New Jersey, under Contract No. 47802-PW-61/72-72.

\* Department of Mathematics.

participles have altogether seventy-five differently spelled forms. A *full-form* dictionary which lists all of these forms as separate entries is extremely wasteful of machine storage space and consequently of lookup time during a machine run. (A dictionary of nontrivial size must be stored in the low speed memory (tape) of the computer and must be searched serially. The search time is approximately proportional to the number of entries.)

A *canonical-form* dictionary (or principal form dictionary) of the type made for human beings is difficult to construct for mechanical use, since there is no simple way to determine the canonical form of a word, or even its part of speech, if some form of the word is given in text. Context often helps the human reader to predict which part of speech to expect, and even which case, mood, tense, number, person, gender, degree, etc., is appropriate. Knowledge of inflectional patterns, both regular and irregular, helps in reducing the number of canonical forms to be tried. It might be possible to write a program to try only certain probable types of word under the assumption that certain preceding or following words have been identified, but this would already involve the syntactic and possibly semantic analysis of the remaining text.

In fact, a human being using a canonical form dictionary must first extract the important part of the word and then must make the proper changes in it to obtain a canonical form. The second part of the process can be eliminated if a *stem* dictionary is used. The *stem* of a word is the portion remaining after all inflectional markers (in Russian, endings) have been removed. A morphological classification number can then indicate which endings are appropriate to a given stem.

However, the use of endings causes several difficulties. First, not every final sequence of letters which looks like an ending actually is one. For example, the infinitives of most Russian verbs end in *-ть*. However, many noun forms also end in *-ть* where only *-ь* is the true ending, the *-т-* belonging to the true stem. To avoid this difficulty caused by the attempt to segment a word into true stem and true ending, our scheme deals only with the *apparent ending* and *apparent stem*; these need not be the true ending and the true stem. Our classification system takes care of all of the strange inflectional patterns which are caused by this convention.

Second, several different parts of speech may share the same apparent or true ending; *-ем* is a common ending of verbs, adjectives and nouns. But even within a part of speech, even in the inflection of a single canonical form, the same ending may have several uses. For example, the ending *-оѡ* in many adjectives may be used as:

nominative singular masculine (animate or inanimate);  
accusative singular masculine (inanimate);  
genitive, dative, instrumental or locative singular feminine (animate or inanimate).

The determination that these are exactly the possibilities is dependent on the classification scheme: If the stem of a text word is given as that of an adjective of class 4, then the ending *-оѡ* has exactly these possibilities. The list of uses is obtained by intersecting the information from the stem entry with the information from the ending tables. The determination that any one of the above uses is the correct use in a given source sentence is not a matter of morphology but rather of syntax and

semantics, and so belongs to another, more difficult part of the translating routine, that of syntactic integration, which is described in [7].

As a result of these considerations, our dictionary consists of two parts: a full-form dictionary (*list 1*), consisting of uninflectable words and word forms which cannot be easily put into our classification scheme; and a stem dictionary (*list 2*), consisting of stems and associated morphological class numbers. List 1 is divided into two parts, list 1a and list 1b, for easier lookup. Furthermore, there are ending tables, which list the usage patterns (by number) which may be associated with a given ending and class, and a surprisingly short list of these usage patterns. The dictionary also contains certain tables for use with source words not found in the dictionary.

## 2. The Glossary Lookup Process

The process by which our program identifies occurrences in text during a translation run consists of several steps. See [7] and [4] for a description of the syntactic integration scheme, which processes the occurrences once they have been identified.

First, a large amount of text is read into the machine, either by punched cards or by an automatic character-recognition device. Occurrences are separated from each other by means of a program which recognizes space marks, punctuation marks and other conventional clues. Occurrences which are obviously not words (such as punctuation marks, intermixed sequences of letters and digits, etc.) are indicated as such. The remaining occurrences, called *apparent words*, are first sorted (for efficiency) and then looked up individually in our short list 1a, which is in the high speed memory of the machine. At present we distinguish neither between accented and unaccented letters nor between e and ë. If an *apparent word* is not found in list 1a, then the *apparent ending*,<sup>1</sup> which need not be a true ending, is removed automatically from the word, and the remainder, the *apparent stem*, is broken up into *apparent prefix(es)*, *apparent root(s)* and *apparent suffix(es)*. (Hereafter, the word "apparent" is omitted except for emphasis.) As an example of the splitting procedure, a comparison with true segmentation for the text words *вращая* and *описать* appears below.

	<i>Apparent</i>	<i>True</i>	<i>Apparent and True</i>
Reflexive ending	(none)	(none)	(none)
Ending	-ая	-я	-ть
Prefix(es)	в-	(none)	о-
Suffix(es)	-ащ-	-а-	-а-
Root	-р-	-вращ-	-пис-

(The items are listed in order of discovery by our routine.)

The root (if there is only one<sup>2</sup>) is looked up in a catalog of roots kept in the high speed memory. This catalog gives the location in the low speed memory for the beginning of the set of glossary entries, if any, having that root. (These glossary entries include items in lists 1b and 2 intermingled.) If the root is found in this catalog, then the stem is looked up under the root. The entries for a given stem are

<sup>1</sup> See table of apparent endings in Appendix 1.

<sup>2</sup> We have a separate routine for recognizing words with multiple roots.

divided into two parts and stored consecutively: those entries with an explicit ending (list 1b), and those with no explicit ending (list 2). If the stem of the text word is found, then a search for the apparent ending is made through the section with an explicit ending, and then a search for the stem alone is made through the section with no explicit ending.

If the word is found in list 1a, or in list 1b alone, then all uses of the word are found with the glossary entry. If the stem of the word is found in list 2 only, then the morphological class information found with the stem entry will determine whether or not the ending found in the word is appropriate. All uses of the word are determined, by a process described later in this report, by intersecting the information from the stem with that from the ending.

If the word is found in both list 1b and list 2, then all its uses are obtained by uniting the information from list 1b with that from list 2.

If the word is not in any of the lists, then it is not in the glossary and is treated by a special routine (see Section 9). In any case, the lookup process is now complete for this word.

After all words have been looked up in the glossary, the items are sorted into text order and the syntactic integration routine is begun.

### 3. Inflectional Forms

To classify a particular word morphologically, it is necessary to know precisely how to inflect the word. Unfortunately there is no completely accurate source of information about inflection. This is due partly to the fact that there are no precise boundaries of language usage, so that a form may be admitted at one time but not at another or may be considered colloquial, bookish or slang by some authorities. The many reference works used by us often differ among themselves on the proper inflection or spelling of a word. In fact, the same source may even contradict itself! Our principal sources are [1, 2, 3, 8], and for verbs especially, [6]. In cases of disagreement, the judgment of the members of our staff, many of whom have been native speakers of Russian, is taken as final.

For each part of speech, a certain form is considered to be the *canonical form*, and certain forms are considered to *belong to* that canonical form.

*Verbs.* The canonical form is the infinitive (of a given aspect). Associated with an infinitive are the following one-word forms of the same aspect:

infinitive;

present indicative (for imperfective verbs), or future indicative (for perfective verbs), in three persons and two numbers;

past indicative, in three genders and two numbers;

imperative, in one person and two numbers;

gerund in -а or -я;

gerund in -в, -вши, or -ши.

The first person plural of the future indicative of perfective verbs is also used as an imperative.

*Participles and adjectives.* The canonical form is the nominative singular mascu-

line form of a given degree if it exists; otherwise the masculine short singular if it exists; or the neuter singular short form. Associated with a given canonical form are:

long forms of the same degree, in six cases, three genders, two numbers, and two animations;  
 short forms of the same degree, in three genders and two numbers;  
 the impersonal form of the same degree;<sup>3</sup>  
 the one-word comparative short form (only if the original degree is positive);  
 the one-word comparative impersonal form<sup>3</sup> (only if the original degree is positive);  
 a theoretical combining form.

*Adverbs.* The canonical form is the normal form for that degree. The only forms associated with an adverb of positive degree are that form itself and the corresponding one-word form of comparative degree. The only form associated with an adverb of comparative or superlative degree is that form itself.

*Nouns.* The canonical form is the nominative singular form if it exists; otherwise the canonical form is the nominative plural form. Associated with a noun having singular forms are its forms in six cases and two numbers (even if the noun is not customarily used in the plural), and a theoretical combining form. Belonging to a noun not having singular forms are its plural forms in six cases and a theoretical combining form. If a noun is declined like an adjective (such as производная or a family name ending in -ов, -ев, or -ский), then it is classified as an adjective, and its canonical form is the nominative singular masculine (unless the word is not used in the singular).

*Pronouns and cardinal numerals.* The canonical form is the first of the following forms which exists:

nominative singular masculine, nominative singular, nominative plural, nominative, genitive, etc.

Belonging to a given canonical form are the forms for all numbers, cases, genders and animations which exist.

*Prepositions, conjunctions, interjections, particles.* For these uninflectable words the canonical form is the form itself, and no other forms belong to it. Alternate spellings are considered as separate words.

*Hyphenated words, phrases, idioms, and abbreviations.* These are treated separately but are considered to have the forms appropriate to the part of speech they replace.

#### 4. Endings and Stems

The lookup procedure in a mechanical translation program determines somewhat the format of a stem dictionary (like our list 2) as well as the admissible endings and stems. If "true" endings are used, then the lookup program may have to attempt to split a word between stem and ending in several ways in order to find the proper dividing point; each split will have to be looked up separately. For example, the word form раздела could be split разде/ла, раздел/а or раздела/. The first two splits yield actual interpretations (from the verb раздеть and the noun разде); the third

<sup>3</sup> The impersonal form of an adjective is called by some Russian grammarians "предикативное наречие" [3, pp. 39, 630-632] or "категория состояния."

split is wasted. In many cases moreover, the correct splitting is a matter of taste, as in the infinitive вести.

Our program permits only *one* splitting. We thus have to use a set of *apparent endings* and *apparent stems*. Our splitting procedure follows.

First, a zero-order reflexive ending -еь or -ея, if any, is removed from a word which is to be identified. The reflexive ending need not be a true one, as in the word ось. We do not distinguish between the two reflexive ending forms -еь and -ея but merely note their presence or absence. This rarely causes confusion. See Section 6 for discussion. If a reflexive ending is removed, then the whole word and the eventually resulting stem are called reflexive, otherwise nonreflexive.

Next, a comparison is made to determine the longest possible match between an entry in our ending table (Appendix 1) and the nonreflexive remainder of the word. The largest match (which may be the null ending) is called the (*apparent*) ending of the word and the remainder of the word is the (*apparent*) stem.

To make our splitting as close to true splitting as possible, our apparent endings include, with few exceptions, all endings which occur in the true inflection of regularly inflected words. Thus -y and -ов come from noun declensions and -y and -ло from verbal conjugations. We have decided to consider the -и- of all forms of words like знание as part of the ending in order to simplify the classification. Of course other sets of apparent endings could have been chosen. These would yield different splittings in many cases and would require different classifications. The apparent ending table grew naturally and is not necessarily optimal. It is believed, however, that for a fixed set of canonical forms a different ending table would not significantly reduce the number of entries in the glossary. Increasing the number of apparent endings will probably increase the number of morphological classes, thus complicating the classification process. The system is adaptable to changes in Russian spelling. For example, prerevolutionary Russian spelling would require certain new endings and modifications of old ones such as -аго, -ья, -ие, -ихъ. The recently proposed spelling reform of Russian may also cause a slight modification (see [Известия 24-25 Sept. 1964]).

*Examples.* Here  $\emptyset$  is the null stem or ending, and a star indicates the presence of an apparent reflexive ending.

Word	Stem	Ending	Reflexive Ending
раздела	разде-	-ля	
разделась	разде-	-ля	*
дастся	даст-	$\emptyset$	*
носясь	нос-	-я	*
нося	н-	-о	*
ему	$\emptyset$	-ему	
ось	$\emptyset$	-о	*

The reflexive endings on *нося* and *ось* are not "true" endings.

##### 5. Criteria for Assignment to List 1 or 2

The criteria given below specify which word forms are to be put into list 1. All other forms are assigned to list 2. We put into list 1:

- (a) all words having an apparent but not true reflexive ending;

(b) all uninflectable words, such as prepositions, particles, interjections and some adverbs;

(c) all pronouns and numerals which are not declined like adjectives; also all personal pronouns however declined;

(d) all special -y or -ю genitives and locatives on nouns;

(e) all abbreviations, and all hyphenated prefixes and suffixes used in the formation of pronouns and indefinite words, such as т.е., -нибудь, кое-;

(f) all word forms coming from stems to which only one ending can be added during the full inflection of its canonical form, such as читая, which is the only form belonging to the canonical form читать and having the stem чит-, since all other forms of читать have the apparent stem чита-;

(g) all word forms homographic to word forms already in list 1a (these are put into list 1a), but word forms homographic to word forms in list 1b but not to any in list 1a are not put into list 1 merely for this reason;

(h) certain other very common words; these are put into list 1a.

Some interesting cases result from the application of these criteria. Suppose that the glossary includes all forms belonging to the canonical forms тот, тема and тема. The genitive plural тем of тема must be put into list 1, since it is the only form of тема with the apparent stem т-. (The word тем itself is also an adverb or conjunction, and so belongs in list 1 for that reason also.) If тем (from тема and the adverb or conjunction) should be put into list 1a, then the genitive plural тем of тема and the instrumental singular masculine and neuter and the dative plural тем of тот also belong with it in list 1a. Now most of the forms of тема have the stem тем-. Two forms, the accusative singular тему and the genitive plural тем, have the stem т-. If тем is in list 1a, then there is only one remaining form belonging to тема with stem т- which is not already in list 1, and so by rule (f) the form тему should also go into list 1 (either part). Note that if тем (from тема and the adverb or conjunction) should be put into list 1b, then the form тем from тот and the forms тем and тему from тема would be left in list 2, since they can be accounted for there.

## 6. Ending Arrangements for List 2

To each stem in list 2 we assign a classification which includes its part of speech, certain subsidiary information (such as gender and animation for nouns, aspects for verbs, degree for adjectives, etc.) and a morphological class number. This class number stands for the *arrangement* of endings which can be added to the stem to form all or part of the complete inflection of its canonical form.

As far as possible, the arrangements allow the reconstruction of just those forms which belong to the canonical form. The few exceptions are stated at the end of this section. Since there is often disagreement about the existence of a form, we use the information described in Section 3.

For example, the noun with canonical form ядро has two stems: ядр- and ядер-. The noun with canonical form число has three stems: числ-, чис-, and чисе-. The ending arrangements associated with these stems to give the complete declensions of these canonical forms are listed below together with the full declensions.

Case	Full Declension		ЯДР-		ЯДЕР-	
	Singular	Plural	Singular	Plural	Singular	Plural
Nominative	ядр/о	ядр/а	-о	-а		
Genitive	ядр/а	ядер/ѳ	-а			ѳ
Dative	ядр/у	ядр/ам	-у	-ам		
Accusative	ядр/о	ядр/а	-о	-а		
Instrumental	ядр/ом	ядр/ами	-ом	-ами		
Locative	ядр/е	ядр/ах	-е	-ах		
Combining Form	ядр/о-		-о-			

Case	Full Declension		ЧИСЛ-		ЧИС-		ЧИСЕ-	
	Singular	Plural	Singular	Plural	Singular	Plural	Singular	Plural
Nominative	чис/ло	чис/ла			-ло	-ла		
Genitive	чис/ла	чисе/л			-ла			-л
Dative	числ/у	числ/ам	-у	-ам				
Accusative	чис/ло	чис/ла			-ло	-ла		
Instrumental	числ/ом	числ/ами	-ом	-ами				
Locative	числ/е	числ/ах	-е	-ах				
Combining Form	числ/о-		-о-					

(The combining form is used in testing for compound words. It is split by a different process. The program for this has not been written.)

Blank spaces in the ending arrangements indicate the *absence* of a form with that stem and ending. The reason that both stems *чис-* and *числ-* are needed is that *-ло* and *-ла* are apparent endings in our list (since they are needed in the regular conjugation of verbs), but *-лу*, *-лом*, etc. are not apparent endings in our list (since they are not needed in any regular inflection).

According to the criteria of Section 5, the forms *ядер* and *чисел* belong in list 1, since their stems *ядер-* and *чисе-* allow only one ending to be added. The remaining stems *ядр-*, *числ-* and *чис-* are put into list 2. The stem *ядр-* has a "normal" declensional arrangement, which we code as Nvi5, which indicates a noun, neuter, inanimate, class 5. However, both *числ-* and *чис-* have "abnormal" patterns. The stem *чис-* is assigned to class Nvi19 which has exactly the proper ending arrangement. At first glance it would seem that a special class would have to be made up for *числ-*. However because of the character of the splitting process, it is possible to assign it to the same class as the stem *ядр-*. This is because the forms *число* and *числа* obtained by using the endings *-о* and *-а* with the stem *числ-* would be split in a different way. In general, if adding an ending to a stem produces a word form which our routine splits to yield a longer ending (and shorter stem), then it is irrelevant whether the class to which the first stem belongs can produce the given word form, since this word form will be considered as belonging to another stem.

The interpretation of the morphological class numbers depends on the part of speech to which a stem belongs.

*Verbs.* A morphological class number describes the same arrangement of endings for all stems having this class number. However, the interpretation of the ending depends on the aspect and reflexivity assigned to the stem. The present indicative of imperfective verbs corresponds to the future indicative of perfective verbs. The first person plural of the future indicative of perfective verbs is also an imperative form. The gerund in *-а* or *-я* is treated as a past gerund for perfective verbs. Re-



flexive verb forms are required to end in *-сь* or *-ся*; nonreflexive verb forms are forbidden to end thus; all exceptions are put into list 1.

*Participles and adjectives.* A morphological class number describes the same arrangement of endings, independently of degree, for each stem having this class number, except as modified by flags. These flags indicate which species of form are found with each stem; *long, short, impersonal, comparative short, comparative impersonal, combining*. Adjective stems differing only in this flag system may have the same morphological class number. Reflexivity is also indicated, since participles are classified in the same way as adjectives.

*Adverbs.* Those adverb stems which are placed in list 2 have morphological class numbers to indicate the endings allowed in the positive and comparative degrees.

*Nouns.* The morphological class numbers are assigned separately for the three genders, but classes differing only in *animation* are described by the same numbers. Thus a class number assigned to a masculine noun stem may have an ending arrangement which is entirely different from that of the identically numbered class for feminine or neuter nouns. Moreover the endings (or even the presence or absence of endings) for the accusative case may differ according to animation without changing the class number. A flag indicates the presence of a combining form. Nouns such as *производная*, which are declined like adjectives, have adjective classification numbers and are treated as adjectives; their use as a noun (of a specific gender) is indicated in other ways.

*Pronouns and cardinal numbers.* These are classified according to type. Those which follow the adjective declensional pattern are classified in the same way as adjectives. All others are put into list 1.

We cannot include all extant ending arrangements in our morphological classification system, since their number is enormous, and most of them are rather rare. However, all ending arrangements appearing for stems in our glossary are included, as well as several additional classes for stems likely to be added if the glossary size should be increased moderately. Our machine space requirements limit us to eight bits (256 possibilities) for class numbers for stems in list 2. This upper bound is more than adequate for present purposes, since no class number exceeds 109 at present. A complete list of our morphological classes is in Appendix 3, and several classification examples are given in Appendix 5.

As mentioned earlier, the morphological classes do not entirely agree with the existence of word forms. There are only three special cases to be noted.

1. The existence of certain forms is doubtful. While we can, and do, make up morphological class numbers for verb stems which are found only in the third person (or even third person singular), for adjectives which have no short forms and for verbs which have no present gerunds, we do assume that every noun, unless used in a special sense, can be declined in the plural, even though such use is unlikely for certain nouns. However, we do not deliberately misinflect words merely to put them into our classification scheme. For example, the noun *англичанин* has the stems *англичанин-* and *англичан-* in the singular and plural, respectively. These stems are classified as  $\mu\alpha 18$  and  $\mu\alpha 1$ , which give precisely the desired forms, and no more. Thus we do not have any trouble with the "artificial homography" which plagues certain other splitting systems, except as described in the next two paragraphs.

2. The routine for assigning reflexive category to a word form ending in *-сь* or *-ся*

(for list 1b and list 2 items) does not distinguish between these two endings. This is rarely serious, for while several examples of confusion caused by this identification of letter combinations can be found (such as *гучь* vs. *гучя*), none have yet been found in our prepared glossary. Any such forms which might be found in later additions to the glossary can be put into list 1a, where the true spelling is used, without any separation into stem, ending and reflexive ending.

3. As a result of using the same morphological class numbers for reflexive and nonreflexive verb stems, an "impossible" past gerund is often introduced for reflexive verbs. Reflexive verbs allow only *-шись*, *-вшись* or *-ившись* as past gerund ending. Thus, any morphological class which allows *-в*, *-ив*, *-ев*, *-ов* or *-нев* as a past gerund ending will be incorrect for reflexive verbs. It is not worthwhile to correct this error by means of programming or by assigning different classification numbers to reflexive verbs, since a word ending in *-всь* or *-всья* cannot be a true reflexive form of any word, and so, if it exists, will be put into list 1.

### 7. Usage Patterns for List 2 Endings

If the stem of a word is found in list 2, then its morphological class numbers must be compared with the ending split off from the word to determine the precise uses (our "temporary choices") of the word. This is done in the following way.

Corresponding to each ending is a section of the high speed memory of the computer. This section contains, in a certain arrangement, a list of those morphological classes for list 2 which admit this ending, together with the uses to which these classes put the ending. This section of the memory is divided up according to part of speech. The section devoted to nouns is further subdivided according to gender. Within each part of speech, or subdivision thereof, we list the different usage patterns (given by number), the modifications given by flags, and the morphological class numbers to which the usage pattern applies. The total number of usage patterns is quite small. See Appendix 4 for a list of them.

*Nouns.* A usage pattern lists the combinations of case and number which a noun may have with a given ending, if it belongs to a given morphological class. For convenience, the usage patterns are numbered independently for each part of speech. Thus noun usage pattern 20 is for an ending which with specified noun classes has exactly the uses: genitive singular, dative singular, locative singular, nominative plural and inanimate accusative plural. Note that the same usage pattern numbers are used for animate and inanimate nouns, and are independent of the gender of the noun. The animation of the noun will direct the search to the proper pattern.

*Verbs.* A usage pattern lists the combinations of person, number, gender, tense and mood which are appropriate to a given ending in a given set of classes. There are only three usage patterns which contain more than one use; of these, two are rare. The interpretation of the uses to which an ending may be put are often dependent on the aspect of the verb. Those usage patterns which are aspect-dependent are marked with an aspect-modifier flag, which is used to make the changes mentioned in Section 6.

*Adjectives* (including participles and certain pronouns and numerals). A usage pattern lists the combinations of case, number, gender and animation which the ending may have if it is used as a long form ending, and the combinations of gender

and number which the ending may have if it is used as a short form ending. The modifications given by flags indicate the species of form allowed to the ending. In determining the exact uses of an adjective form, only those uses which are common to the information from the stem and that from the modification flags of the ending may be used. Only a few usage patterns have more than one flag. Since the impersonal, comparative short and comparative impersonal forms are described completely by these descriptions alone, no separate pattern need be made to describe them, and the blank pattern will suffice. Because of the format in which the precise uses of a long form adjective are put during the running of our translation program, the same ending, with the same morphological class number, may give rise to several usage pattern numbers. This is because the case, number, gender and animation possibilities are stated separately rather than jointly. Thus the common ending *-и* is listed as belonging to usage patterns 1 and 4 for many morphological classes.

*Adverbs.* No numbering of usage patterns is necessary since there are only a few patterns which are very simple and depend only on the degree assigned to the ending. The actual uses of the ending for the classes listed with it are given. Actually, there is only one possibility: positive and comparative degree. However, because in an earlier version of the classification system many adverbs with only one form having a given stem were put into list 2, there are now a few superfluous classes which have a pattern specifying only one usage.

The numbers assigned to the morphological classes have not been assigned arbitrarily but have been assigned in such a way that the lookup process is considerably simplified. As far as possible, morphological classes differing only slightly have been given consecutive numbers. Thus the morphological classes associated with the same usage pattern (for a given ending) may be given in the form of a few "from-to" sets, rather than as an apparently random sequence of numbers. Thus the ending *-ом* for masculine nouns has the noun usage pattern 10 in classes 5 to 11, and classes 14 to 18 (all intervals "inclusive").

## 8. Usage Patterns for List 1 Words

To each item (i.e., full word) in list 1 we assign a classification which differs formally from that assigned to list 2 stems only in that the morphological class number is replaced by the usage pattern. Thus the complete information about the usage of the word is stored with it. At present, the actual usage pattern, rather than a number referring to it, is used. Since in our files we have classified by pattern number all list 1 words belonging to inflectable parts of speech, a list of these patterns is given at the end of Appendix 3. They are distinguished from other class and pattern numbers by the presence of the letter X.

## 9. Nondictionary Items

If a word is not found in the glossary during the lookup process, then we may assign to it a set of usages by looking at its ending and final suffix. A separate "unglossary" is searched to see if it has an entry for the given final suffix (or no suffix) and ending, in just the same way that lists 1b and 2 are searched. For example, if the word *агребранческим* occurs in text, it will not be found in the glossary, which, with few exceptions, does not include words in the international vocabulary which

can be translated merely by transliterating a portion of the word and adding the proper termination (and possibly prefix). However, this word has the ending *-им* and the suffix *-чек*. A portion of the unglossary devoted to words with the suffix *-чек* will have an entry (as in list 2) for an adjective of class 9, long forms only. The ending *-им* is appropriate to this class and flag system. Hence the word will be found in the unglossary. Similarly, the adverb *алгебраически* will be found in the section of the unglossary devoted to the suffix *-чек*, but this time it will be found with the actual ending *-и* (as in list 1b).

If a word is found in the unglossary, then it is treated from here on, in nearly all particulars, like a word found in the glossary. Of course, a flag is used to indicate that the item comes from the unglossary, and the word is tentatively translated by using the stem of the word, less final suffix, plus a termination which is determined by the unglossary entry. For example, the adjectives found with the suffix *-чек* use the suffix *-с*, and the adverbs with the suffix *-чек* use the suffix *-cally*. Thus the translations of *алгебраическим* and *алгебраически* would be "algebrai-c" and "algebrai-cally" respectively.

Of course, the use of the unglossary is not foolproof. For example, the noun with canonical form *том* should be put into the regular glossary if it is expected to be found in text, since its forms *том* and *тому* are identical with forms of the pronominal adjective *тот*, which is essential in any glossary of Russian. As another example, the indeclinable noun *кентуры* will probably be treated, like *стрелуры*, as an accusative singular feminine form of a noun ending in *-ура*, and so will not be treated properly. The hardest items, however, will be those of Slavic or mixed origin, since these will not make much sense if transliterated and may have arrangements of endings which are not like those of international words.

We have not yet decided how to treat items which are in neither the glossary nor the unglossary.

### 10. Implementation of the Classification System

At the present time, we have several sample glossaries in various stages of completion.

First, we have on cards (3 x 5) a morphological classification of the entire American Mathematical Society's *Vocabulary* [5]. In fact many additional words related to those actually listed in the *Vocabulary* are in our files. We have about 3900 stems and 1200 list 1 items. Actually, our entire classification system was based on the items found in the *Vocabulary*, and so is incomplete, especially for nouns; the verb and adjective classifications contain virtually every allowable arrangement of endings. There is no unglossary for this glossary.

Second, we have on coding sheets a glossary containing every item found in the text [9]. This glossary is ready to be put into operation with the translation program as soon as certain minor changes are made in the latter. This glossary distinguishes between lists 1a and 1b as follows: Words in list 1 which have fewer than seven letters (and so fit into one machine word of the IBM 704, etc.) are put into list 1a; all others are put into list 1b. Furthermore, all adverbs having the same stem as a list 2 adjective are put into list 2 (instead of 1b) in many cases; they may have classification numbers which yield only one ending. The unglossary has several entries for list 2 only.

Third, we have a glossary consisting of the items found in the first two sentences of [9]. This glossary is operating with the machine translation program as presently constituted. The differences between this glossary and the one described in this report do not involve morphological classifications, but rather other information needed by the program. This glossary has no words in list 1b. There is a short unglossary for list 2 items only.

*Acknowledgments.* Many people have participated in the morphological classification process and in the preparation of this report. The following have been of especial help: Luba Ross, Renée Robinson, Valentina Monroe,<sup>4</sup> Owen McArdle, and Barbara Cummins. Extremely important has been the help of Ida Rhodes, who explained to me the requirements of the mechanical translation code, so that the morphological classification process could be made efficient, and on whose ideas the mechanical translation scheme developed at the National Bureau of Standards is based.

<sup>4</sup> Deceased.

#### APPENDIX 1. *Table of Apparent Endings*

Reflexive endings:	(2)
-сь -ся	
Four-letter endings:	(2)
-ивши -иями	
Three-letter endings:	(25)
-ами -вши -его -ему -ете -ешь -ев -ей -ней -нем -нею -ила -или -ило -ими -ите -ить -ишь	
-иям -иях -йте -ого -ому -ыми -ьте -ями	
Two-letter endings:	(43)
-ам -ат -ах -ая -ев -ее -ей -ем -ет -ею -ив -ие -ии -ий -ил -им -ит -их -ию -ия -ла -ли -ло	
-ов -ое -ой -ом -ою -ть -ут -ую -ши -ые -ый -ым -ых -ью -ют -юю -ям -ят -ях -яя	
One-letter endings:	(12)
-а -в -с -и -й -л -о -у -ы -ь -ю -я	
Zero-letter ending:	(1)
∅	

The above table lists 83 nonreflexive endings.

#### APPENDIX 2. *Abbreviations Used in Later Appendices*

a	accusative	K	comparative imper-	r	reflexive
C	comparative short		sonal	S	short
c	comparative	L	long	s	singular; superlative
cf	combining form	l	locative	V	verb
D	adverb	N	noun	X	list 1 pattern number
d	dative	n	nominative; nonre-	μ	masculine
g	genitive; gerund		flexive	ν	neuter
I	impersonal	p	plural; perfective;	φ	feminine
i	instrumental; inani-		positive	1	first (person)
	mate; imperfective	pf	present or future	2	second (person)
inf	infinitive	pr	present	3	third (person)
ip	imperative	ps	past	*	reflexive stem
J	adjective				



## B. List 2 Feminine Nouns

Class	1	2	3	4	5	6	7	8	9	10
ns			а	а	а	а	а	а	а	а
gs			ы	ы	ы	ы	ы	и	и	и
ds			е	е	е	е	е	е	е	е
as			у	у	у	у	у	у	у	у
is			ой	ой	ой	ой	ей	ей	ой	ой
ls			ою	ою	ою	ою	ею	ею	ою	ою
			е	е	е	е	е	е	е	е
np	и	ы	ы	ы	ы	ы	ы	и	и	и
gp		ø			ø	ø	ø	ø	ø	ø
dp	ам	ам	ам	ам	ам	ам	ам	ам	ам	ам
ap	и	ы	ы	ы	ы	ы	ы	и	и	и
apa		ø			ø	ø	ø	ø	ø	ø
ip	ами	ами	ами	ами	ами	ами	ами	ами	ами	ами
lp	ах	ах	ах	ах	ах	ах	ах	ах	ах	ах
cf			е-	о-	о-	е-	е-	е-	е-	о-
Class	11	12	13	14	15	16	17	18	19	20
ns	а	и	я	я	я				ь	ь
gs	и	и	и	и	и				и	и
ds	е	е	е	е	е	и	и	и	и	и
as	у	ю	ю	ю	ю				ь	ь
is	ой	ей	ей	ей	ей			ью	ью	ью
ls	ою	ею	ею	сю	сю	и	и	и	и	и
	е	е	е	е	е					
np	и	и	и	и	и	и	и	и	и	и
gp		й		ь	ей	ей	ей	ей	ей	ей
dp	ам	ям	ям	ям	ям	ям	ям	ям	ям	ям
ap	и	и	и	и	и	и	и	и	и	и
apa		й		ь	ей	ей	ей	ей	ей	ей
ip	ами	ями	ями	ями	ями	ями	ами	ями	ями	ями
lp	ах	ях	ях	ях	ях	ях	ах	ях	ях	ях
cf	о-	е-	е-	е-	е-	е-	е-	е-	е-	о-
Class	21	22	23	24	25	26	27	28	29	
ns	ь	ь	ия	ля					и.ля	
gs	и		ии							
ds	и		ии		ете			ее		
as	ь	ь	ию			ему	ью	ею		
is	ью	ью	ией							
ls	и		иею							
			ии		ете			ее		
np	и		ии							
gp	ей		ий	л	ет	ем	ей	ей	ил	
dp	ам		иям							
ap	и		ии							
apa	ей		ий	л	ет	ем	ей	ей	ил	
ip	ами		иями							
lp	ах		иях							
cf	е-		ие-							

## C. List 2 Neuter Nouns

Class	1	2	3	4	5	6	7	8	9	10	11	12
ns	о	о	о	о	о	о	о	о		е	е	
gs	я	я	я	я	я	я	я	я	и	я	я	
ds	у	у	у	у	у	у	у	у	и	у	у	
as	о	о	о	о	о	о	о	о		е	е	
is	ом	ом	ом	ом	ом	ом	ом	ом	см	ем	ем	
ls	е	е	е	е	е	е	е	е	и	е	е	
np		и	и	я	я	я	я	я	я	я	я	и
gp		й	ов	ов			й	й	й	й		ей
dp		ам	ам	ам	ам	ам	ам	ам	ам	ам	ам	ам
api		и	и	я	я	я	я	я	я	я	я	и
apa		й	ов	ов			й	й	й	й		ей
ip		ами	ами	ами	ами	ами	ами	ами	ами	ами	ами	ами
lp		ах	ах	ах	ах	ах	ах	ах	ах	ах	ах	ах
cf	о-	о-	о-	о-	о-	е-	е-	о-		е-	е-	е-

  

Class	13	14	15	16	17	18	19	20	21	22	23
ns		е	е	е		ие	лю	лю	нлю		
gs		я	я	я		ия	ля	ля	иля		
ds		ю	ю	ю		ию					бю
as		е	е	е		ие	лю	лю	нлю		
is		см	см	см		нем					
ls		е	е	е		ин				ете	
np	и	л	л	я	я	ил	ля	ля	иля		
gp	ей	ей		ев	ев	ий		л	ил	ет	ей
dp	ям	ям	ям	ям	ям	иям					
api	и	я	я	я	я	ия	ля	ля	иля	ет	ей
apa	ей	ей		ев	ев	ий		л	ил	ет	ей
ip	ями	ями	ями	ями	ями	иями					
lp	ях	ях	ях	ях	ях	иях					
cf	е-	е-	е-	е-	е-	ие-					

## D. List 1 Noun Patterns

Class	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23
ns	x	x	x	x											x								
gs					x	x	x	x					x		x	x	x		x	x			x
ds														x	x	x	x						
as		x	x		x		x								x								
is															x	x	x			x	x	x	x
ls															x	x	x						
np					x	x			x	x		x			x	x	x						
gp			x	x							x				x								
dp						x				x		x			x								
ap				x											x			x					
ip															x								
lp															x								
cf															x								





E. List 2 Adjectives, *continued*

Class	13	14	15	16	17	18	19	20	21	22	23	24
<i>L<sub>μ</sub> ns</i>			ий	ий	ий	ий	ий	ий	ий	ий		
gs	ого	его	его	его	его	его	его	его	его	его	его	его
ds	ому	ому	ему	ему	ему	ему	ему	ему	ему	ему	ему	ему
asi			ий	ий	ий	ий	ий	ий	ий	ий		
asa	ого	его	его	его	его	его	его	его	его	его	его	его
is	ем	ем	им	им	им	им	им	им	им	им	им	им
ls	ом	ем	ем	ем	ем	ем	ем	ем	ем	ем	ем	ем
<i>L<sub>φ</sub> ns</i>	а		ая	ая	ая	ая	ая	ая	ая	ая	я	а
gs	ой	ей	ей	ей	ей	ей	ей	ей	ей	ей	ей	ей
ds	ой	ей	ей	ей	ей	ей	ей	ей	ей	ей	ей	ей
as	у	ю	ую	ую	ую	ую	ую	ую	ую	ую		у
is	ой	ей	ей	ей	ей	ей	ей	ей	ей	ей	ей	ей
ls	ою	ею	ею	ею	ею	ею	ею	ею	ею	ею	ею	ею
	ой	ей	ей	ей	ей	ей	ей	ей	ей	ей	ей	ей
<i>L<sub>ν</sub> ns</i>	о	е	ее	ее	ее	ее	ее	ее	ее	ее	е	е
gs	ого	его	его	его	его	его	его	его	его	его	его	его
ds	ому	ему	ему	ему	ему	ему	ему	ему	ему	ему	ему	ему
as	о	е	ее	ее	ее	ее	ее	ее	ее	ее	е	е
is	ем	ем	им	им	им	им	им	им	им	им	им	им
ls	ом	ем	ем	ем	ем	ем	ем	ем	ем	ем	ем	ем
<i>L np</i>	е	е	ие	ие	ие	ие	ие	ие	ие	ие	и	
gp			их	их	их	их	их	их	их	их	их	их
dp	ем	ем	им	им	им	им	им	им	им	им	им	им
api	е	е	ие	ие	ие	ие	ие	ие	ие	ие	и	
apa			их	их	их	их	их	их	их	их	их	их
ip			ими	ими	ими	ими	ими	ими	ими	ими	ими	ими
lp			их	их	их	их	их	их	их	их	их	их
<i>S<sub>μ</sub> s</i>			ø	ø	ø	е			ь			
φ s			а	а	а	е	я	я	я	а		
ν s			о	о	е	е	е	е	е	е,о		
p			и	и	и	е	и	и	и	и,ы		
<i>I</i>			о	о	е	е	о	е	е	о		
<i>C</i>			ее	ее	ее	ее	ее	ее	ее	ее		
<i>K</i>			ее	ес	ее	ее	ее	ее	ее	ее		
<i>cf</i>		е-	о-	е-	е-	е-	о-	е-	е-	е-,о-	е-	

E. List 2 Adjectives, *continued*

Class	25	26	27	28	29	30	31	32	33	34	35	36	37
L <sub>μ</sub> ns	ей	ый	Ø	Ø	Ø	Ø	ий	ей	ой				ам
gs	его	его	а	а,ого	а	а							
ds	ему	ему	у	у,ому	у	у							
asi	ей	ый	Ø	Ø	Ø	Ø	ий	ей	ой				ам
asa	его	его	а	а,ого	а	а							
is	им	ым	ым	ым	ым	ом							
ls	ем	ем	ом	ом	е	е							
L <sub>φ</sub> ns	ия	ая	а	а	а	а							
gs	ей	ей	ой	ой	ой	ой							
ds	ей	ей	ой	ой	ой	ой							
as	ию	ую	у	у	у	у	ью	ью	ою				
is	ей	ей	ой	ой	ой	ой							
ls	ею	ею	ою	ою	ою	ою							
	сй	ей	ой	ой	ой	ой							
L <sub>ν</sub> ns	ие	ее	о	о	о	о			ое				
gs	его	его	а	а,ого	а	а							
ds	ему	ему	у	у,ому	у	у							
as	ие	ее	о	о	о	о			ое				
is	им	ым	ым	ым	ым	ом							
ls	ем	ем	ом	ом	е	е							
L <sub>π</sub> np	ии	ые	ы	ы	ы	ы							ами
gp	их	ых	ых	ых	ых	ых							
dp	им	ым	ым	ым	ым	ым							
api	ии	ые	ы	ы	ы	ы							ами
apa	их	ых	ых	ых	ых	ых							
ip	ими	ыми	ыми	ыми	ыми	ыми							
lp	их	ых	ых	ых	ых	ых							
S <sub>μ</sub> s		Ø											
φ s		а									ла	ла	ила
ν s		е									ло	ло	ило
p		ы											
I		е									ло	ло	ило
C		ее											
K		ее											
cf		е-	о-	о-	о-	о-							



## G. List 2 Adverbs

Class	1*	2	3	4	5	6*	7*	8*	9*
(p)	о	о	о,е		е	е	и	ло	ило
c		ее	ее	ее	ее				

Notes. (p). The degree is actually specified by the stem and is not altered by using this form. Thus the stem *хорош-* is listed as positive. The form *хорошо* is a class 1 adverb and is positive. The stem *лучш-* is listed as comparative. The form *лучше* is a class 6 adverb and is comparative. The stem *наилучш-* is listed as superlative. The form *наилучше* is a class 6 adverb and is superlative. The stem *бедн-* is for a class 2 positive adverb. The form *беднее* comes from this stem and is comparative.

\* These classes admit only one form and are relics of a previous classification system for adverbs. The forms *хорошо*, *лучше* and *наилучше* mentioned in the previous note would now be listed in list 1b.

There are no list 1 patterns for adverbs. The degree (positive, comparative or superlative) is listed directly by number.

## H. List 2 Verbs

Class	1	2	3	4	5	6	7	8	9	10	11	12
inf	ь	ь						ть	ть			
pf 1s		ешь	ешь	ешь	ю	ю	ю	ю	ю	ю	ю	у
2s		ет	ет	ет	ет	ет	ет	ет	ет	ет	ет	ешь
3s	ет	ет	ет	ет	ет	ет	ет	ет	ет	ет	ет	ет
1p		ем	ем	ем	ем	ем	ем	ем	ем	ем	ем	ем
2p		ете	ете	ете	ете	ете	ете	ете	ете	ете	ете	ете
3p					ют	ют	ют	ют	ют	ют	ют	ут
ip 2s						й	й	й	й	и	и	и
2p						йте	йте	йте	йте	ите	ите	ите
ps $\mu$ s								ли	ли			
$\phi$ s								ля	ля			
$\nu$ s								лю	лю			
p								ли	ли			
pr g		а					я	я			я	я
ps g							в	в	в			
							вши	вши	вши			
Class	13	14	15	16	17	18	19	20	21	22	23	24
inf		ить				и						
pf 1s	у	у	у	у	у	у	у	у	у	у	у	у
2s	ешь	ешь	ешь	ешь	ешь	ешь	ешь	ешь	ешь	ешь	ешь	
3s	ет	ет	ет	ет	ет	ет	ет	ет	ет	ет	ет	
1p	ем	ем	ем	ем	ем	ем	ем	ем	ем	ем	ем	
2p	ете	ете	ете	ете	ете	ете	ете	ете	ете	ете	ете	
3p	ут	ут	ут	ут	ут	ут	ут	ут	ут	ут	ут	ут
ip 2s	и	и	и	и	и	и	и	и	ь	ь		ь
2p	ите	ите	ите	ите	ите	ите	ите	ите	ьте	ьте		
ps $\mu$ s		ь	ь	ь								
$\phi$ s		ля	ля	ля								
$\nu$ s		лю	лю	лю								
p		ли	ли	ли								
pr g	я	я	я					а	а			
ps g	ши	ши	ши	ши	ши							



H. List 2 Verbs, *continued*

Class	59	60	61	62	63	64	65	66	67	68
inf	ить	ить	ить	ить			ить	ить	ить	ить
pf 1s		ю	ю	ю	ю	ю	ю	ю		
2s	ишь	ишь	ишь	ишь	ишь	ишь	ишь	ишь	ишь	ишь
3s	ит	ит	ит	ит	ит	ит	ит	ит	ит	ит
1p	им	им	им	им	им	им	им	им	им	им
2p	ите	ите	ите	ите	ите	ите	ите	ите	ите	ите
3p	ят	ят	ят	ят	ят	ят	ят	ят	ят	ят
ip 2s	ь	ь		й	й	й	й			
2p	ьте	ьте		йте	йте	йте	йте			
ps $\mu$ s	ил	ил	ил	ил			ил	ил	ил	ил
$\phi$ s	ила	ила	ила	ила			ила	ила	ила	ила
$\nu$ s	ило	ило	ило	ило			ило	ило	ило	ило
p	или	или	или	или			или	или	или	или
pr g						я	я	я	я	
ps g	ив	ив	ив	ив			ив	ив	ив	ив
	ивши	ивши	ивши	ивши			ивши	ивши	ивши	ивши

Class	69	70	71	72	73	74	75	76	77	78	79	80
inf	ить											ить
pf 1s	ю	ю	ю							у	у	у
2s	ишь	ишь	ишь	ишь			ишь	ишь	ишь	ишь	ишь	ишь
3s	ит	ит	ит	ит			ит	ит	ит	ит	ит	ит
1p	им	им	им	им	им	им	им	им	им	им	им	им
2p	ите	ите	ите	ите	ите	ите	ите	ите	ите	ите	ите	ите
3p	ят	ят	ят	ят	ят	ят	ят		ат	ат	ат	
ip 2s	и	и	и	и	и						и	и
2p	ите	ите	ите	ите	ите						ите	ите
ps $\mu$ s	ил											ил
$\phi$ s	ила											ила
$\nu$ s	ило											ило
p	или											или
pr g									а	а	а	а
ps g	ив		ев	ев	ев	ев	ев					ив
	ивши											ивши

Class	81	82	83	84	85	86	87	88	89
inf	ить	ить	ить	ить	ить				
pf 1s	у	у	у	у	у	у			
2s	ишь	ишь	ишь	ишь	ишь	ишь			
3s	ит	ит	ит	ит	ит	ит	ит	ит	ит
1p	им	им	им	им	им	им			
2p	ите	ите	ите	ите	ите	ите			
3p	ат	ат	ат	ат	ат	ат	ат	ат	ат
ip 2s	и,ь	ь	ь	и,ь	и	и			
2p	ите,ьте	ьте	ьте	ите,ьте	ите	ите			
ps $\mu$ s	ил	ил	ил	ил	ил				
$\phi$ s	ила	ила	ила	ила	ила				
$\nu$ s	ило	ило	ило	ило	ило				
p	или	или	или	или	или				
pr g	а	а						а	
ps g	ив	ив	ив	ив	ив				
	ивши	ивши	ивши	ивши	ивши				

H. List 2 Verbs, *continued*

Class	90	91	92	93	94	95	96	97	98	99
inf	ить	ить	ить	ить	ить	ить	ить	ить		
pf 1s						ью	ью		ею	сю
2s										
3s	ит	ит	ит	ит						
1p										
2p										
3p	ят	ят	ят	ят						
ip 2s						ей	ей	ей	ей	ей
2p										
ps $\mu$ s	ил	ил	ил	ил	ил	ил	ил	ил		
$\phi$ s	ила	ила	ила	ила	ила	ила	ила	ила		
$\nu$ s	ило	ило	ило	ило	ило	ило	ило	ило		
p	или	или	или	или	или	или	или	или		
pr g	а		я			ия				
ps g	ив	ив	ив	ив	ив	ив	ив	ив		ев
	ивши	ивши	ивши	ивши	ивши	ивши	ивши	ивши		
Class	100	101	102	103	104	105	106	107	108	109
inf					ть	ть	ть			и
pf 1s	ем	ою	ою	у					ю	
2s	ешь			ешь					ешь	
3s				ет					ет	ет
1p									ем	
2p									ете	
3p									ют	ут
ip 2s	ешь	ой	ой						и	
2p									ите	
ps $\mu$ s								ø		
$\phi$ s								ла		
$\nu$ s						лю	лю	лю		
p								ли		
pr g										
ps g	ев	ев			в	в			ов	
					вши	вши				

## I. List 1 Verb Patterns

Class	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
inf								x			
pr 1s	x										
2s		x									
3s			x	x							
1p											
2p				x	x						
3p											
ip 2s						x					
2p							x				
ps $\mu$ s									x		
$\phi$ s											
$\nu$ s											
p											
pr g										x	
ps g											x







APPENDIX 5. *Examples of Classification*

## A. Nouns

See Section 6 for the classification of the sample nouns with canonical forms ядро and число.

## B. Verbs

Canonical form: БЫТЬ						
Forms	Stems:	БЫ-	БУД-	ЕС-	СУ-	БУДУЧ-
inf	быть	ть				
pf 1s	буду		у			
2s	будешь		ешь			
3s	будет; есть		ет	ть		
1p	будем		ем			
2p	будете		ете			
3p	будут; есть, суть		ут	ть	ть	
ip 2s	будь		ь			
2p	будьте		ьте			
ps $\mu$ s	был	л				
$\phi$ s	была	ла				
$\nu$ s	было	ло				
p	были	ли				
pr g	будучи					и
ps g	быв	в				
	бывши	вши				
Classes:		Vpn 35	Vpn 22	Vin X4	Vin X5	Vin X10

Note: The forms of быть are distributed between the aspects purely for convenience in classifying. Since the future forms of быть are simple, we assign all nonpresent forms to the perfective aspect and all present forms to the imperfective aspect. There is no need for other coding to take care of the anomalous present forms.

## C. Adverbs

Canonical form: БЫСТРО  
 Forms { positive: быстро  
           comparative: быстрее  
 Stem: БЫСТР-  
 Class: Dp2

D. Adjectives

Canonical form: ПОЛНЫЙ			
Forms	Stems:	ПОЛН-	ПОЛОН-
$I_{\mu}$ ns	полный	ый	
gs	полного	ого	
ds	полному	ому	
asi	полный	ый	
asa	полного	ого	
is	полным	ым	
ls	полном	ом	
$I_{\phi}$ ns	полная	ая	
gs	полной	ой	
ds	полной	ой	
as	полную	ую	
is	полной	ой	
ls	полною	ою	
	полной	ой	
$I_{\nu}$ ns	полное	ое	
gs	полного	ого	
ds	полному	ому	
as	полное	ое	
is	полным	ым	
ls	полном	ом	
L np	полные	ые	
gp	полных	ых	
dp	полным	ым	
api	полные	ые	
apa	полных	ых	
ip	полными	ыми	
lp	полных	ых	
$S_{\mu}$ s	полон		ø
$\phi$ s	полна	а	
$\nu$ s	полно	о	
p	полны	ы	
I			
C	полнее	ее	
K			
cf	полно-	о-	
Classes		JpSLC 1	Jp XI

D. Adjectives, *continued*

Canonical form: BECB						
Forms	Stems:	B-	BC-	ø	BCEX-	BCEM-
I <sub>μ</sub> ns	весь	e*				
gs	всего		его			
ds	всему		ему			
asi	весь	e*				
asa	всего		его			
is	всем		ем			
ls	всем		ем			
I <sub>φ</sub> ns	вся			в*		
gs	всей		ей			
ds	всей		ей			
as	всю		ю			
is	всей		ей			
	всею		ею			
ls	всей		ей			
L <sub>ν</sub> ns	все		е			
gs	всего		его			
ds	всему		ему			
as	все		е			
is	всем		ем			
ls	всем		ем			
L np	все		е		ø	
gp	всех					
dp	всему		ем			
api	все		е			
apa	всех				ø	
ip	всем					и
lp	всех				ø	
S, I, C and K forms do not exist.						
cf	все-		е-			
Classes:		Jp X4* and X5*	JpL 14	Jp X9*	Jp X6 and X7	Jp X8

\* The starred entries belong in list 1a because they have an apparent reflexive entry which is not a true one.

Several entries in list 1 require two classes because no single pattern can accommodate all of their uses.

## F. Examples of Classification for Each Class

List 2 Masculine Nouns			List 2 Feminine Nouns	
Stem Example	Canonical Form	Class	Stem Example	Canonical Form
англичан-	англичанин	1	сутк-	сутки
телят-	теленок	2	ножниц-	ножницы
вес-	весы	3	овц-	овца
зародыш-	зародыш	4	судьб-	судьба
луч-	луч	5	труб-	труба
признак-	признак	6	мац-	маца
доктор-	доктор	7	таблиц-	таблица
том-	том	8	задач-	задача
закон-	закон	9	меж-	межа
кузнец-	кузнец	10	книг-	книга
раз-	раз	11	клетк-	клетка
месяц-	месяц	12	стру-	струя
палец-	палец	13	петл-	петля
отец-	отец	14	недел-	неделя
угл-	угол	15	дол-	доля
кусек-	кусок	16	любв-	любовь
теленк-	теленок	17	рж-	рожь
клин-	клин (s)	18	матер-	мать
муж-	муж	19	дроб-	дробь
пут-	путь	20	чест-	честь
будн-	будни	21	реч-	речь
клинъ-	клин (p)	22	любов-	любовь
другъ-	друг	23	линь-	линия
учител-	учитель	24	иг-	игла
вихр-	вихрь	25	пом-	помета
корн-	корень	26	т-	тема
случа-	случай	27	стат-	статья
кра-	край	28	ид-	идея
ручь-	ручей	29	с-	сила
княз-	князь	30		
критер-	критерий	31		
мужчин-	мужчина	32		
дедушк-	дедушка	33		
юнош-	юноша	34		
дяд-	дядя	35		
судь-	судья	36		
отв-	ответ	37		
изл-	излом	38		
объ-	объем	39		
преде-	предел	40		
сло-	слово	41		
к-	кит	42		
руч-	ручей	43		
муз-	музей	44		
суд-	судья	45		

E. Examples of Classification For Each Class, *continued*

List 2 Neuter Nouns			List 1 Nouns		
Class	Stem Example	Canonical Form	Class	Full Word Example	Canonical Form
1	звен-	звено	X1	отец/ѳ	отец
2	яблок-	яблоко	X2	уго/л	угол
3	очк-	очко	X3	ботинок/ѳ	ботинок
4	облак-	облако	X4	солд/ат	солдат
5	ядр-	ядро	X5	перепе/за	перепел
6	кольц-	кольцо	X6	кр/ая	край
7	лиц-	лицо	X7	ор/ла	орел
8	мест-	место	X8	случ/ал	случай
9	времен-	время	X9	искате/ли	искатель
10	кладбищ-	кладбище	X10	делите/ли	делитель
11	сердц-	сердце	X11	бук/в	буква
12	уш-	ухо	X12	друз/ей	друг
13	колен-	колени (р)	X13	юно/ши	юноша
14	пол-	поле	X14	при/ему	прием
15	ружь-	ружье	X15	анть/е	антье
16	плать-	платье	X16	ѳ/вши	вошь
17	звень-	звено	X17	це/ли	цель
18	движен-	движение	X18	стр/ую	струй
19	чис-	число	X19	ча/ю*	чай
20	те-	тело	X20	вид/у*	вид
21	прав-	правило	X21	кра/ю*	край
22	реш-	решето	X22	кра/ю	край
23	руж-	ружье	X23	вид/у	вид

\* Older classification system only. The dative form is in list 2 in the newer system, unless the whole word form is put into list 1a.

E. Examples of Classification For Each Class, *continued*

List 2 Adjectives					
Class	Stem Example	Canonical Form	Class	Stem Example	Canonical Form
1	известн-	известный	19	поздн-	поздний
2	быстр-	быстрый	20	много-	много-
3	ветрян-	ветряной, ветряной		-сторонн-	-сторонний
4	прям-	прямой	21	син-	синий
5	прикладн-	прикладной	22	искренн-	искренний
6	больш-	большой	23	мо-, чь-, треть-	мой, чей, третий
7	плох-	плохой	24	наш-	наш
8	заводск-	заводский, заводской	25	с-	сей
9	высок-	высокий	26	куц-	куцый
10	кратк-	краткий	27	ньютон-ов-	ньютон-ов
11	эт-, одн-	этот, один	28	сестрин-	сестрин
12	сам-	сам	29	Пушкин-	Пушкин
13	т-	тот	30	Псков-	Псков
14	вс-	весь	31	трет-	третий
15	хорош-	хороший	32	ч-	чей
16	свеж-	свежий	33	м-	мой
17	общ-	общий	34	теп-	теплый
18	больш-, лучш-	большой, лучший	35	тяже-	тяжелый
			36	м-	милый
			37	с-	сам

List 1 Adjectives			List 2 Adverbs		
Class	Full Word Example	Canonical Form	Class	Stem Example	Canonical Form
X1	полон/∅	полный	1	хорош-	хорошо
X2	легч/е	легкий (с)	2	быстр-	быстро
X3	пол/ого	пологий	3	искренн-	искренно, искренне
X4	тот/∅	тот	4	бол-	более
X5	тот/∅	тот	5	куц-	куце
X6	тех/∅	тот	6	лучш-	лучше (с)
X7	тех/∅	тот	7	коническ-	конически
X8	тем/и	тот	8	теп-	тепло
X9	∅/в/ся	весь	9	м-	мало
X10	хоро/ши	хороший			
X11	на/ши	наш			
X12	на/ши	наш			



List 2 Verbs			List 2 Verbs <i>continued</i>		
Class	Stem Example	Canonical Form	Class	Stem Example	Canonical Form
1	вытеч-	вытечь	64	сто-	стоять
2	влеч-	влечь	65	стро-	строить
3	береж-	беречь	66	сто-	стоять
4	моч-	мочь	67		
5	да-	давать	68	понадоб-	понадобить
6	осну-	основать	69	дополн-	дополнить
7	действиу-	действовать	70	погон-	погнать
8	влад-	владеть	71	обозр-	обозреть
9	сдела-	сделать	72	потерп-	потерпеть
10	поколебл-*	поколебаться	73	захот-	захотеть
11	колебл-*	колебаться	74	хот-	хотеть
12	бер-	брать	75	завид-	завидеть
13	вед-	вести	76	избеж-	избежать
14	ошиб-	ошибиться	77	беж-	бежать
15	вез-	везти	78	слыш-	слышать
16	внес-	внести	79	содерж-	содержать
17	введ-	ввести	80	служ-	служить
18	раст-	расти	81	корч-	корчить
19	верн-	вернуть	82	множ-	множить
20	ищ-	искать	83	знач-	значить
21	реж-	резать	84	упорядоч-	упорядочить
22	буд-	быть	85	включ-	включить
23	схд-	схвать	86	выдерж-	выдерживать
24	леч-	лечить	87	надлеж-	надлежать
25	бег-	бежать	88	обсто-	обстоять
26	жг-	жечь	89	содерж-*	содержаться
27	влек-	влекать	90	гранич-	граничить
28	мог-	мочь	91	наруш-	нарушить
29	вытек-	вытечь	92	свод-*	сводиться
30	возник-	возникнуть	93	близ-*	близиться
31	ид-	идти	94	ж-	жить
32	предпоч-	предпочесть	95	б-	бить
33	ве-	вести	96	доб-	добить
34	подвергну-	подвергнуть	97	разб-	разбить
35	бра-	брать	98	закл-	заклеить
36	упа-	упасть	99	влад-	владеть
37	да-	дать	100	Ø	есть
38	дава-	давать	101	п-	петь
39	удава-*	удаваться	102	выстр-	выстроить
40	име-*	иметься	103	хоч-	хотеть
41	исчерпа-*	исчерпаться	104	простере-	простереть
42	потребу-	потребовать	105	надлежа-	надлежать
43	требу-	требовать	106	недоста-	недостать
44	свед-	свести	107	исчез-	исчезнуть
45	простр-	простереть	108	пол-	положить
46	дад-	дать	109	выраст-	вырасти
47	сѣд-	сѣсть			
48	ед-	есть			
49	вид-	видеть			
50	терп-	терпеть			
51	смотр-	смотреть			
52	гон-	гнать			
53	говор-	говорить			
54	вер-	верить			
55	готов-	готовить			
56	вод-	водить			
57	закреп-	закрепить			
58	выстав-	выставить			
59	встрет-	встретить			
60	измер-	измерить			
61					
62	закле-	заклеить			
63	насто-	настоять			

  

List 1 Verbs		
Class	Full Word Example	Canonical Form
X1	вож/у	водить
X2	даш/ь	дать
X3	даст/Ø	дать
X4	ес/ть	быть
X5	су/ть	быть
X6	встре/ть	встретить
X7	добе/йтс	добить
X8	моч/ь	мочь
X9	ше/л	идти
X10	блужд/ая	блуждать
X11	вошед/ши	войти

## F. Example of Glossary Arrangement

As illustrations of the method of associating stems and endings for input word forms, we analyze several word forms having the richly homographic stem *пол-*. It is assumed that our glossary contains all words having this stem found in [1] and that no word form occurs in list 1a.

For reference, we first list these words by canonical form, in alphabetical order, together with their translations, stems and morphological classifications.

Several assumptions are made about the existence of forms. First, it is assumed that perfective verbs, like *полаять*, *полаяться*, *полить*, *политься*, have no present gerund and no present participles (active or passive). Second, it is assumed that reflexive verbs, like *политьсяя*, have no passive participles. Third, it is assumed that all nouns have plural forms. Fourth, it is assumed that the adjectives *полево́й* and *полово́й* have no short forms. Fifth, it is assumed that the adjectives *пологий* and *полый* have short forms, comparative short forms and corresponding adverbs in the positive and comparative degrees. Sixth, it is assumed that the glossary contains no proper names, not even that of the famous mathematician *По́лия*, or *По́ля* (*Pólya*). Seventh, it is assumed that the verbs *полоть* and *полоться* have no present gerunds.

пол	"sex"				политься	"be watered"				
по-		N	μ	i	40	пол-*	V	p	r	96
пол-		N	μ	i	9	поле/йте*	V	p	r	X7
						поль-*	V	p	r	5
пол	"floor"					поло	"polo"			
по-		N	μ	i	40	по/ло	N	ν	i	X15
пол-		N	μ	i	9					
пол/у		N	μ	i	X22	поло	"hollowly"			
(special locative)						по/ло	D	p		
						пол/ее	D	c		
поля	"tail (of a coat)"					полова	"chaff"			
по-		N	φ	i	24	пол/ов	N	φ	i	X11
пол-		N	φ	i	5	полов-	N	φ	i	5
поле	"field"					пологий	"slanting"			
пол-		N	ν	i	14	пол/ого	J	p		X3
полевать	"hunt"					полог-	J	p	LS	9
пол/ую		V	i	n	X1	полож/е	J	c		X2
полева-		V	i	n	35					
полю-		V	i	n	7	полого	"slantingly"			
полет	"flight"					пол/ого	D	p		
пол-		N	μ	i	37	полож/е	D	c		
полет-		N	μ	i	9	полом	"fracture"			
полив	"watering"					пол-	N	μ	i	38
пол/ив		N	μ	i	X2	полом-	N	μ	i	9
полив-		N	μ	i	9	полоть	"weed"			
полива	"glaze"					по/ли	V	i	n	X6
пол/ив		N	φ	i	X11	пол-	V	i	n	108
полив-		N	φ	i	5	поло-	V	i	n	35
политый	"watered"					полоться	"be weeded"			
пол/ит		J	p		X1	пол-*	V	i	r	42
полит-		J	p	LS	2	поло-*	V	i	r	35
полить	"water"					(third person only)				
пол-		V	p	n	96	полый	"hollow"			
полс/йте		V	p	n	X7	по-	J	p	S	35
поль-		V	p	n	5	пол-	J	p	LSC	2

Next we list the items with stem *пол-* which go into list 1.

пол/ее	D c		(поло, adverb)
пол/ив	N $\mu$ i	X2	(полиив)
	N $\phi$ i	X11	(полиива)
пол/ит	J p	X1	(политый)
пол/ов	N $\phi$ i	X11	(полова)
пол/ого	D p		(полого)
	J p	X3	(пологий)
пол/у	N $\mu$ i	X22	(пол, "floor")
пол/юю	V i п	X1	(полевать)

Next we list the items with stem *пол-* which go into list 2. The list is arranged by part of speech and subdivision thereof.

пол-	V p n	96	(полить)
	V p r	96	(политься)
	V i n	108	(полоть)
	V i r	42	(полоться)
	J p	LSC 2	(полый)
	N $\mu$ i	9	(пол, "sex")
	N $\mu$ i	9	(пол, "floor")
	N $\mu$ i	37	(полет)
	N $\mu$ i	38	(полом)
	N $\phi$ i	5	(пола)
	N $\nu$ i	14	(поле)

Altogether, 43 different endings may be split from word forms with the stem *пол-*. Of these, six belong to three canonical forms each (not distinguishing between the two nouns *пол*):  
е, ив, ов, ого, ом, ы.

Thirteen belong to two canonical forms each:

ам, ами, ах, ее, ей\*, ем, ет\*, ете, ой, ому, ою, у, ю.

Twenty-four endings belong to one canonical form each:

ая, ешь, ивший\*, ил\*, ила\*, или\*, ило\*, ит, ите, ить\*, ое, ую, ые, ый, ым, ыми, ых, ью\*,  
ют, юю, я, ям, ями, ях.

(The endings marked with an asterisk are also used with one reflexive form.) Because of the nature of our splitting process, the four endings  $\emptyset$ , -а, -и, -о can never be split off from a word form to give a stem ending in -л- like *пол-*.

As an example of the way that our glossary is used, suppose that the word form *полов* is encountered in text. The splitting process recognizes that the word form has the ending -ов and the nonreflexive stem *пол-*. (The splitting process must be done here, since it is assumed that this word form is not in list 1a.) The dictionary search through list 1b finds the use of the word form as the genitive plural of *полова*. (The list 1b search requires only that the ending be explicitly listed in the glossary. Hence the word forms in list 1b with different endings are passed over.) Then list 2 is searched. Each item in list 2 is tested to see if the ending -ов is compatible with it. Here compatibility is found only as the past gerund of *полоть* and as the genitive plural of the two nouns *пол*. Thus, four uses ("temporary choices") are found for this text occurrence.

RECEIVED MARCH 3, 1965

## REFERENCES

1. Академия Наук СССР, Институт Языкознания. Словарь русского языка (4 vols.). Государственное Издательство Иностраных и Национальных Словарей, Moscow, 1957-1961. [USSR ACADEMY OF SCIENCES, INSTITUTE OF LINGUISTICS. *Dictionary of the Russian Language* (4 vols.). State Publishing House for Foreign-Language and National Dictionaries. Moscow, 1957-1961.]

2. —, Институт Русского Языка. Словарь современного русского литературного языка (vols. 5-9). Издательство Академии Наук СССР, Moscow-Leningrad, 1956-1959. [—, INSTITUTE OF THE RUSSIAN LANGUAGE. *Dictionary of the Contemporary Russian Literary Language* (vols. 5-9). USSR Academy of Sciences Publishing House, Moscow-Leningrad, 1956-1959.]
3. —. Грамматика русского языка, vol. 1 (Фонетика и морфология), под редакцией академика В. В. Виноградова. Издательство Академии Наук СССР, Moscow, 1960. [—, *Grammar of the Russian Language, vol. 1* (Phonetics and Morphology). V. V. Vinogradov (Ed.), USSR Academy of Sciences Publishing House, Moscow, 1960.]
4. ALT, F. L., AND RHODES, I. Recognition of clauses and phrases in machine translation of languages. Proc. 1961 Int. Conf. on Machine Translation of Languages and Applied Language Analysis, vol. 1, pp. 125-142, Nat. Phys. Lab. Symp. 13, London, 1962.
5. AMERICAN MATHEMATICAL SOCIETY. *Russian-English Vocabulary with a Grammatical Sketch*. Amer. Math. Soc., New York, 1950.
6. DAUM, E., AND SCHENK, W. *Die Russischen Verben*. VEB Bibliographisches Inst., Leipzig, 1954.
7. RHODES, I. A new approach to the mechanical syntactic analysis of Russian. *Mech. Transl.* 6 (1961), 33-50.
8. Ушаков, Д. Н. Толковый словарь русского языка (4 vols.). Tokyo, 1953. [USHAKOV, D. N. *Explanatory Dictionary of the Russian Language* (4 vols.). Tokyo, 1953.]
9. Чебышев, П. Л. Полное собрание сочинений, vol. 5, pp. 27-38. (Beginning with "Показать" and ending with "изложения теории вероятностей," excluding three sentences containing mathematical symbols.) Издательство Академии Наук СССР, Moscow, 1951. [CHEBYSHEV, P. L. *Complete Collected Works, vol. 5*, pp. 27-28. (Beginning with "Показать" and ending with "изложения теории вероятностей," excluding three sentences containing mathematical symbols.) USSR Academy of Sciences Publishing House, Moscow, 1951.]