Modelling Inflections in Romanian Language by P Systems with String Replication

Artiom Alhazov[1,2], Elena Boian[1], Svetlana Cojocaru[1], Yurii Rogozhin[1,3]

[1] Institute of Mathematics and Computer Science
Academy of Sciences of Moldova
Academiei 5, Chişinău MD-2028 Moldova
{artiom,lena,sveta,rogozhin}@math.md

[2] IEC, Department of Information Engineering
Graduate School of Engineering
Hiroshima University, Higashi-Hiroshima 739-8527 Japan

[3] Research Group on Mathematical Linguistics, Rovira i Virgili University
Av. Catalunya, 35, Tarragona 43002 Spain

Summary. The aim of this article is the formalization of inflection process for the Romanian language using the model of P systems with cooperative string replication rules, which will make it possible to automatically build the morphological lexicons as a base for different linguistic applications.

1 Introduction

Natural language processing has a wide range of applications, the spectrum of which varies from a simple spell-check up to automatic translation, text and speech understanding, etc. The development of appropriate technology is extremely difficult due to the specific feature of multidisciplinarity of the problem. This problem involves several fields such as linguistics, psycholinguistics, computational linguistics, philosophy, computer science, artificial intelligence, etc.

As in many other fields, solving of a complex problem is reduced to finding solutions for a set of simpler problems. In our case among the items of this set we find again many traditional compartments of the language grammar. The subject of our interest is the morphology, and more specifically, its inflectional aspect.

The inflectional morphology studies the rules defining how the inflections of the words of a natural language are formed, i.e., the aspect of form variation (of the inflection, which is the action of words modification by gender, number, mood, time, person) for various expressing grammatical categories.

In terms of natural language typology the morphological classification can be *analytical* and *synthetic*. Of course, this classification is a relative one, having, however, some irrefutable poles: Chinese, Vietnamese, as typical representatives of the analytical group, and Slavic and Romance languages serving as examples of synthetic ones. The English language, with a low degree of morpheme use, is often among the analytical ones, sometimes is regarded as synthetic, indicating however that it is “less synthetic” comparatively with other languages from the same group. It is evident that it is the inflectional morphology of synthetic languages that presents special interest, being a problem more complex comparatively with analytical class.

The object of our studies is the Romanian language, which belongs to the category of synthetic flective languages. The last notion stresses the possibility to form new words by declension and conjugation. Moreover, the Romanian language is considered a highly inflectional language, because the number of word-forms is big enough.

The inflection simplicity in English makes that the majority of researchers in the field of computational linguistics neglect the inflection morphology. For efficient processing of other natural languages, including Romanian, it is necessary to develop suitable computational models of morphology of each language. In the case of Romanian language, some inflectional models are known [24], [18],[6].

In [24] it is certified an advanced number of morpho-syntactic specifications for Romanian language, namely 34 for nouns, 44 for verbs, 24 for adjectives, 15 for pronouns, etc. The aim of our paper is to describe the process of inflection (i.e. the process of obtaining both the derivative words and their morphological attributes) by P systems [16].

2 Description of the inflection process

To develop a formalism for the inflection process description we invoke a number of definitions and notions which allow us to understand the essence of this process. Inflection is a part of morphology - the science which “includes the rules considering the word forms and the formal modifications of the words” [23]. From the morphological point of view the words are classified corresponding to the part of speech, and their structure is described in terms of inflection, derivation and composition. Inflection is the systematic variation of the word form which allows to obtain different semantic and syntactic functions [9]. The words combine in themselves two components: a *constant* and a *variable* [11]].

The root of primary lexical units is called the *constant.* For the derivative ones the term *lexical theme* is used. Since in our study this distinction does not play any role, for both cases we use a single term “*root*”.

The *variable* is the bearer of grammatical meanings, it consists of one or more morphemes being called also *flective.* This term will be used in exposure below. In accordance with [23] we identify three ways of achieving the inflections:

analytical: the flective is a free morpheme (separated from root) and the root remains invariable (e.g., adverb, *bine* – *mai bine* (engl. well - better));

synthetic: the flective is a conjunctive morpheme (group of morphemes), related to the root (e.g., for noun, pronoun; *studentă – studente – studentei; care-căreia-căruia-cărora* (engl. *student – students – student's, who-whose-whom*), etc.).

synthetic and analytical: the flective consists of free and conjunctive morphemes (e.g., adjective, verb, *frumos – frumoasă – mai frumoasă; cântasem – am cântat* (engl. beautiful – beautiful – more beautiful, singing – I sang), etc.).

In the following we will deal with the synthetic method, the analytical one is effectuated relatively easy through a set of simply formulated rules. Following the model from [9] we present in Figure 1 the classification of Romanian language parts of speech in terms of the inflection process.

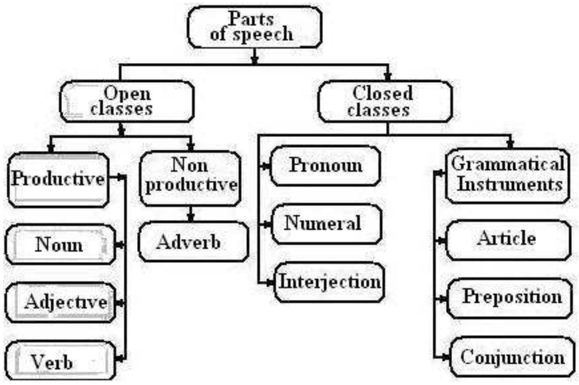

Fig. 1. The classification of the Romanian language parts of speech (in terms of the inflection process.)

The class of opened productive parts of speech is the most interesting in terms of inflection, and it will be the primary object of our investigations.

Indeed, opened classes, containing tens of thousands of elements, are characterized by a productive process of inflection, derivation and composition, while the closed ones include a reduced number of items (practically excluding the possibility of the new ones apparition), because the morphological processes of word formation are poorly productive [12]. Moreover, in the case of opened classes the problem is complicated not only because we cannot enumerate the elements, existing at the moment, but also because a successful formalism should be able to "serve" the future neologisms that could occur in language development process. In the following we will operate with the paradigms of inflection, by which we imply the systematic arrangement of all inflection forms of a word [13].

For our purposes we will work not with the whole words, but with their variable parts. Hereinafter by paradigm we mean a list of flectives.

For each flective we can put into correspondence a set of morphological attributes.

Example. Let us examine the morphological attributes for masculine nouns of Romanian language [24].

N m	noun (part of speech), masculine gender,
s p	singular number, plural number,
d o v	direct (nominative – accusative cases), oblique (genitive – dative cases), vocative case,
y n	yes – definiteness, no – definiteness.

(Given that the Romanian forms for nominative and accusative cases coincide, as well as for the genitive and dative ones, we reduced the paradigm merging both word forms, and respective attributes.)

Thus, the list of flectives $F = \{-, -, -, \text{ul, ului, ule, i, i, i, ii, ilor, ilor}\}$, where "$-$" denotes the empty word, can be regarded as a morphologically annotated one.

$$\begin{aligned} F_{morf} = \{\ & (-, Nmsdn), (-, Nmson), (-, Nmsvn), \\ & (\text{ul}, Nmsdy), (\text{ului}, Nmsoy), (\text{ule}, Nmsvy), \\ & (\text{i}, Nmpdn), (\text{i}, Nmpon), (\text{i}, Nmpvn), \\ & (\text{ii}, Nmpdy), (\text{ilor}, Nmpoy), (\text{ilor}, Nmpvy)\}. \end{aligned}$$

Let us mention the use of paradigmatic model for the Romanian language [7, 8, 19, 20, 21].

We will refer also to the works [17] and [10], which treat the subject of generation of the flectioned forms for the Romanian language. The authors do not provide the inflection algorithms, but offer some useful suggestions for generation of flectioned forms. In paper [17] it is proposed a method of encoding vowel and consonant alternations in the root, taken by the authors from researches of acad. G. Moisil, namely: each alternation is presented in the root by a distinct code. In paper [10] it is found a (incomplete) set of rules, which indicates the way of concatenation of flective for nouns and adjectives without concerning the problem of the alternations in the root. Therefore, having the aim to achieve the synthetic model of inflection, we must develop a formalism, which should include two processes:

- making the alternation in the root, and
- concatenation of a flective.

The starting point of our approach was the dictionary [12], in which the flective words of Romanian language are classified according to the way of inflections formation. There were set 100 groups of inflection for masculine nouns, 273 – for verbs, etc. A dictionary of about 30,000 words with the specification of the number of the group was constructed. The classification was made taking into account all linguistic aspects, e.g. accents. In our case we will focus only on the way of writing a word, which in equal measure simplifies and complicates the problem. However this classification is extremely useful suggesting us the idea of defining a special class of grammars to formalize the inflection process [1, 2, 3, 4].

In general case, from a whole variety of inflection groups, we can identify two classes:

– without alternations, and

– with alternations.

In the first case the inflection is made in the following manner. Let $\Im$ be a set formed from lists of flectives, $F = \{f_1, f_2, \cdots, f_n\}$, $w = w'\alpha$ is a word-lemma, where $|\alpha| \geq 0$. In the simplest case the inflected words will be those of the form $w'f_i$, $f_i \in F$, $(i = 1, \cdots, n)$.

General case: Let $w = w_1 a_1 w_2 a_2 \cdots w_m \alpha$. The inflected words will be of the form:

$$\begin{array}{l} w^{(1)} = w_1 \;\; a_1 \;\; w_2 \;\; a_2 \;\; \cdots w_m f_{i_1}, \\ w^{(2)} = w_1 \;\; u_1^{(2)} \;\; w_2 \;\; u_2^{(2)} \;\; \cdots w_m f_{i_2}, \\ \quad \cdots \\ w^{(s)} = w_1 \;\; u_1^{(s)} \;\; w_2 \;\; u_2^{(s)} \;\; \cdots w_m f_{i_s}, \end{array}$$

where w_i, $a_i \in V^+$, $u_i^{(j)} \in V^*$, $f_{i_1} \in F^{(1)}, \ldots, f_{i_s} \in F^{(s)}$, and $F^{(1)} \cup \ldots \cup F^{(s)}$ forms a complete paradigm.

Note: the analysis of inflection rules allowed us to ascertain that for the Romanian language $m \leq 4$, $s \leq 3$.

Example 1. Inflection of masculine nouns without alternations.

Let $F = \{-, -, -, ul, ului, ule, i, i, i, ii, ilor, ilor\}$ – a list of flectives, where '-' denotes the empty word. Let $w =$'stejar' (engl. *oak*), $|\alpha| = 0$, $|F| = 12$. The set of inflected words supplied by morphological attributes will be:

$$\begin{array}{llll} \{ & (\text{stejar}, Nmsdn), & (\text{stejar}, Nmson), & (\text{stejar}, Nmsvn), \\ & (\text{stejarul}, Nmsdy), & (\text{stejarului}, Nmsoy), & (\text{stejarule}, Nmsvy), \\ & (\text{stejari}, Nmsdn), & (\text{stejari}, Nmpon), & (\text{stejari}, Nmpvn), \\ & (\text{stejarii}, Nmpdy), & (\text{stejarilor}, Nmpoy), & (\text{stejarilor}, Nmpvy) \; \}. \end{array}$$

Taking advantage of paradigmatic ordering of the elements from the list of flectives, in what follows we will omit the explicit writing of morphological attributes implying their conformity to respective flectives.

Example 2. Inflection of masculine nouns with alternations.

Let $w =$*tânăr* (engl. *young*), $|\alpha| = 0$. The vowel alternations *â*→ *i* and *ă*→ *e* will be used. The obtained roots $w =$*'tânăr'* and $w' =$*'tiner'* are respectively annexed by the endings: $F_1 = \{-, -, \text{ul, ului, ule}\}$ and $F_2 = \{e, \text{i, i, i, ii, ilor, ilor}\}$, $|F_1| + |F_2| = 12$.

{ (tânăr, $Nmsdn$), (tânăr, $Nmson$), (tânărule, $Nmsvy$),
(tânărul, $Nmsdy$), (tânărului, $Nmsoy$), (tinere, $Nmsvn$),
(tineri, $Nmsdn$), (tineri, $Nmpon$), (tineri, $Nmpvn$),
(tinerii, $Nmpdy$), (tinerilor, $Nmpoy$), (tinerilor, $Nmpvy$) }.

Note: In most cases (for 80 groups of inflexion from [12]), when declining the masculine noun, 12 words are obtained. Exceptions are the following nouns:

– irregular, for example, those which can not have the plural definite form (instance, the word *gnu*);

– those which are singularia tantum (nouns which appear only in the singular form), *ianuarie* etc.;

– those which are pluralia tantum (nouns that appear only in the plural and do not have a singular form), for example, *ochelari, pantaloni* etc.

In general, the 100 groups of inflection of masculine nouns in relation to the number of words produced at inflection, present the following table:

Forms of the lemma	Number of forms	Number of groups
all forms	12	80
singularia tantum	6	13
pluralia tantum	6	4
irregular	6-8	3

Modern dictionaries contain hundreds of thousands of words–lemma. Their forms of inflexion (the amount of which exceeds millions) are needed for developing various applications based on natural language: from the spell-checker up to the systems understanding the speech. Obviously, to solve the problem of creating a dictionary with a morphologically representative coverage, as well as to build various applications based on it, effective mechanisms are needed, especially those that allow parallel processing. One of the possible ways to perform parallel computation is based on biological models.

Let us mention a series of works that used the biological calculation approaches for solution of linguistic problems. In [14] are presented some attempts to construct linguistic membrane systems and some applications related to analyze of conversational acts, a bio-inspired for dealing with semantics. In [15] two parsing methods using P automata are presented. The first method uses P automata with active membranes for parsing natural language sentences into dependency trees. The second method uses a variant of P automata with evolution and communication rules for parsing Marcus contextual Languages[13].

Our paper tries to expand the area of potential applications of P systems to linguistics problems, introducing a formalism to capture inflections with their morphological attributes.

To formalize the inflection process for the Romanian language the model of cooperative membrane P systems with replication will be used [16].

3 P systems with string replication and input

Let us recall the basics of P systems with string objects and input. The membrane structure μ is defined as a rooted tree with nodes labeled $1, \cdots, p$. The objects of the system are strings (or words) over a finite alphabet O. A sub-alphabet $\Sigma \subseteq O$ is specified, as well as the input region i_0, $1 \leq i_0 \leq p$. In this paper we need to use cooperative rewriting rules (i.e. string rewriting rules, not limited by context-free ones) with string replication and target indications.

A rule $a \to u_1$, where $a \in O^+$ and $u_1 \in O^*$, can transform any string of the form $w_1 a w_2$ into $w_1 u_1 w_2$. Application of a rule $a \to u_1 || u_2 || \cdots || u_k$ transforms any string of the form $w_1 a w_2$ into the multiset of strings $w_1 u_1 w_2$, $w_1 u_2 w_2$, $\cdots$, $w_1 u_k w_2$. If in the right side of the rule (u_i, t) is written instead of some u_i, $1 \leq i \leq k$, $t \in \{out\} \cup \{in_j \mid 1 \leq j \leq p\}$, then the corresponding string would be sent to the region specified by t.

Hence, such a P system is formally defined as follows:

$$\begin{aligned} &\Pi = (O, \Sigma, \mu, M_1, \cdots, M_p, R_1, \cdots, R_p, i_0), \text{ where} \\ &M_i \quad \text{is the multiset of strings initially present in region } i, 1 \leq i \leq p, \\ &R_i \quad \text{is the set of rules of region } i,\ 1 \leq i \leq p, \\ &\qquad \text{and } O, \Sigma, \mu, i_0 \text{ are described above.} \end{aligned}$$

The initial configuration contains the input string(s) over Σ in region i_0 and strings M_i in regions i. The computation consists in non-deterministic application of the rules of a region to a string of that region, in parallel to all the strings in the system. The computation halts when no rules are applicable. The result of the computation is the set of all words sent out of the outermost region (called skin).

4 Describing the inflection process by P systems

Let us define the P system performing the inflection process. Let L be the set of words which form opened productive classes. We will start by assuming that the words in L are divided into groups of inflection, i.e. for each $w \in L$ the number of inflection group is known [12]. The inflection group is characterized by the set $G = \{\alpha, R_G, F_G\}$, where $|\alpha| \geq 0$ is the length of ending which is reduced in the process of inflection, F_G is the set of the lists of flectives, the assembly of which forms complete paradigm, R_G is the set of the rules, which indicate vowel/consonant alternation of type $a \to u$, $a \in V^+$, $u \in V^*$, and also the conformity of the roots obtained by the lists of flectives from F_G. To each group of inflexion a membrane system Π_G will be put into correspondence.

As it was mentioned earlier, we will investigate two cases:
– without alternations, and
– with vowel/consonant alternation.

The first model is very simple. For any group $G = (\alpha, \emptyset, \{f_{1_G}, f_{2_G}, \cdots, f_{n_G}\})$ of inflection without alternation,

$$\begin{aligned} \mathit{\Pi}_G &= (O, \Sigma, [\ \ ]_1, \emptyset, R_1, 1), \text{ where} \\ O &= \Sigma = V \cup \{\#\}, \\ V &= \{\mathrm{a}, \cdots, \mathrm{z}\} \text{ is the alphabet of the Romanian language, and} \\ R_1 &= \{\alpha\# \to (f_{1_G}, out)||(f_{2_G}, out)|| \cdots ||(f_{n_G}, out)\} \end{aligned}$$

If this system receives as an input the words $w'\alpha\#$, where $w'\alpha$ corresponds to the inflection group G, then it sends all its inflected words out of the system in one step. Clearly, $\mathit{\Pi}_G$ is non-cooperative if $\alpha = \lambda$, but non-cooperativeness is too restrictive in general, since then the system would not be able to distinguish the termination to be reduced from any other occurrence of α.

The general model will require either a more complicated structure, or a more sophisticated approach. Let G be an arbitrary inflection group, with $m-1$ alternations $a_1 = a_1^{(1)} a_2^{(1)} \cdots a_{n_1}^{(1)}, \cdots, a_m = a_1^{(m)} a_2^{(m)} \cdots a_{n_m}^{(m)}$. Let the set of flectives consist of s subsets, and for subset $F_{k_G} = \{f_1^{(k)}, \cdots, f_{p_1}^{(k)}\}$, $1 \le k \le s$, the following alternations occur: $a_1 \to u_1^{(k)}$, $\cdots$, $a_m \to u_m^{(k)}$ (the alternations are fictive for $k = 1$), and $\bigcup_{k=1}^{s} F_{k_G}$ corresponds to a complete paradigm. For instance, Example 2 corresponds to $s = 2$ sublists (singular and plural), and $m - 1 = 2$ alternations.

The associated P system should perform the computation

$$w\# = \prod_{j=1}^{m-1} (w_j a_j)\, w_m \alpha\# \Rightarrow^* \left\{ \prod_{j=1}^{m-1} \left(w_j u_j^{(k)} \right) w_m f_{i_k} \mid 1 \le k \le s,\ f_{i_k} \in F^{(k)} \right\},$$

where $u_j^{(1)} = a_j$, $1 \le j \le m$.

The first method is assuming the alternating subwords a_j are present in the input word in just one occurrence, or marked. Moreover, we assume that carrying out previous alternations does not introduce more occurrences of the next alternations.

For modeling such process of inflection for the group G we define the following P system with $1 + (s-1)m$ membranes

$$\begin{aligned} \mathit{\Pi}'_G &= (O, \Sigma, \mu, \emptyset, \cdots, \emptyset, R_1, \cdots, R_{1+(s-1)m}, 1), \text{ where} \\ \Sigma &= V \cup \{\#\}, \\ O &= \Sigma \cup E, \\ \mu &= [\,[\ \ ]_2 [\ \ ]_3 \cdots [\ \ ]_{1+(s-1)m}\,]_1, \\ E &= \{\#_k \mid 2 \le k \le s\} \cup \{A_{k,j} \mid 1 \le k \le s,\ 1 \le j \le m\}, \\ V &= \{\mathrm{a}, \cdots, \mathrm{z}\} \text{ is the alphabet of the Romanian language,} \end{aligned}$$

(V can be extended by marked letters if needed), and the rules are given below.

$$\begin{aligned} R_1 &= \{\alpha\# \to A_{1,m}||(\#_2, in_2)|| \cdots ||(\#_s, in_s)\} \\ &\cup \{A_{k,j} \to (\lambda, in_{k+(s-1)j}) \mid 2 \le k \le s,\ 1 \le j \le m-1\} \\ &\cup \{A_{k,m} \to (f_1^{(k)}, out)|| \cdots ||(f_{p_m}^{(k)}, out) \mid 1 \le k \le s\}, \\ R_{k+(s-1)(j-1)} &= \{a_j \to (u_j^{(k)} A_{k,j}, out)\},\ 2 \le k \le s,\ 1 \le j \le m-1, \\ R_{k+(s-1)(m-1)} &= \{\#_k \to (A_{k,m}, out)\},\ 2 \le k \le s. \end{aligned}$$

The work of P system Π'_G is the following. First, s copies of the string are made, and the first one stays in the skin, while others enter regions $2, \cdots, s$. Each copy in region k is responsible to handle the k-th subset of inflections. The first one simply performs a replicative substitution in the end, and sends the results out, in the same way as Π_G works. Consider a copy of the input in region k, $2 \le k \le s$. When j-th alternation is carried out, the string returns to the skin, and symbol $A_{k,j}$ is additionally produced. This symbol will be used to send the string in the corresponding region to carry out alternation $j+1$. Finally, if $j = m$, then the system performs a replicative substitution in the end, and sends the results out.

Assuming $s \ge 2$, the system halts in $2m+1$ step, making an efficient use of scattered rewriting with parallel processing of different inflection subsets. For instance, the inflection group from Example 2 would transform into a P systems with 4 membranes, halting in 7 steps. Notice that this system is non-cooperative if $\alpha = \lambda$ and $|a_j| = 1$, $1 \le j \le m$. It is also worth noticing that it is possible to reduce the time to $m+1$ steps by using tissue P systems with parallel channels.

The second method avoids the limiting assumptions of the first methods. More exactly, it performs the first alternation at its leftmost occurrence, the second alternation at its leftmost occurrence which is to the right of the first one, etc. Formally, such a P system discovers the representation of the input string as $\prod_{j=1}^{m-1} (w_j a_j)\, w_m \alpha$, where a_j has no other occurrences inside $w_j a_j$ except as a suffix.

A theoretical note: overlapping occurrences or occurrences with context can be handled by rules with a longer left-hand side. A different order of occurrences of the alternations can be handled by renumbering the alternations. Should the specification of a group require, e.g., second-leftmost occurrence for $a \to u$, this can be handled by inserting a fictive substitution $a \to a$ before $a \to u$, etc. Therefore, this is the most general method.

We construct the following P system, which takes the input in the form

$$\#_l w \#_r = \#_l \prod_{j=1}^{m-1} (w_j a_j)\, w_m \alpha \#_r.$$

$$\begin{aligned} &\Pi''_G = (O, \Sigma, [\ \ ]_1, \emptyset, R_1, 1),\ \text{where} \\ &\ \Sigma = V \cup \{\#_l, \#_r\}, \end{aligned}$$

$$O = \Sigma \cup E,$$
$$E = \{A_{k,j} \mid 1 \leq k \leq s,\ 0 \leq j \leq m\},$$
$V = \{a, \cdots, z\}$ is the alphabet of the Romanian language,

and the rules are given below.

$$\begin{aligned} R_1 = \{ &\#_l \to A_{1,0}|| \cdots ||A_{s,0}\} && (1)\\ &\cup \{A_{k,j-1}\gamma \to \gamma A_{k,j-1} \mid \gamma \in V \setminus \{a_1^{(j)}\},\ 1 \leq k \leq s,\ 1 \leq j \leq m\} && (2)\\ &\cup \{A_{k,j-1}a_1^{(j)}v\gamma \to a_1^{(j)}A_{k,j-1}v\gamma \mid a_1^{(j)}v \in Pref(a_j),\\ &\quad |v| < |a_j| - 1,\ \gamma \in V \setminus \{a_1^{(|v|+2)}\},\ 1 \leq k \leq s,\ 1 \leq j \leq m\} && (3)\\ &\cup \{A_{k,j-1}a_j \to u_j^{(k)}A_{k,j} \mid 1 \leq k \leq s,\ 1 \leq j \leq m\} && (4)\\ &\cup \{\alpha A_{k,m}\#_r \to (f_1^{(k)}, out)|| \cdots ||(f_{p_m}^{(k)}, out) \mid 1 \leq k \leq s\}. && (5) \end{aligned}$$

The rules are presented as a union of 5 sets. The rule in the first set replicates the input for carrying out different inflection subsets. The symbol $A_{k,j}$ is a marker that will move through the string. Its index k corresponds to the inflection subset, while index j tells how many alternations have been carried out so far.

The rules in the second set allow the marker to skip a letter if it does not match the first letter needed for the current alternation. The rules in the third set allow the marker to skip one letter if some prefix of the needed subword is found, followed by a mismatch. The rules in the fourth set carry out an alternation, and the last set of rules perform the replicative substitution of the flectives.

This system halts in at most $|w| + 2$ steps.

5 Determining the inflection group

The rules of the systems described above define, in fact, the way of inflection at algorithmic level:

– deleting the given number of symbols at the end of the word (α),

– obtaining the roots by making substitutions (vowel and consonant alternations),

– attachment of the respective endings to each root.

But this method can be applied only for the case when the number of the inflexion group is known. Otherwise there appears the problem of inflexion model establishing, knowing the graphical representation of the word. Is it possible to solve algorithmically this problem? The answer is negative. The first obstacle is the determination of part of speech: there are several examples of homonyms which mean different parts of speech. (Example: *abate* – masculine noun (*abbat*) and verb (*to divert*). In English this phenomenon is very common, and most nouns are the verbs too.) Let us restrict the formulation of the problem: is it possible to establish the model of inflection (in the conditions indicated above) knowing the part of speech? The answer is negative in this case too. For confirmation we can bring a list of examples, which show us that without invoking phonetic information or the etymological one we cannot determine the model of inflection. Let us illustrate this assertion by analyzing female noun *masă*. Following the meaning of furniture object we will form plural *mese*, using the model with vowel alternation $a \to e$. But if you are following the meaning "compact crowd of people" [22], the plural *mase* will be produced without alternation. The origin of this phenomenon is etymological: in the first case the origin of the word is from Latin *mensa*, and in the second – from the French word *masse* [22]. But the problem can be tackled in another way: we can set certain criteria that allow us as a result of analysis of the word structure to conclude, if it is possible to determine the inflection model or not. If so, we determine precisely which is the respective model.

In [5] the algorithm had been proposed, which, analyzing the dictionary of classification into morphological groups with entries of type (w, σ), where w is a word in natural language, and σ – number (label) of inflection group, constructs two groups of sets $A = \{A_1, A_2, \ldots, A_k\}$ and $P = \{P_1, P_2, \ldots P_s\}$, $\cap_{i=1}^{k} A_i = \emptyset$, $\cap_{i=1}^{s} P_i = \emptyset$. $A_i \cap P_j = \emptyset$.

These sets consisted of subwords α_i of the words $w = w'\alpha_j$, where $1 \leq |\alpha_j| \leq |w|$. In [5] it is shown that for certain categories of words it is possible to construct such sets A_i, that from the fact that $\alpha_j \in A_i$ it results unequivocally that the word w belongs to the single inflection group σ, and these words being named "absolutely regular". With the help of the same algorithm there are constructed also such sets P_i, that from the fact that $\alpha_j \in P_i$ it results that $w = w'\alpha_j$ can belong to several inflection groups $\sigma_1, \ldots, \sigma_m$, and the respective words being named "partially regular".

So, in the case of an arbitrary word w, using the algorithm mentioned above, the inflection group is established at first, and then with the help of membrane system described above, the inflection is carried out obtaining word forms (with respective morphological attributes).

6 Conclusions

The membrane system to describe the inflexional process when the inflexional morphological model is known is investigated in this article.

In the case when the model is not known in advance, it can be determined by using the algorithm from [5]. The membrane systems presented in this paper can be also adapted for other natural languages with high level of inflection, such as Italian, French, Spanish etc., having structured morphological dictionaries, similar to the Romanian one.

Future work: we plan to also consider the problem of representation of the algorithm determining the inflection group by membrane systems.

Acknowledgments

The authors acknowledge the support of the Science and Technology Center in Ukraine, project 4032 "Power and efficiency of natural computing: neural-like P (membrane) systems". The first author also acknowledges the support of the Japan Society for the Promotion of Science, and the Grant-in-Aid, project 20·08364. The fourth author gratefully acknowledges the support of the European Commission, project MolCIP, MIF1-CT-2006-021666.

References

1. E. Boian, S. Cojocaru. The Inflexion Regularities for the Romanian Language. *Computer Science Journal of Moldova*, 4, 1, 1996, 40–58.
2. E. Boian, S. Cojocaru, L. Malahova. Tools for Linguistic Applications (Instruments pour Applications Linguistiques). in: *La terminologie en Roumanie et en Republique de Moldova*, Hors serie, N4, 2000, 42–44 (in French).
3. E. Boian, A. Danilchenco, L. Topal. The Automation of Speech Parts Inflexion Process. *Computer Science Journal of Moldova*, 1(2), 1993, 14–26.
4. S. Cojocaru. Romanian Lexicon: Tools, Implementation, Utilization. in *Language and Technology.* (Lexicon român: instrumentar, implementare, utilizare. In: *Limbaj şi tehnologie*), Academia Română, Bucureşti, 1996, 37–40 (in Romanian).
5. S.Cojocaru. The Ascertainment of the Inflexion Models for Romanian. *Computer Science Journal of Moldova*, 14, 1(40), 2006, 103–112.
6. S. Cojocaru, M. Evstiunin, V. Ufnarovski. Detecting and Correcting Spelling Errors for Romanian Language. *Computer Science Journal of Moldova*, 1(1), 1993, 3–22.
7. C. Coşman. Paradigmatic Morphology of Romanian language. Environment of development – actualization. (Morfologia paradigmatică a limbii române. Mediu de dezvoltare-actualizare. Teză de licenţă), Facultatea de Informatică, Universitatea "A.I.Cuza", Iaşi, 2002. (http://consilr.info.uaic.ro) (in Romanian).
8. D. Cristea, C. Forăscu. Linguistic Resources and Technologies for Romanian Language. *Computer Science Journal of Moldova*, 14, 1(40), 2006, 34–73.
9. R. Hausser. Foundations of Computational Linguistics. Human-Computer Communication in Natural Language. 2nd edition, revised and extended. Springer, 2001.
10. T. Hristea, C. Moroianu. Generation of Flexional Forms of Nouns and Adjective for Romanian Language (Generarea formelor flexionare substantivale şi adjectivale în limba română). in: *Building Awareness in Language Technology.* F.Hristea, M.Popescu (eds.), Editura Universităţii din Bucureşti, 2003, 443–460 (in Romanian).
11. D. Irimia. The Grammar of Romanian Language (Gramatica limbii române). Ed.II-a. Polirom, Bucureşti, 2004 (in Romanian).
12. A. Lombard, C.Gâdei. Morphological Romanian Dictionary (Dictionnaire morphologique de la langue roumaine). Bucureşti, Editura Academiei, 1981 (in French).
13. S. Marcus, Gh. Păun , C. Martn-Vide, Contextual grammars as generative models of natural languages, *Computational Linguistics*, v.24 n.2, June, 1998, 245–274.
14. G. Bel Enguix, M. D. Jimenez Lopez. Linguistic Membrane Systems and Applications. in *Applications of Membrane Computing.* G. Ciobanu, M. J.Prez-Jimnez, Gh.Păun, (Eds.) 2006, 347–388.

15. R.Gramatovici, G. Bel Enguix, Parsing with P automata. in *Applications of Membrane Computing.* G. Ciobanu, M. J.Prez-Jimnez, Gh.Păun, (Eds.) 2006, 389–436.
16. Gh. Păun. Membrane Computing: an Introduction. Springer, 2002.
17. L. Peev, L. Bibolar, E. Jodal. A Formalization Model of Romanian Morphology. in *Language and Technology* (Un model de formalizare a morfologiei limbii române. în: *Limbaj şi Tehnologie.*) Editura Academiei Române, Bucureşti, 1996, 67–72 (in Romanian).
18. L. Peev, F. Şerban. Methods of Romanian Text Linguistic for Terminological Extraction. In Tools and Resources. (Metode de analiză lingvistică a textelor în limba română pentru extragerea terminologică. Instrumente şi resurse.) - in http://dtil.unilat.org/seminar_bucuresti_2008/actes/peev_serban.htm (in Romanian)
19. D. Tufiş. Paradigmatic Morphology Learning. *Computers and Artificial intelligence* 9(3), 1990, 273–290.
20. D. Tufiş, A. M. Barbu, V. Pătraşcu, G. Rotariu, C. Popescu. Corpora and Corpus-Based Morpho-Lexical Processing. In: D.Tufiŝ, P.Andersen (eds.). *Recent Advances in Romanian Language Technology*, Editura Academie Române, Bucureşti, 1997, 115-128.
21. D. Tufiş, L. Diaconu, C. Diaconu, A. M. Barbu. Morphology of Romanian Language, a Reversible and Reusable Resource. In *Language and Technology* (Morfologia limbii române, o resursă lingvistică reversibilă şi reutilizabilă. În: *Limbaj şi Tehnologie*). Editura Academiei Române, Bucureşti, 1996, 59–65 (in Romanian).
22. *The explanatory Romanian Dictionary (Dicţionarul explicativ al limbii române.)* Academia Română, Institutul de Lingvistică "Iorgu Iordan", Editura Univers Enciclopedic, 1998 (in Romanian).
23. *The Grammar of Romanian language (Gramatica limbii române)*, vol.I, Editura Academiei Republicii Populare Române, Bucureşti, 1963 (in Romanian).
24. nl.ijs.si/ME/V3/msd/html/
25. http://www.thefreedictionary.com/paradigm/