

Automation of Word-forming Process in the Romanian Language

Elena Boian, Anna Danilchenco and Ludmila Topal

Institute of Mathematics of the Academy of Sciences

5, Academiei str.,

Kishinev 277028,

THE REPUBLIC OF MOLDOVA

Abstract: The article describes the Romanian language words automatic inflexion system, which the Romanian texts spelling-checker ROMSP is based on. Verb conjugation, and noun and adjective declination are detailed.

Keywords: Word-forming process, inflexion system, natural language processing, spelling-checker, word-forms, formalization of the inflexion process, generating inflexion, the nouns and adjectives inflexion algorithm, the procedure of verb inflexion.

Elena Boian was born in Kishinev, The Republic of Moldova in 1954. She graduated from the Department of Mathematics and Cybernetics of the Kishinev University in 1979. She obtained her doctoral degree in Computer Science from the University of Kiev (Ukraine) in 1992. Since 1979 she has been working at the Institute of Mathematics of the Academy of Sciences of The Republic of Moldova. Her present position is senior research worker. Dr. Elena Boian published more than 30 scientific papers. Her research interests include compiler construction, natural language processing.

Anna Danilchenco was born in Chita, Russia. She graduated the Pedagogical Institute in 1956. She obtained her doctoral degree in Mathematical Logic from the Institute of Mathematics in Kishinev in 1979. Since 1964 she has been working at the Institute of Mathematics of the Academy of Sciences of The Republic of Moldova. Her present position is senior research worker. Dr. Anna Danilchenco published more than 30 scientific papers. Her research interests include multivalued logic, compiler construction, natural language processing.

Ludmila Topal was born in Kalarash, The Republic of Moldova in 1956. She graduated from the Department of Mathematics and Cybernetics of the Kishinev University in 1979. Since 1979 she has

been working at the Institute of Mathematics of the Academy of Sciences of the Republic of Moldova. Her present position is engineer programmer. Ludmila Topal published more than 10 scientific papers. Her research interests include compiler construction, natural language processing.

1. Introduction

Text editors' capability of orthographically correcting, and automated translation systems are supported by a special dictionary. A dictionary's structure in case of a highly inflexional language is a topical and difficult problem. Two methods for saving information are known: declarative (to memorize all inflexions) and procedural (to obtain the necessary inflexions in a special way, starting from one basic word or from some basic words). Such a dictionary can be enriched by using a declarative method which needs all word-forms be obtained.

Similarly, using a procedural method for inflecting a new word asks for inflectional rules to be set up.

To manually introduce all word-forms with the morphological information for series applications is a rather complicated, time-consuming and bulky activity. So, automating this process is a "must".

Pronouns, articles and numerals generate few flexions and can be introduced in the dictionary first time the spelling-checker is created.

However, the problem of completely formalizing the inflectional process is very difficult, and at least a partial solving is aimed at, but in special cases, users' additional information about the morphological categories, and the verb consonant and vowel alternations are considered.

The package has been implemented under Turbo-Pascal.

2. Speech Parts Inflecting Regularities

Six parts of speech (out of eleven) can be inflected. Nouns, adjectives, articles, numerals and pronouns decline in accordance with case, number and form. Verb conjugates in accordance with tense, mood, person, etc. Adverb, preposition, conjunction, particle and interjection are invariable [1, 2, 3]. Inflecting regularities of verbs, nouns and adjectives are now discussed.

The word "affix" is used for denoting the notion of both a suffix as an ending, and a suffix and an ending in accordance with the context.

2.1. The Verb

The package for generating verb inflexion produces the generated words-form list in accordance with the Infinitive of a verb. This list contains **Imperativul**, **Participiul**, **Gerunziul** flexions and verb flexions of those tenses only, which differ in accordance with person (i.e. **Indicativ Prezent**, **Conjunctiv Prezent**, **Imperfectul**, **Perfectul Simplu**, **Mai mult ca Perfectul**).

Literature has been generous in analysing the different schemes of verb conjugation. For example, in [5] the verb set is classified into 5 types, each of them having its characteristic features. Each type is divided into some classes, characterized by a suffix in accordance with the indicated mood or tenses of verbs. Each class (or type, if there is no division into classes) is in its turn divided into schemes, depending on the specific vowel alternation, into word roots, difference of suffixes, etc.

For example, the first type of verbs contains the verbs of the first grammatical group (that is, verbs with an ending "a" the Infinitive) and some verbs of the second grammatical group (of which the Infinitive ending is in "ea") which are conjugated by the suffix "ez" (Present tense, Indicative mood, Singular, and first person), for example, *a veghea* (to watch). This type of verbs yields the following characteristics:

- character "a" presence at the end of the word in some cases (**Infinitiv**, **Indicativ Prezent** and **Conjunctiv Plural** for some persons, **Imperfectul**, **Perfectul Simplu** and **Mai mult ca Perfectul** have the form: theme + "a" + affix);

- **Indicativ Prezent** and **Conjunctiv Singular** have the form : theme + suffix (0/"ez") + affix, where 0 means that a suffix is not needed;
- **Gerunziul** has the form : theme + "ind", (note that in accordance with [5], this type contains verbs such as *a muia* (to soak), *a infoia* (to swell), *a tăia* (to cut), etc., which have their corresponding **Gerunziu** in: *muind*, *infoind*, *tăind*, etc.).

Corresponding to the above-mentioned type of verbs, the two classes differ by the suffixes which **Indicativ Prezent Singular** takes (the suffixes 0 and "ez" for the verbs in the first and the second class respectively). The first class is divided into 15 schemes, the second one into 4 schemes.

To get a clear image of these schemes, we will study a representative case (Figure 1) in any of the 15 schemes of the first class, and show its conjugation in accordance with **Prezent Singular Indicativ**. The difference of conjugation in these schemes will thus appear. The consonant and vowel alternations are indicated in the order of their occurrence in a word. The flexion is obtained by dropping the ending "a" of the verb Infinitive, and by executing a set of alternations, and adding the necessary affixes.

Note: in cases 9 and 10 for II person there is an alternation *k — ch*, which is a phonetic one, but, graphically, it has an invariant form.

Figure 1 shows the specific form of conjugation for each scheme.

Other types, classes and schemes of conjugation do present similar characteristics. The second grammatical group (verbs with an ending "ea") contains one class, which is divided into 3 schemes, the third grammatical group (verbs with an ending "e") contains 2 types, 3 classes and 19 schemes, the fourth grammatical group includes one type, consisting of 4 classes and 15 schemes. As shown in [5], the regular verbs set is divided into 56 groups having in appendix the verbs set (about 4,000), with their conjugation scheme number. Although [5] makes an excellent start for developing an algorithm, it still contains some inaccuracies. For example, no scheme for the type of verbs *a apropia* (to draw), *a invia* (to revive) is provided. It proposes that these verbs are conjugated in accordance with the scheme of the verb *a încuia* (to lock), which is incorrect, because **Indicativ Prezent**,

Verb	p.I		p.II		p.III		
	alt.	af.	alt.	af.	alt.	af.	af.
a cînta (to sing)	—	—	t → ț		i	—	ă
a căuta (to seek)	ă → a	—	ă → a, t → ț		i	ă → a	ă
a arăta (to show)	—	—	t → ț		i	ă → a	ă
a apăra (to defend)	—	—	ă → e		i	—	ă
a apăsa (to press)	—	—	ă → e, s → ș		i	ă → a	ă
a căpăta (to obtain)	ă → a	—	ă → a, ă → e		i	ă → a	ă
a asemăna (to compare)	—	—	t → ț		i	e → ea	ă
a boteza (to baptize)	—	—	—		i	—	e
a convoca (to convene)	—	—	—		i	o → oa	ă
a juca (to dance)	u → o	—	u → o		i	u → oa	ă
a afla (to learn)	—	u	—		i	—	ă
a încuia (to lock)	—	—	—		i	—	e
a tăia (to cut)	ă → a	—	ă → a		—	ă → a	e
a înfoia (to swell)	—	—	—		—	o → oa	e
a muia (to soak)	u → o	—	u → o		—	o → oa	e
a învia (to revive)	—	i	—		i	—	e

Figure 1: Representative Elements of the First Class of Verbs Conjugation in Accordance with Present Singular Indicativ

I person, is *eu încui* (I lock), but — *eu învii* (I revive). That is the reason why Figure 1 includes a new string.

Imperativul Singular for such verbs as *a minți* (to lie), *a simți* (to feel), *a crede* (to believe), etc. has a single form *mințe*, *simțe*, *crede* in [5], but also the verbs have forms such as *minți*, *simți*, *crezi*. The verb *a însufla* (to inspire) is wrongly classified by the conjugation scheme of the verb *a cânta* (to sing), etc.

New aspects are put forward by [6], where several verbs conjugation schemes are indicated (about 300), mainly deriving from a lot of forms for one morphological category (including diverse regionalisms and seldom used forms). For example, such verbs as *a apropia* (to bring) and *a învia* (to revive) in [6] are classified into different conjugation schemes, because of a double form of the verb *a apropia*: *eu apropii* and *eu apropii*, and of the only form of the verb *a învia*, *eu învii*. [6] used both forms of **Imperativ Plural** of some verbs such as the verb *a naște* (to give birth) (for example *nașteți* and *nășteți*).

[5] was a basic reference. In the authors' opinion, types and classes description does not play a great part. All the information about a verb being included in one grammatical group or another, which is procedurally determined by the verb ending in the Infinitive has been utterly considered. In case of the verbs under grammatical groups I and IV, information should be available whether verb conjugates with a suffix or not. It is essential to know whether a verb is personal or impersonal. Given one mood or another, 35 or 12 flexions are generated. Answers can be provided by dialoguing with the system user. The information about including a verb in any group corresponds, according to [5], to including it in any type, but the division of a grammatical group into parts (with or without a suffix) corresponds to classes. Thus, we will investigate 6 classes of regular verbs, and one class of irregular verbs (*a fi* (to be), *a lua* (to take), *a avea* (to have), etc.). For each class, the procedures have been specified.

Our division of verbs into 6 classes overlaps that made in [5], and so, it is possible that the division into schemes as in [5] should be transferred into classes.

2.2. Nouns and Adjectives

As a part of speech, the noun modifies its form according to case and number, but keeps invariable in gender [4].

The following morphological categories are characteristic to noun: gender (*GF* — feminine, *GM* — masculine and *SN* — neuter noun), case (*CNA* — nominative-accusative, *CGD* — genitive-dative, *CV* — vocative), number (*NS* — singular, and *NP* — plural, singularia-tantum — nouns have only singular and pluralia-tantum — they have only plural), form (*FI* — indefinite and *FD* — definite), common and proper.

The non-registered affixes have, after examining the Romanian nouns, been grouped in a lot of series. The affixes series characterizing, for example, the neuter nouns is presented in Figure 2. The analogous affixes series is studied for feminine and masculine nouns.

Homonymy types ranging from A to C will be identified by examining the above-discussed series of affixes and based on gender importance in noun inflexions in Romanian, namely:

Homonymies types	a	b	c	d	e	f	for noun (gender)
A	+	+	+	+	+		feminine
B		+	+	+	+	+	masculine
C	+	+				+	neuter noun

where a) — f) are the following homonymies:

- a) $(CN = CA)$;
- b) $(CG = CD) = (CN = CG = CD = CA)$;
- c) $(CN = CG) = (CD = CA)$;
- d) $(CN = CG = CD = CA)$;
- e) $(CG = CD)$;
- f) $(CN = CG = CD = CA) = (CN = CG = CD = CA)_{NP}$.

All the varieties of nouns examined above belong to the common nouns set.

The declination and articulation of proper nouns coincide with common nouns inflexing with analogous morphological categories [3].

So, the classification of all nouns was performed based on the affixes got. The nouns inflexion program implementation derives from this classification.

N	series affixes	Case	NS	NP	Example
1	vocală+u	CNA CGD CV	u u ule	uri uri urilor	tablou (picture)
2	consoană	CNA CGD CV	consoană consoană consoană+ule	uri uri urilor	tren (train)
3	u	CNA CGD CV	u u ule	uri uri urilor	lucru (work)
4	o	CNA CGD CV	o o oule	ale ale alelor	caro (caro)
5	e	CNA CGD CV	e e e	e e elor	n nume (name)

Figure 2: The Affixes Series Characterizing the Neuter Nouns

2.3. The Adjective

Following the defined noun, adjectives decline in gender, number and case.

M. Manoliu in [1] classifies adjectives in 19 large groups of inflexion. Our classification is even larger, since it contains 26 groups. Adjectives' inflexions forming process is correlated with the investigated affixes series.

In the adjective inflexion process *GF* gets 10 forms, the same as *GM*. If an adjective precedes a noun, then it takes a definite article and declines as the respective noun does.

If an adjective follows after a noun, then it changes as the (definite or indefinite) form and number of the noun do. Only feminine adjectives have special forms of case (*CGD*).

The adjective affix vocative coincides with the corresponding forms of masculine and feminine nouns.

Given the gender and studying the adjective affixes, homonymies of the types A-B are identified, namely:

Homonymies types	a	b	c	d	e	f	for adj. (gender)
A	+		+	+	+	+	fem.
B	+	+	+	+	+		masc.

where a) — f) are homonyms to nouns (see 2.2). The analysis above helped in implementing the adjective inflexion program.

3. The Consonant and Vowel Alternations in Speech Parts Inflexion Process

One of the particularities of declination and conjugation in the Romanian language is consonant and vowel alternation. It results from variations of roots and affixes. The noun and adjective declination makes use of several regularities in consonant and vowel alternations.

The sets of alternations are the following:

- for masculine nouns SM (A1) : $t \rightarrow \text{ț}$, $d \rightarrow z$, $s \rightarrow \text{ș}$, $x \rightarrow \text{cș}$, $\text{str} \rightarrow \text{ștr}$, $z \rightarrow j$, $l \rightarrow 0$, $\text{sc} \rightarrow \text{șt}$, $\text{ea} \rightarrow \text{a}$, $\text{ean} \rightarrow \text{en}$, $\text{ăt} \rightarrow \text{eț}$, $\text{înt} \rightarrow \text{inț}$, $\text{eaz} \rightarrow \text{ej}$, $\text{iac} \rightarrow \text{iec}$, $\text{ăn} \rightarrow \text{en}$, $\text{ian} \rightarrow \text{ien}$, $\text{ar} \rightarrow \text{ăr}$, $\text{ia} \rightarrow \text{ie}$ (where 0 is the empty affix);
- for feminine nouns SF (A2) : $\text{sc} \rightarrow \text{șt}$, $\text{șc} \rightarrow \text{șt}$, $\text{ea} \rightarrow \text{e}$, $\text{a} \rightarrow \text{e}$, $\text{ean} \rightarrow \text{en}$, $\text{ian} \rightarrow \text{ien}$, $\text{ăn} \rightarrow \text{en}$, $\text{ead} \rightarrow \text{ez}$, $\text{ăt} \rightarrow \text{eț}$, $\text{ăr} \rightarrow \text{er}$, $\text{ăd} \rightarrow \text{ed}$, $\text{ic} \rightarrow \text{el}$, $\text{t} \rightarrow \text{ț}$, $\text{d} \rightarrow \text{z}$, $\text{eat} \rightarrow \text{eț}$, $\text{ad} \rightarrow \text{ez}$, $\text{east} \rightarrow \text{eșt}$, $\text{asc} \rightarrow \text{ășt}$, $\text{easc} \rightarrow \text{eșt}$, $\text{ia} \rightarrow \text{ie}$, $\text{st} \rightarrow \text{șt}$, $\text{ar} \rightarrow \text{ăr}$, $\text{as} \rightarrow \text{ăș}$, $l \rightarrow 0$, $\text{oas} \rightarrow \text{os}$, $\text{tate} \rightarrow \text{tățî}$, $\text{oa} \rightarrow \text{o}$, $\text{ias} \rightarrow \text{ies}$, $\text{iaț} \rightarrow \text{ieț}$, $\text{eanc} \rightarrow \text{enc}$, $\text{ianc} \rightarrow \text{ienc}$, $\text{șanc} \rightarrow \text{șenc}$, $\text{în} \rightarrow \text{in}$, $\text{eal} \rightarrow \text{el}$, $\text{jeal} \rightarrow \text{jel}$, $\text{șeal} \rightarrow \text{șel}$, $\text{easc} \rightarrow \text{eșt}$;
- for neuter nouns SN (A3, A0) : $\text{ea} \rightarrow \text{e}$, $\text{înt} \rightarrow \text{int}$, $\text{ăn} \rightarrow \text{en}$, $\text{ăt} \rightarrow \text{et}$, $\text{o} \rightarrow \text{oa}$, $\text{ia} \rightarrow \text{ie}$;
- for masculine adjectives AM (A4) : $t \rightarrow \text{ț}$, $d \rightarrow z$, $s \rightarrow \text{ș}$, $x \rightarrow \text{cș}$, $\text{st} \rightarrow \text{șt}$, $\text{xt} \rightarrow \text{cșt}$, $z \rightarrow j$, $\text{sc} \rightarrow \text{șt}$, $\text{ea} \rightarrow \text{e}$, $\text{ian} \rightarrow \text{ien}$, $\text{ean} \rightarrow \text{en}$, $\text{ăn} \rightarrow \text{en}$, $\text{eat} \rightarrow \text{eț}$, $\text{ead} \rightarrow \text{ez}$, $\text{ăt} \rightarrow \text{eț}$, $\text{ăd} \rightarrow \text{ez}$, $\text{os} \rightarrow \text{oș}$, $\text{est} \rightarrow \text{eșt}$, $\text{oal} \rightarrow \text{ol}$, $l \rightarrow 0$;
- for feminine adjectives AF :
 - (A5) $\text{ăt} \rightarrow \text{eaț}$, $\text{o} \rightarrow \text{oa}$, $\text{os} \rightarrow \text{oas}$, $\text{eț} \rightarrow \text{eaț}$, $\text{or} \rightarrow \text{oar}$, $\text{el} \rightarrow \text{ea}$, $\text{esc} \rightarrow \text{easc}$, $\text{iesc} \rightarrow \text{iasc}$, $\text{ieț} \rightarrow \text{iaț}$, $\text{e} \rightarrow \text{ea}$;
 - (A6) $\text{at} \rightarrow \text{et}$, $\text{ăd} \rightarrow \text{ed}$, $\text{int} \rightarrow \text{int}$;
 - (A7) $\text{ian} \rightarrow \text{ien}$, $\text{sc} \rightarrow \text{șt}$, $\text{ea} \rightarrow \text{e}$, $\text{ean} \rightarrow \text{en}$, $\text{es} \rightarrow \text{eas}$, $\text{el} \rightarrow \text{ea}$, $\text{el} \rightarrow \text{ic}$, $\text{ăn} \rightarrow \text{en}$, $\text{esc} \rightarrow \text{eșt}$.

Consonant alternations are present in roots and affixes variations. The word root ends with a consonant or with a group of consonants, which, while being inflexed, are substituted for others.

Nb.	FI						FD					
	NS			NP			NS			NP		
C.	NA	GD	V	NA	GD	V	NA	GD	V	NA	GD	V
SM	—	—	A0	A3	A3	A3	—	—	A0	A3	A3	A3
SF	—	—	—	A1	A1	A1	—	—	—	A1	A1	A1
SN	—	A2	—	A2	A2	A2	—	A2	—	A2	A2	A2
AM	—	—	—	A4	A4	A4	—	A2	—	A4	A2	A2
AF	A5	A6, A7	—	A7	A7	A7	A5	A6, A7	—	A7	A7	A7

where A0 is the set of alternation for CV (affix — "le").

On examining the A1 — A7 sets, the following consonant alternations can be selected:

- t / ț ex.: *burete* — *bureți*, *boltă* — *bolți*;
- d / z ex.: *brad* — *brazi*, *coadă* — *cozi*;
- s / ș ex.: *urs* — *urși*;
- (ș)c / (ș)t, (ș)c / (ș)t ex.: *pușcă* — *puști*, *muscă* — *muște*;
- s(t) / ș(t), s(tr) / ș(tr) ex.: *artist* — *artiști*, *ministru* — *miniștri*;
- z / j ex.: *obraz* — *obraji*;
- l / i ex.: *cal* — *cai*, *gol* — *goi*;
- x / cș ex.: *sfinx* — *sfinși*, *fix* — *fiși*.

If comparing the above set of consonant alternations, based on the graphic principle, with that based on the phonetic principle [3], we will find out that they differ from one another by the following consonant alternations:

- c / ci ex.: *sărac* — *săraci*, *ac* — *ace*;
- g / gi ex.: *vargă* — *vergi*, *cîrlig* — *cîrlige*.

Being identical from a graphical point of view, the alternations will not be examined for automatic inflexion purpose.

Vowel alternations can be inside the root and in the affixes got in the process of word inflexion. They occur in both tonic (stressed) and atonic (unstressed) syllables. We have the following set of vowel alternations [3]. (Note that " ' " denotes the accent in the word).

1. The vowel alternations occurring in tonic syllables :

- a' / e' ex.: *băia't* — *băie'ți*, *șar'pe* — *șe'rpi*;
- ă' / e' ex.: *vă'r* — *ve'ri*, *mă'r* — *meri*;
- î' / i' ex.: *cuvi'nt* — *cuvi'nte*, *vi'nă* — *vi'ne*;
- a' / ă' ex.: *ca'le* — *că'i*, *la'mpă* — *lă'mpi*;
- oa' / o' ex.: *soa're* — *so'ri*, *moa'ra* — *mo'ri*;
- o' / oa' ex.: *ogo'r* — *ogoa're*, *ro'd* — *roa'de*;
- ea' / e' ex.: *crea'stă* — *cre'ste*; *sea'ra* — *se'ri*;
- î' / îi' ex.: *mî'nă* — *mîi'ni*;
- e' / i' ex.: *fugă're'ț* — *fugă'ri'ță*;
- e' / ea' ex.: *des* — *dea'să*.

2. The vowel alternations occurring in atonic syllables:

- ă / e ex.: *sî'mbătă* — *sî'mbete*;
- o / oa ex.: *mijlo'c* — *mijloa'ce*.

3. The vowel alternations presupposing a change of the accent place :

4. o' / u ex.: *so'ra* — *suro'ri*, *no'ra* — *nuro'ri*.

If atonic or tonic syllables in a word can be procedurally determined, then the problem of the word inflexion process automation, from the vowel alternations point of view, can be solved for good. Unfortunately, it is difficult (or even impossible) to indicate a procedure of determining stressed or unstressed syllables. Vowel alternations in word inflexing are analysed.

The set of vowel alternations can be divided into three groups: absolute regularities, partial regularities and irregularities.

The following will be absolute regularities:

- common to nouns and adjectives: ea → e, ia → ie;
- for nouns: ă → e;
- for neuter nouns: o → oa;
- for adjectives: o → oa, e → ea, ie → ia.

The following are partial regularities:

- common to nouns and adjectives: î → i;
- for feminine nouns: a → ă, a → e, oa → o.

An user will deal with irregularities in the process of the proposed system inflexion correction. For example, the word "om" will be declined by the system: *FI : NS : CNA — om, CGD — om, CV — om;*

NP : CNA — omi, CGD — omi CV — omi;
FD : NS : CNA — omul, CGD — omului, CV — omule;

NP : CNA — omii, CGD — omilor, CV — omilor.

The user will make the corrections for *FI, NP, CNA, CGD* and *CV*: *omi* will be changed in *oameni*, but for *FD NP, CNA, CGD* and *CV*: *omii* and *omilor* will be changed in *oamenii* and *oamenilor*, respectively.

Some root may be affected by both vowel and consonant alternations. They are the following:

Alt.vow. alt.cons.	a'/e'	ă'/e'	î'/i'	a'/ă'	o'/oa'	e'/ea'	ă/e	e/a
t/ț	+		+	+	+	+		+
d/z	+			+	+			
s/ș					+	+	+	
st(r)/șt(r)					+			+
sc/șt						+		
z/j						+		
l/				+		+		

4. The Procedure of Verb Inflexion

We will now refer the procedure of verb inflexion, taking the activity of one rather complicated module, that generates inflexions for the

first grammatical group of verbs, of which conjugation needs no suffix.

Every procedure work is reflected in the list *L* of complicated structures, where information on different parts of speech is provided.

Most frequently list *L* includes 35 elements (**Infinitivul** — 1, **Indicativ Prezent** — 6, **Perfectul Simplu** — 6, **Mai mult ca Perfectul** — 6, **Imperfectul** — 6, **Conjunctiv Prezent** — 6, **Imperativul** — 2, **Participiul** — 1, **Gerunziul** — 1). Some verbs of **Imperativ Singular** or **Plural** have more than 2 forms. For example, **Imperativ Singular** of the verb *a veni* (to come) has the forms *vino, vin, vină*, and **Imperativ Plural** of the verb *a naște* (to be born) has the forms: *nașteți, nașteți*.

There is a similar situation with **Conjunctiv Prezent**, III person, for some verbs having 2 forms. For example, the verb *a da* (to give): *să dea, să deie*. For impersonal verbs, there are only 12 inflexions.

The implementation of procedures using some functions of the Turbo-Pascal language meant for strings updating, is described :

- Length (*S*) — indicates the string *S* length;
- Copy (*S, n, k*) — gives substring of length *k* from string *S* in position *n*;
- Insert (*S1, S, n*) — the string *S1* is inserted in the string *S* in position *n*;
- Delete (*S, n, k*) — in the string *S* of position *n k* characters are deleted.

This will do for the realization of the schemes of conjugation. We introduce the notion of "word root".

Definition 1 The word root is calculated by the formula :

$R = \text{Copy}(S, 1, \text{Length}(S) - k)$,
where *S* is the string which contains The conjugation verb Infinitive,
k = 2 for the verbs set which is the element of a second grammatical group (ending in "ea"),
k = 1 for the rest of the verbs.

If during the verbs conjugation process, word root consonant or vowel alternations turn in, then some roots denoted by *R, R1, R2, R3, ...* are obtained.

The set of affixes consists of 6 strings, which are attached to the corresponding root for I —

VI persons (I — III make the singular and IV — VI — make the plural) for getting inflected. For example, in order to obtain the conjugation scheme of the verb *a cânta* (to sing) $R = \text{"cânt"}$, the affixes set forming **Indicativ Prezent** will be (" ", "i", "ă", "ăm", "ați", "ă"), but in forming, for example, **Indicativ Imperfect**, the affixes set ("am", "ai", "a", "am", "ați", "au", etc.) will be needed.

The conjugation scheme of the verb *a căuta* (to seek) will produce the roots $R = \text{"căut"}$ and $R1 = \text{"caut"}$. Some remarks should be made. One question that might arise is the following: why do we only use one root of the verb *a cânta* (to sing) on updating inflexions, when there are other roots as well: *cînt* (*eu cînt, tu cînti*)? The answer is the following: such consonant alternations as $t \rightarrow \text{ț}$, $d \rightarrow z$, $sc \rightarrow \text{șt}$, $st \rightarrow \text{șt}$, $șc \rightarrow \text{șt}$, which need be executed for getting the II person flexions, **Indicativ Prezent Singular** and **Conjunctiv** (sometimes, the III person flexions, for example, *eu urăsc, tu urăști, el urăște*), are automated by two auxiliary procedures. It is possible that these alternations should be executed since they are independent of the conjugated verb scheme. They only depend on the presence of one of the left parts for I person, **Indicativ Prezent Singular**. These are unique alternations, not requiring the introduction of a new root.

Note the difficulties to be encountered in executing the conjugation schemes. There are many exceptions even in one conjugation scheme. For example, the verb *a scădea* (to subtract) is conjugated as the verb *a cădea* (to fall), but **Imperativ Singular** of the verb *a cădea* coincides with **Indicativ Prezent Singular**, II person — *cazi*, while the verb *a scădea* has **Imperativ Singular** — *scade*, which coincides with **Indicativ Prezent Singular**, III person.

In programming the given conjugation scheme, all these exceptions have been considered. Unfortunately, these exceptions are not always referred in [6]. Some collisions are deleted by consulting [5, 6], ORTO [7] or philologists. Another difficulty to be coped with in executing the verbs inflexion procedures, is how to formalize the verbs set description as an element of one conjugation scheme or another. The notion of "root predaffix" is introduced here and an example of this formalization is given.

Definition 2 *Predaffix with the length n ($n \leq \text{Length}(R)$) is the substring Copy ($R, \text{Length}(R) - (n - 1), n$), that is S is the root R "tail" of length n .*

The verbs *a tăia* (to cut), *a înfoia* (to swell), *a muia* (to soak), *a apropia* (to approach) have a lot of common properties, but they differ strongly as far as their conjugation schemes are concerned. The schemes converge into one case.

1. If predaffix of the length 1 is constant "I", then 2 affixes subsets variants appear in forming **Indicativ Prezent** (we are examining it, because other flexion tenses forms can be achieved very easily):

if $R[\text{Length}(R) - 1]$ does not coincide with one of the characters "a", "u", "o", "e", then the following set of affixes is used: "i", "i", "e", "em", "ați", "e";

else (if it does) — the following set: " ", " ", "e", "em", "ați", "e" is used. If the set of roots needs be clarified, then the definition of root R is to be remembered, considering the following subcases:

- if predaffix of the length 3 coincides with string "tăi" (for example, the verbs *a tăia*, *a străția*, etc.), then we get one new root ($R1$), substituting the last but one vowel "ă" for "a";
- if $R[\text{Length}(R) - 1]$ coincides with character "o", then we get $R1 = \text{Insert("a", } R, \text{Length}(R))$;
- if $R[\text{Length}(R) - 1] = \text{"u"}$, but $R[\text{Length}(R) - 2]$ does not coincide with some of the characters "c", "e", "ț" and $R \neq \text{"împui"}$, the root $R1$ is got on substituting the character "u" for "o";
- the rest of verbs in this case has one root.

2. How to formalize the verbs set which conjugates in accordance with the verb scheme *a afla* (to hear) and forms **Indicativ Prezent**, is considered. The formalization of this set is possible in the following way. If the predaffix of length 2 coincides with one of the strings "FL", "CR", "PL", "NU", "TR", "BL", "RL", "TU" (for example, *a însufla* (to inspire), *a consacra* (to consecrate), *a contempla* (to contemplate), *a intra* (to enter), *a umbla* (to walk), *a urla*

(to howl), *a continua* (to continue) and so on), then the affixes set will be : "u", "i", "ă", "ăm", "ați", "ă". The roots set (except for the verb *a lătra* (to bark), which has 2 roots : $R = "lătr"$ and $R1 = "latr"$) consists of one single element. For the verb *a lătra* (to bark), 2 roots are obtained.

3. The verbs set which conjugates in accordance with the conjugation scheme of the verb *a apăsa* (to press), has to satisfy the following condition. The predaffix of length 3 must coincide with one of the strings: "PĂS", "FĂS", "FĂT", "BĂR", "MĂT", "BĂT", "VĂȚ", "FĂȚ", "ĂRS", "PĂL", or "SĂL". For example, the verbs *a învăța* (to study), *a înfășa* (to swaddle), *a vărsa* (to spill), and so on.
4. The verbs set of which conjugation schemes coincide with those of the verb *a căpăta* (to receive), satisfies the following condition :

$(Length(R) > 3)$ and $(R[Length(R) - 1] = "ă")$ and
 $((R[Length(R) - 3] = "ă")$ or $(R[Length(R) - 4] = "a"))$.

Such verbs are *a se cățăra* (to climb), *a vătăma* (to harm), *a scărmana* (to snatch), *a scăpăra* (to sparkle), and so on.

5. The verbs of which conjugation schemes coincide with those of the verb *a asemăna* (to compare), have to satisfy the condition: the predaffix of length 3 should coincide with one of the following strings: "PĂD", "MĂN", "TĂN", "PĂN", "MĂT", "ZEM", "PĂR", "RĂȚ", "MĂR".

Examples of such verbs are: *a depăna* (to spin), *a fremăta* (to rustle), *a pieptăna* (to comb), *a rezema* (to lean), *a lepăda* (to drop), *a legăna* (to rock), and so on.

In an analogous way, other 5 sets, having to satisfy vaster conditions, can be studied, but we do not indicate them here.

The rest of the verbs of the first grammatical group (which does not satisfy the conditions of cases I — X), conjugates in accordance with the conjugation scheme of the verb *a cânta* (to sing).

We need say that, as part of one and the same conjugation scheme, that is, as part of each of the 10 cases, subcases are imagined for

verbs specific conjugation subgroups, entering this case because of given criteria. Such verbs as *a se cățăra* (to climb), *a se văicări* (to lament), *a vătăma* (to harm), *a tărăgăna* (to dally), and perhaps others, belonging to the II case, for **Indicativ Present**, II person **Singular**, do not substitute the last but one vowel "ă" for "e" (*cațări*, *valămi*, and so on), but, in this case, the rest of the verbs has the vowel alternation $ă \rightarrow e$ (*capeți*, *scarmeni*).

The formalization of cases and subcases is a rather difficult problem, but if we consider the development and extension of a natural language, using different forms belonging to the same grammatical categories (for example, *comenzi* and *comanzi* referred to by different authors), then, as the natural working procedure goes, all the verbs flexions (or other parts of speech) and definite corrections made by an user, will be displayed on computer monitor.

5. The Nouns and Adjectives Inflexion Algorithm

The inflexion algorithm of nouns and adjectives is the following.

The inflexion word is introduced, and the part of speech the word represents, is indicated. If the word is a noun, then gender and number are specified.

The basic word is indicated for this word — *CNA* and *NS*. Splitting the word into a root and an affix, is made in the following way. The affixes specified (Figure 2) are arranged in a decreased order of length. For each affix, a special procedure of noun inflexion is determined. The affix in a word is found in the decreased order of length. If it coincides with the one specified, then it calls for the corresponding procedure of inflexion. In an automatic inflexion process, it should produce a consonant alternation, as shown below:

- t / ț ex. : *burete* — *bureți*, *boltă* — *bolți*;
- d / z ex. : *brad* — *brazi*, *coadă* — *cozi*;
- s / ș ex. : *it urs* — *urși*;
- (ș)c / (ș)t, (s)c / (ș)t ex. : *pușcă* — *puști*, *muscă* — *muște*;
- s(t) / ș(t), s(tr) / ș(tr) ex. : *artist* — *artiști*, *ministru* — *miniștri*;

- z / j ex. : obraz — obraji;
- l / i ex. : cal — cai, gol — goi;
- x / cș ex. : sfinx — sfincși, fix — ficși.

While the vowel alternations, which are absolute regularities for both nouns and adjectives (ea → e, ia → ie), happen with no dialogue, the alternation î → i, which is a partial regularity, requires a dialogue. (For example, *tînăr — tineri, pămînt — pămînturi*).

The vowel alternation, which is characteristic to nouns, and represents an absolute regularity, (in most of the cases) is : ă → e.

For feminine nouns, the following vowel alternations, a partial regularity, can be distinguished: a → ă a, → e, oa → o. For example, *cană — căni, masă — mese, floare — flori*, but *casă — case, masă — mase, coasă — coase*.

The masculine nouns have vowel alternations, which are absolute regularities only, and the above-mentioned consonant alternations (see above).

The following vowel alternations, an absolute regularity, can be specified for neuter nouns: o → oa. For example, *cojoc — cojoace*.

If a word is an adjective, then it indicates the basic word — *GM, CNA* and *NS*. Splitting a word into a root and an affix is similar to splitting a noun. All the adjective affixes are arranged in the decreased order of their length. The procedure of selecting all possible affixes is analogous to that applied to a noun. If an affix coincides with the specified one, then we call for the corresponding procedure for adjective declination. In an automatic inflexion process, consonant alternations occur (see above).

For adjectives, the following vowel alternations can be distinguished: o → oa, e → ea, ie → ia. These alternations are frequent in adjective inflexion (feminine gender forms).

For example, *frumos — frumoasă, des — deasă, măiestru — măiastră*.

This information and additional information will be required for enriching a dictionary, and for a spelling-checker. The dictionaries [6, 7] are the dictionaries we worked with.

REFERENCES

1. TOȘA, A., *Elemente de morfologie* (Morphological Elements), Scientific and Encyclopaedic Publishing House, Bucharest 1983.
2. ZDRENGEA, M., *Limba română contemporană. Morfologia* (Contemporary Romanian Language. Morphology), Cluj 1977.
3. GUȚU-ROMALO, V., *Morfologie structurală a limbii române* (Structural Morphology of the Romanian Language), The Academic Publishing House, Bucharest 1968.
4. IORDAN, I., GUȚU-ROMALO, V. and NICULESCU, A., *Structura morfologică a limbii române contemporane* (Morphological Structure of the Contemporary Romanian Language), Scientific Publishing House, Bucharest 1967.
5. IRIMIA, D., *Verbul* (The Verb), Junimea Publishing House, Jassy 1977.
6. LOMBARD, A. and GADEI, C., *Dictionnaire morphologique de la langue roumaine*, The Academic Publishing House, Bucharest 1981.
7. ORTO — *Dicționar ortografic cu elemente de ortoepie și morfologie* (An Orthographical Dictionary, Also Including Orthoepy and Morphology Elements), Editors-in-Chief of the Soviet Moldavian Encyclopaedia, Kishinev 1990.