[*International Conference on Machine Translation of Languages and Applied Language Analysis*,
National Physical Laboratory, Teddington, UK, 5-8 September 1961]

THE GRAMMATICAL INTERPRETATION OF RUSSIAN INFLECTED FORMS USING A STEM DICTIONARY

by

J. McDANIEL and S. WHELAN,
National Physical Laboratory,
Teddington, England

INTRODUCTION

THE NPL Russian-English automatic dictionary is organised on a stem-paradigm basis wherein there is for most nouns and adjectives a single entry for all their inflected forms and for most verbs only one or two entries. This is in contrast to the full-form type of dictionary organisation wherein each inflected form of every word has a separate entry. The decision to organise our dictionary on this basis was made so as to be able to accommodate it on the magnetic tape store available to us on the ACE digital electronic computer of our laboratory, and, further, to minimise the look-up time per word on the computer without complicating the look-up procedure too much or investing too much programming effort in its compilation. The word content of the dictionary initially is to be 15,000 words from the Harvard University Automatic Dictionary. Out dictionary will have an average of about 1.5 entries per word, whereas a full-form dictionary would have about ten times that average.

The operation of our stem-paradigm dictionary involves two extra processing steps as compared with the full-form type dictionary. Firstly, words referred to the dictionary are reduced to their stems so that they may be matched against the corresponding dictionary stem entries and, secondly, after matching of stems, that part of the referred word split off to give the stem requires interpretation to determine its grammatical significance for that stem. The first process is known as affix-splitting and consists of matching the end of a referred word against a list of recognised affixes having grammatical significance. The process is fully described in a companion paper to this,[1,2]. We shall refer to the results of these papers where necessary. The second process, affix interpretation, is the subject of this paper. The extra grammatical propertles of the referred word revealed by affix identification, in addition to those identifiable in the stem of the word are as follows, for nouns, adjectives and verbs:-

NOUNS:- Number and case

ADJECTIVES:- Number, case, gender, short or long form

VERBS:- Person, number, tense, gender, mood, voice, and, for participles only, case and short or long form.

Of course, not all combinations of these properties can occur. The majority of pronoun forms are treated like adjectives. The remaining pronoun forms and all indeclinable words are referred to full-form type dictionary entries, and do not participate in affix identification, although they undergo the splitting process.

Affix interpretation is necessary for all stem type entries as its results form the basis of systems of syntactic analysis designed to improve a word-for-stem type "translation" of Russian into English. Rules of English inflection, insertion of prepositions and auxiliaries, suppression of Russian equivalents and variations of word order will all require the affix interpretation results.

2. PRINCIPLE OF INTERPRETATION

THE splitting process consists in matching the endings of text words against a list of affixes, and splitting off any matched affixes, so that the interpretation problem may be stated as the problem of giving a grammatical significance to each of these recognised affixes when they are found. Now some of the affixes will have varying significance depending on the stem from which they have been split. For instance, one of the affixes in the list is —A, and this can have five different interpretations:-

1. Genitive singular when split from some masculine noun stems.

2. Genitive singular and nominative plural when split from some other masculine noun stems and from neuter noun stems.

5. Nominative singular when split from feminine noun stems.

4. Feminine short form when split from adjective and participle stems.

5. Present Gerund when split from verb stems.

So for these ambiguous affixes (they are mostly noun affixes) it is necessary to check the stem type from which the affix has been split before giving the grammatical significance.

There is a further check, on the ***validity*** of a given split, which can be conveniently made during interpretation. This is to check that the matched dictionary stem includes the split-off affix in the declension or conjugation intended to be associated with it in the dictionary compilation stage. We call this check reconciliation of stem and affix, and it is necessary because of the occurrence of stem homographs and also because of the possibility of a text word whose true stem is not entered in the dictionary being falsely split and the resulting stem matching with a dictionary stem.

We combine interpretation and reconciliation in one operation, making use of a paradigm indicator associated with each stem, and one or more role indicators associated with each affix. By speaking of the paradigm of a stem, we mean that set of our recognised affixes, all of which combine with that stem to form valid inflectional forms of one Russian word. Thus each stem entry in the dictionary contains a computer word, known as the paradigm indicator word (PIW), which indicates by a binary pattern the paradigm of that stem. There are three different formats for the PIW for noun, adjective and verb stems. The verb format is used for two types of verb stems, but in each case it represents a different set of endings. This was only necessary in practice because one computer word (the ACE word is 48 binary digits (bits) long) is not long enough to represent all the verbal affixes. We shall consider the noun format of the PIW as a specific example.

The word is divided into fields, one for each of the case and number combinations of nouns. Accusative plural is excluded, as its endings follow those of nominative plural or genitive plural depending on the animation of the noun. In each field, a bit position is associated with each affix that conveys the significance of that field with a noun stem. The noun format is shown in ***Figure 1****.* (# is our symbol for the null affix). In the accusative singular field, only the feminine affixes are shown, the masculine and neuter affixes being implicit from the nominative singular, and genitive singular fields and the animation marker in bit position 43. We could have repeated the masculine and neuter, nominative and genitive singular endings in the accusative singular field, but this would have required more bit positions than are available in an ACE word. So simply by indicating the animation of a noun stem, we can restrict the paradigm format to within one ACE word.

The PIW for a particular noun stem is formed in general by inserting a binary digit 1 in the bit position corresponding to the appropriate affix in each field. For example, consider the stem entry and PIW resulting from the Russian word whose nominative singular is СТОЛ (table). The stem entry will be СТОЛ– and the set of affixes which give all the inflected forms of СТОЛ is #, А,У,Е,ОМ,Ы,ОВ,АМ,АХ,АМИ.
The PIW will thus have "ones" in positions 1, 11, 15, 19, 21, 26, 32, 37,

NOMINATIVE SINGULAR							ACCUSATIVE SINGULAR			GENITIVE SINGULAR				DATIVE SINGULAR				LOCATIVE SINGULAR		INSTRUMENTAL SINGULAR					NOMINATIVE PLURAL				GENITIVE PLURAL						DATIVE PLURAL		LOCATIVE PLURAL		INST. PLURAL		X	A				TYPE.	
#	Й	Ь	А	Я	О	Е	У	Ю	Ь	А	Я	Ы	И	У	Ю	Е	И	Е	И	ОМ	ЕМ	ОЙ	ЕЙ	ЬЮ	Ы	А	И	Я	Ь	Й	ОВ	ЕВ	#	ЕЙ	ЯМ	АМ	ЯХ	АХ	ЯМИ	АМИ						0	0
1				5					10					16					20					25					80					35					40					45			48

Fig. 1. Noun-Type Paradigm Indicator Word Format.

ADJECTIVAL TYPE FORMAT.

MASCULINE SINGULAR											FEMININE SINGULAR						NEUTER SINGULAR										PLURAL										SHORT						ADV.			TYPE.	
NOM.			GEN.		DAT.		LOC.		INST.		NOM.		ACC.		G., D., L. & I.		NOM & ACC.		GEN		DAT.		LOC		INST.		NOM.		GEN.		DAT		LOC.		INST.		MAS.	FEM.	NEUT.		PLURAL						
ЫЙ	ИЙ	ОЙ	ОГО	ЕГО	ОМУ	ЕМУ	ОМ	ЕМ	ЫМ	ИМ	АЯ	ЯЯ	УЮ	ЮЮ	ОЙ	ЕЙ	ОЕ	ЕЕ	ОГО	ЕГО	ОМУ	ЕМУ	ОМ	ЕМ	ЫМ	ИМ	ЫЕ	ИЕ	ЫХ	ИХ	ЫМ	ИМ	ЫХ	ИХ	ЫМИ	ИМИ	#	А	О	Е	Ы	И	О			1	0
1				5					10					15					20					25					30					35					40					45			48

“И” VERB TYPE FORMAT.

INFINITIVE			PRESENT								IMPERATIVE						GERUNDS				PAST					ACTIVE PARTICIPLES				PASSIVE PARTICIPLES					SHORT PASSIVE PARTICIPLE											TYPE	
			1ST SING.		2ND SING	3RD SING	1ST PLUR	2ND PLUR	3RD. PLUR.		SING.			PLURAL			PRESENT		PAST		MASC.		FEM.	NEUT.	PL.	PRESENT		PAST		PRESENT		PAST															
ИТЬ			У	Ю	ИШЬ	ИТ	ИМ	ИТЕ	АТ	ЯТ	И	Й	Ь	ИТЕ	ЙТЕ	ЬТЕ	А	Я	ИВ OR ИВШИ		ИЛ		ИЛА	ИЛО	ИЛИ	АЩ	ЯЩ	ИВШ		ИМ			ЕНН	ИТ		ЕН										0	1
1				5					10					15					20					25					30					35					40					45			48

“E” VERB TYPE FORMAT.

INFINITIVE			PRESENT								IMPERATIVE						GERUNDS				PAST					ACTIVE PARTICIPLES				PASSIVE PARTICIPLES					SHORT PASSIVE PARTICIPLE											TYPE.	
			1ST. SING.		2ND SING	3RD SING	1ST. PLUR.	2ND PLUR	3RD. PLUR.		SING.			PLURAL			PRESENT		PAST		MASC.		FEM	NEUT	PL.	PRESENT		PAST		PRESENT		PAST															
ТЬ	ТИ	Ь	У	Ю	ЕШЬ	ЕТ	ЕМ	ЕТЕ	УТ	ЮТ	И	Й	Ь	ИТЕ	ЙТЕ	ЬТЕ	А	Я	В & ВШИ	ШИ	Л	#	ЛА	ЛО	ЛИ	УЩ	ЮЩ	ВШ	Ш	ЕМ	ОМ	НН	ЕНН	Т	Н	ЕН										1	1
1				5					10					15					20					25					30					35					40					45			48

Fig. 2. Adjective and Verb Type Paradigm Indicator Word Formats.

39 and 41. The absence of a "one" in bit position 43 indicates the inanimate nature of the stem and hence implicitly indicates the accusative singular and accusative plural endings. A stem which takes alternative affixes in a given field will have "ones" in the bit positions of both affixes e.g. the stem ВОЛОС (hair) has the alternative affixes Ы and А in the nominative plural form. Where a stem is not common to all inflected forms of a word, only those fields to which that stem applies will have a "one" in them e.g. the stem БРАТ- (brother) applies to the singular inflected forms only (1, 11, 15, 19, 21, 43) while the stem БРАТЬ- applies to the plural forms (29, 33, 36, 38, 40, 43).

The formats for adjectives and verbs are shown in ***Figure 2*** and in principle are similar to the noun format. They all have more fields than the noun format, but have much less variety of affixes within each field. The two verb formats have identical fields, but mostly different affixes in those fields. They include fields for participle affixes, but the affixes in these fields are only the participle stem-building affixes. However, as participle adjectival endings follow a perfectly regular pattern, they need not be explicitly stated in the PIW.

Nearly all nouns and adjectives will require only one stem and PIW to represent all their inflected forms. Approximately 2/3 of Russian verbs will need only one stem, most of the rest requiring two stems, and only the irregular verbs more than two.

The PIW are complied by the computer from data sheets (dictionary entry forms) one of which is manually completed for each word to be entered into the dictionary. There is a different data sheet for each of several broad classes of noun declension, so as to limit the linguistic decisions to be made in completing the sheets, but all noun data sheets refer to the one standard format for the noun PIW. There are similar data sheets for adjectives and the two types of verbs, in these cases only one type of data sheet per format, because of the lesser variety of inflection.

With the provision of a PIW in each stem entry in the dictionary, the problem of interpretation of an affix which has occurred on a given stem as a text word, is resolved into spotlighting the occurrences (if any) of that affix in the PIW for that stem and noting the fields (grammatical properties) in which they occur. This is most easily done by having, for that affix, a masking pattern containing a "one" bit corresponding to each occurrence of it in the PIW format. Then, by performing a "logical and" operation between this mask and the PIW of the given stem, the result will contain a "one" bit in each field where that affix has significance for the given stem. Of course, if the result was zero, this would mean that the affix and stem were incompatible i.e. the stem did not combine with the affix in any meaningful inflection. This situation may arise with stem homographs and with words whose true stems are not yet complied into the dictionary and are falsely split. In the latter case the PIW would not contain the falsely split affix.

The masking pattern referred to above we call the role indicator word (RIW) for the given affix. Some affixes have significance with more than one of the PIW formats, and for these there will need to be more than one RIW e.g. И has significance for and appears in each of the four PIW formats, so it will have four RIW. In order to be able to match the appropriate RIW to a given PIW in an interpretation, the format types are given a type number (digits 47 and 48) and the RIW which relate to these types are given the corresponding type no. There are identical И and E verb RIW for each of 10 verbal affixes (У , Ю , И , Й , Ь , ЙТЕ , ЬТЕ , А , Я , ЕНН) and so we save some space in storing the RIW by having only one verb RIW for each of these 10 and indicating its dual utility.

Let us consider two examples of interpretation of noun forms АВТОМОБИЛИ and НЕДЕЛИ, which would be matched against the dictionary stems АВТОМОБИЛ- and НЕДЕЛ- respectively, with И as the affix to be interpreted in both cases. The PIW for the noun stem АВТОМОБИЛ- and the noun type RIW for И would be as shown in ***Figure 3.*** The "logical-and" of these two computer words would give a "one" bit in position 28 only i.e. in the nominative plural field. The PIW for НЕДЕЛ- is also shown in ***Figure 3*** and the result of "anding" this word with the RIW for И would be a "one" bit in positions 14 and 28 i.e. in the genitive singular and nominative plural fields. In both cases, the nominative plural indication would be extended by the interpretation routine to indicate also the accusative plural, on checking that the animation digit (position 43) in the PIW indicated "inanimate".

3. ORGANISATION OF INTERPRETATION

We have explained above the principle of interpretation of the significance of a given affix with a given dictionary stem. The actual organisation of the interpretation process in the NPL dictionary may best be explained by reference to the flow chart of the process shown in ***Figure 4***.

The starting point for this flow-chart is the successful match of the stem resulting from splitting a text word, with a full dictionary entry. When such a match is made, the address in the computer storage of the RIW of the affix resulting from the split of the text word is available in the affix identifier word (AIW), which is part of the output of the splitting process. The matching operation may actually alter the position of the original split of the text word, which will always be

PIW FOR NOUN STEM "АВТОМОБИЛ-".

1				5					10					15					20					25					30					35					40					45			48
0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0

NOUN RIW FOR "-И".

1				5					10					15					20					25					30					35					40					45			48
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

PIW FOR NOUN STEM "НЕДЕЛ-".

1				5					10					15					20					25					30					35					40					45			48
0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

Fig. 3. PIW & RIW for Interpretation Examples.

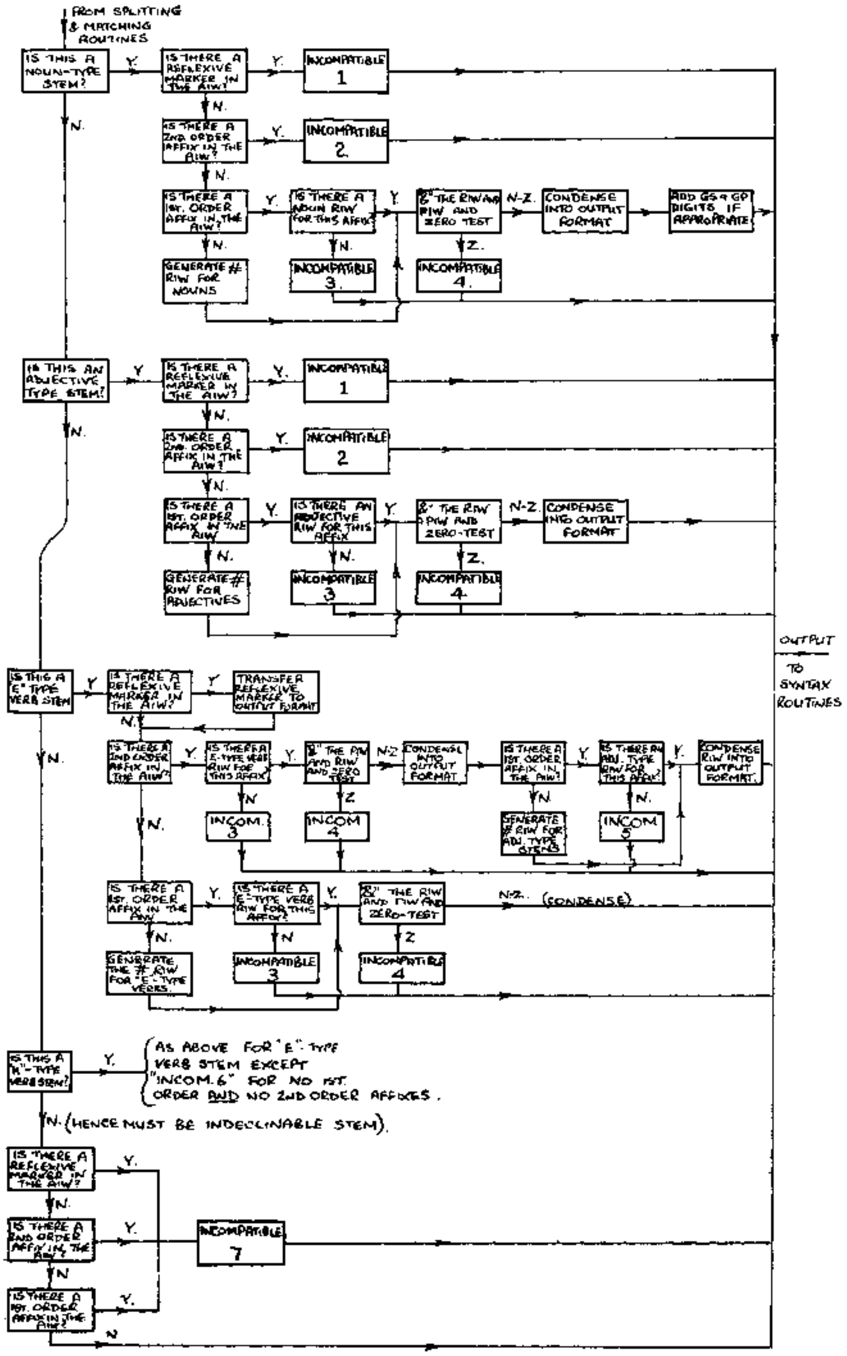
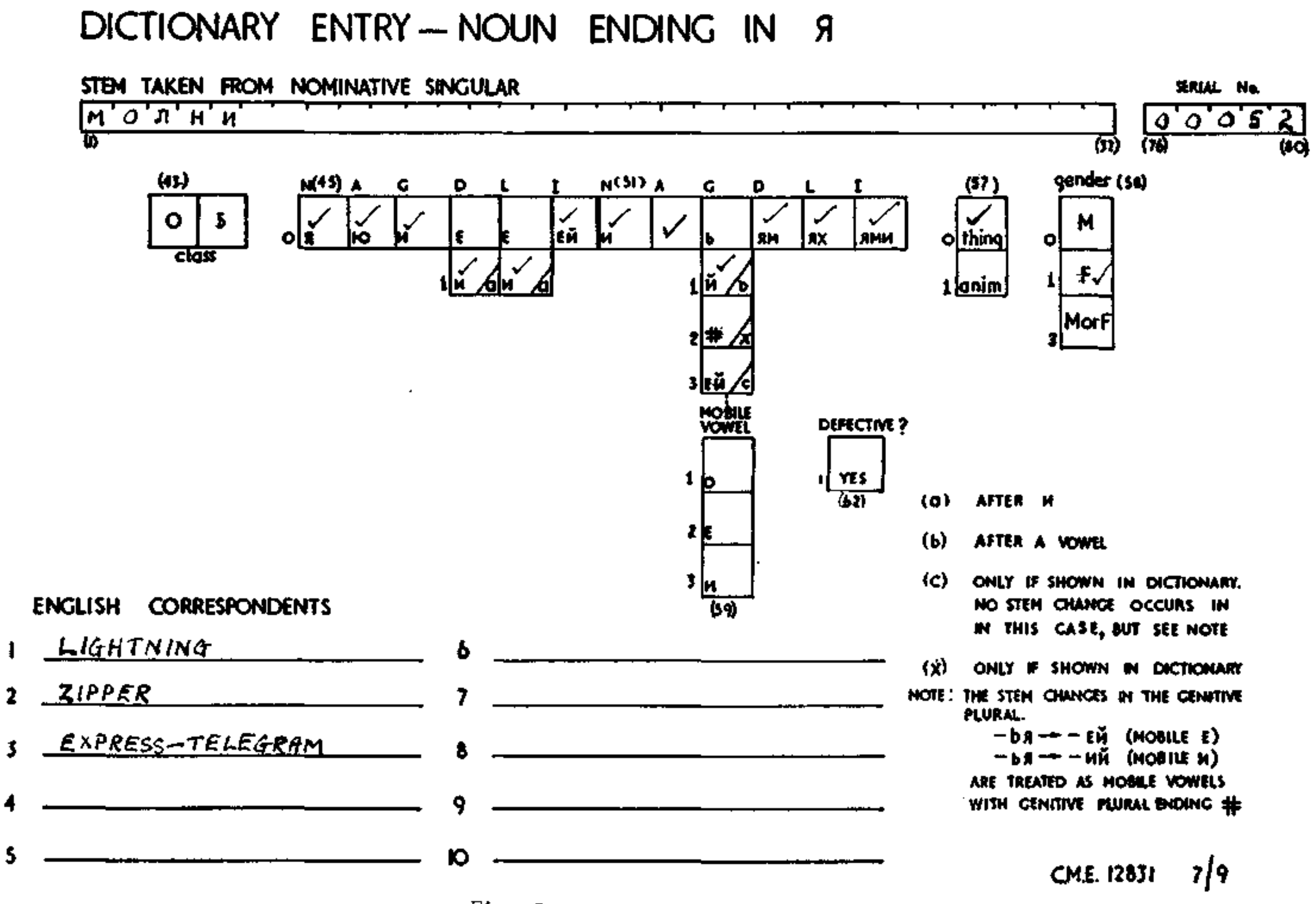

FROM SPLITTING & MATCHING ROUTINES
IS THIS A NOUN-TYPE STEM?
Y.
IS THERE A REFLEXIVE MARKER IN THE AIW?
Y.
INCOMPATIBLE 1
N.
N.
IS THERE A 2ND. ORDER AFFIX IN THE AIW?
Y.
INCOMPATIBLE 2.
N.
IS THERE A 1ST. ORDER AFFIX IN THE AIW?
Y.
IS THERE A NOUN RIW FOR THIS AFFIX?
Y.
"&" THE RIW AND PIW AND ZERO TEST
N-Z.
CONDENSE INTO OUTPUT FORMAT
ADD GS & GP DIGITS IF APPROPRIATE
N.
GENERATE # RIW FOR NOUNS
N.
INCOMPATIBLE 3.
Z.
INCOMPATIBLE 4.
IS THIS AN ADJECTIVE TYPE STEM?
Y
IS THERE A REFLEXIVE MARKER IN THE AIW?
Y.
INCOMPATIBLE 1
N.
N.
IS THERE A 2ND. ORDER AFFIX IN THE AIW?
Y.
INCOMPATIBLE 2
N.
IS THERE A 1ST. ORDER AFFIX IN THE AIW
Y.
IS THERE AN ADJECTIVE RIW FOR THIS AFFIX
Y.
"&" THE RIW PIW AND ZERO-TEST
N-Z.
CONDENSE INTO OUTPUT FORMAT
N.
GENERATE # RIW FOR ADJECTIVES
N.
INCOMPATIBLE 3
Z.
INCOMPATIBLE 4
OUTPUT TO SYNTAX ROUTINES
IS THIS A "E" TYPE VERB STEM
Y
IS THERE A REFLEXIVE MARKER IN THE AIW?
Y
TRANSFER REFLEXIVE MARKER TO OUTPUT FORMAT
N.
N.
IS THERE A 2ND ORDER AFFIX IN THE AIW?
Y.
IS THERE A E-TYPE VERB RIW FOR THIS AFFIX
Y.
"&" THE RIW AND RIW AND ZERO TEST
N-Z
CONDENSE INTO OUTPUT FORMAT
IS THERE A 1ST. ORDER AFFIX IN THE AIW?
Y.
IS THERE AN ADJ. TYPE RIW FOR THIS AFFIX?
Y.
CONDENSE RIW INTO OUTPUT FORMAT.
N
INCOM. 3
Z
INCOM 4
N.
GENERATE # RIW FOR ADJ. TYPE STEMS
N.
INCOM 5
N.
IS THERE A 1ST. ORDER AFFIX IN THE AIW
Y.
IS THERE A E-TYPE VERB RIW FOR THIS AFFIX?
Y.
"&" THE RIW AND PIW AND ZERO-TEST
N-Z. (CONDENSE)
N.
GENERATE THE # RIW FOR "E"-TYPE VERBS.
N
INCOMPATIBLE 3
Z
INCOMPATIBLE 4
IS THIS A "K"-TYPE VERB STEM?
Y.
AS ABOVE FOR "E"-TYPE VERB STEM EXCEPT "INCOM. 6" FOR NO 1ST. ORDER AND NO 2ND ORDER AFFIXES.
N. (HENCE MUST BE INDECLINABLE STEM).
IS THERE A REFLEXIVE MARKER IN THE AIW?
Y.
N.
IS THERE A 2ND ORDER AFFIX IN THE AIW?
Y.
INCOMPATIBLE 7
N
IS THERE A 1ST. ORDER AFFIX IN THE AIW?
Y.
N

Fig. 4. Flow-Chart of Interpretation Routine.

after the largest affix in the affix list. Such an alteration will always be so as to increase the stem length and decrease the affix length. However, the AIW contains the addresses of the RIW of all potential affixes so that the appropriate one of them can always be indicated, corresponding to the enlarged stem, if this type of matching is performed (for full details of splitting and matching see Davies[1], and Davies and Day[2]). These RIW addresses will always refer to the first of the group of RIW for the affix concerned if it is a multi-role affix and the final RIW of the group will be distinguished by a special digit (digit position 46).

The splitting process is applied to each text word in three stages. Firstly, the reflexive affixes СЬ and СЯ are split off, if occurring. Secondly, the longest first-order affix (all remaining affixes except participle stem-building affixes are first-order affixes) is split off and thirdly, the longest second-order affix (all participle stem-building affixes are second-order) is split off. The results of all stages are recorded in the AIW.

Referring now to the flow chart, it is divided into four main sections depending on the stem type of the matched dictionary entry as indicated in the PIW of that entry. The occurrence of a reflexive marker or address of a second-order affix in the AIW for a word matched against a noun or adjective type dictionary entry could not possibly be given a valid interpretation, and so these occurrences are shown as INCOMPATIBLE 1. They can arise from words whose stems have not been compiled into the dictionary but whose stems are homographic with one or other dictionary entries. We have not attempted to discover words which would cause such incompatibility but for the sake of a watertight routine we show paths in it should they occur. The interpretation output for such words will show their type of incompatibility. For all other affixes occurring, the first check must be whether the affix has a RIW of the same type as the matched stem type (Noun, adjective, И-verb, E-verb). If it does not, then there has occurred another type of stem-affix incompatibility, due again to the stem-homography of an uncompiled word. If the affix identifier word does not indicate a first-order affix, then the RIW for the null affix appropriate to the type of the matched stem must be extracted from the RIW list and used in interpretation. Then interpretation may be carried out with RIW and PIW of the same type as described in principle above. If the result of the "and" operation is zero, then a third type of incompatibility has occurred, due to the same causes as before. The non-zero result of the "and" operation (i.e. a valid interpretation) at this stage is in the same format as the PIW. We may conveniently condense it to a format where there is only one digit position for each of the fields of grammatical significance in the PIW format. For now we are only interested in whether or not there has been an interpretation



DICTIONARY ENTRY — NOUN ENDING IN Я

STEM TAKEN FROM NOMINATIVE SINGULAR (1)–(32): М О Л Н И

SERIAL No. (76)–(80): 0 0 0 5 2

class (43): O 5

	N(45)	A	G	D	L	I	N(51)	A	G	D	L	I
0	Я ✓	Ю ✓	И ✓	Е	Е	ЕЙ ✓	И ✓	✓	Ь	ЯМ ✓	ЯХ ✓	ЯМИ ✓
1				И ✓ / a	И ✓ / a				Й ✓ / b			
2									# / x			
3									ЕЙ / c			

(57): 0 thing ✓ · 1 anim

gender (58): 0 M · 1 F ✓ · 3 M or F

MOBILE VOWEL (59): 1 О · 2 Е · 3 И

DEFECTIVE? (62): 1 YES

(a) AFTER И

(b) AFTER A VOWEL

(c) ONLY IF SHOWN IN DICTIONARY. NO STEM CHANGE OCCURS IN IN THIS CASE, BUT SEE NOTE

(x) ONLY IF SHOWN IN DICTIONARY

NOTE: THE STEM CHANGES IN THE GENITIVE PLURAL.
—ЬЯ → —ЕЙ (MOBILE Е)
—ЬЯ → —ИЙ (MOBILE И)
ARE TREATED AS MOBILE VOWELS WITH GENITIVE PLURAL ENDING #

ENGLISH CORRESPONDENTS

1 LIGHTNING

2 ZIPPER

3 EXPRESS–TELEGRAM

4

5

6

7

8

9

10

C.M.E. 12831 7/9

Fig. 5

of affix significance in a given field, and one bit per field will suffice. Thus the noun interpretation output may be condensed to a 12-bit group (one bit for each case-number combination), the adjective output to a 23-bit group and the verb output to a 20-bit group. At the moment, the formats of these groups has followed naturally from the PIW formats. It may eventually prove preferable to have other formats, depending on the syntactic routines which will make use of the interpretation output. Parallelism between the noun and adjective formats would facilitate agreement checks.

When a second-order affix occurs with a verbal stem, it is interpreted first and then the first-order affix is checked as to whether it is adjectival. If it is, then the affix need not be interpreted, as the paradigms for participle stems are perfectly regular and participle affixes do not have variable significance depending on the stem. The RIW can be condensed directly and inserted into the output format. There is one exception to this last statement. The RIW for ОЙ specifies a nominative singular significance (in addition to others) but this does not occur with participle stems. A special participle RIW for ОЙ excluding the nominative singular role will solve this problem.

Finally, if the matched stem is of the indeclinable type, it will actually be the whole indeclinable word and so there will be no affixes that will require interpretation with it. If however, the affix identifier word indicates any affixes, then these must give an incompatible output. The grammatical significance of the indeclinable word will be transferred directly to the interpretation output from the stem entry, if no affixes are indicated.

4. COMPILATION OF THE PARADIGM INDICATOR WORDS

For purposes of PIW compilation (which is part of the general dictionary compilation procedure), nouns are divided into nine main classes.

To each of these classes there corresponds a dictionary-entry form. The classification (with the possible exception of class 9) is more or less conventional and merely one of many possible classifications. Each of these main classes will have various subclasses and the manner in which the various subdivisions are catered for in the entry form for the main class will best be explained if we consider one specific class. Our class 5, i.e. nouns ending in -Я, will illustrate the method.

The dictionary-entry form for class 5 is shown in ***Figure 5*** and has been completed for the particular noun МОЛНИЯ. Clearly note (a) in the form applies to this noun whereas, in the case of the noun НОЗДРЯ say, note (a) does not apply and hence the dative and locative singular of the latter noun end in —Е. Note (b) applies to the case of СУДЬЯ and we indicate that there is an Е/Ь mobility to get its genitive plural СУДЕЙ. ВИШНЯ, on the other hand, has a mobile Е and its genitive plural ends in #. КУХНЯ has a mobile О to form its genitive plural which also ends in #. In the case of СВАТЬЯ, note (x) applies for the formation of its genitive plural. РОДНЯ, for example, presents no difficulties in the singular, but it is defective in that it has no plural. The form caters for all these divergencies. All the other classes behave in a somewhat similar fashion and a series of like devices enable the various subclasses to be incorporated in the entry-form for the main class.

It may be that our final class 9 needs a special word of explanation. This class contains all indeclinable nouns and nouns having only a partial declension. Indeclinables need no further mention, since their paradigm is invariable and equivalent to the nominative singular. There are nouns which have an essentially partial declension in the sense that, for example, they may have no plural or it may be that they are used only in the plural and, hence, have no singular. The noun ТЕМЯ, for instance, is not used in the plural but its paradigm in the singular can easily be fitted into the entry form for nouns ending in -Я as having no entries in the plural and marked as 'defective', which the entry-form permits. The noun ТЕМЯ therefore, is not a candidate for our class 9. Class 9 will contain all those nouns which have more than one stem in the formation of their paradigms and which cannot be conveniently fitted into any of the paradigms associated with any of the remaining classes. The noun ВРЕМЯ will amply illustrate the principle. As Я, И, ЕМ, А,#, АХ, and АМИ are affixes which are in the list we use for our splitting routine, it follows that the paradigm of ВРЕМЯ will have two distinct stems, viz. ВРЕМ- for the nominative and accusative singular and ВРЕМЕН- for all the other cases of singular and plural. (We regard Е and Ë as equivalent for machine purposes and, hence, the genitive plural ВРЕМËН is equivalent to ВРЕМЕН). ВРЕМЯ will be entered in a class 9 entry form as occurring in the nominative and accusative singular and is hence treated as a noun possessing a partial declension containing just those two cases; similarly ВРЕМЕНИ will be entered in a subsequent class 9 entry form as a noun having a partial paradigm in the genitive, dative and prepositional singular, and so on for ВРЕМЕНЕМ, ВРЕМЕНА, etc. All nouns having irregularities in declension can be easily fitted into this class.

Adjectives do not present many difficulties of classification and, in fact, there is but one entry form for all adjectives. This entry form covers nouns which decline like adjectives (as well as surnames in -ОВ, -ЕВ or ИН which decline like adjectives). Adverbial forms ending in —О (as well as forms in -И when the stem ends in -СК, or-ЦК) are incorporated in the form as also regular comparatives in -ЕЕ.

Verbs are divided into the conventional classification of И- and Е— types. Such characteristics as perfective, imperfective, perfective/imperfective, momentary and iterative are accounted for as well as whether the verb exists in a reflexive form only. In addition to the infinitive stem, many verbs will have a second stem used to form the present tense and those participial and other forms derived from it. The И—type verb БРОСИТЬ will exemplify the point. The infinitive stem is БРОС-, since -ИТЬ is one of the list of affixes we use in our splitting routine. This stem БРОС- is taken as the standard or canonical form and is used to form the past tense and those other verbal forms which are grammatically dependent on the infinitive stem. The first person of the present tense, is, however, БРОШУ. We indicate this fact on the entry form by indicating that the last letter (viz. -С in this case) is taken from the infinitive stem and the letter Ш substituted for it. That the letter Ш does not occur throughout the present tense is taken care of, and those forms which depend on the first person singular of the present tense are noted on the form. If, on the other hand, we consider the Е- type verb СОВЕТОВАТЬ, we indicate on the entry form that the infinitive stem is СОВЕТОВА- , since -ТЬ is also an affix we use in our splitting routine. We then indicate on the form that three letters (viz.—ОВА) are removed from this stem before adding the letter У to form the present stem СОВЕТУ- and we further indicate that this У is retained throughout the present tense and those verbal forms grammatically derivable from it. These examples serve to illustrate how the various verbal subclasses can be incorporated in the two main classes.

Finally the paradigm information on the dictionary-entry forms has to be transformed into the PIW formats. This is carried out simply by programmes which assemble and process the information contained in the entry-forms. It is to be noted that the entry-forms themselves contain all the necessary linguistic and grammatical decisions.

5. CONCLUSIONS

We have described that part of the operation of the NPL Russian-English automatic dictionary which interprets the grammatical significance of the part split off from a referred word in order to match it against the stem-type entries of the dictionary. A simple method for listing the endings (affixes) which may associate with stems to match referred words has been described and has the main feature of complete flexibility to represent any type of declension or conjugation. Interpretation of the significance of an affix and checking its valid association with a given stem is performed at the same time in one operation. We are confident that no invalid interpretation can be produced for a referred word. We have described the clerical organisation of data for the compilation of the endings lists (paradigms) for stem entries.

The principles of the method of interpretation of the recognised affixes of Russian may be applied without alteration to the interpretation of the affixes of any inflected language for which the stem type automatic dictionary gives worthwhile economies over the full-form type dictionary.

6. REFERENCES

1. DAVIES, D.W., The organization of a Russian-English Stem Dictionary on Magnetic Tape. *Language and Speech* 1960, **3**, Pt.4, 193-222.

2. DAVIES, D.W., and DAY, A.M., A technique for consistent splitting of Russian words. (this conference)

3. The Russian Verb, Nevill Forbes, Oxford, 1955.

4. MAZON, Andre, Grammaire de la Langue Russe, Institut d'Etudes Slaves, Paris, 1949.

5. TESNIERE, Lucien, DIDIER, Henri, Petit Grammaire Russe, Paris, 1945.