Rule-Based Automatic Phonetic Transcription for the Romanian Language

Article · November 2009

DOI: 10.1109/ComputationWorld.2009.59

2 authors:

Stefan-Adrian Toma
Technical Military Academy of Bucharest

Doru Munteanu
Lumina - The University of South-East Europe

Rule-based Automatic Phonetic Transcription for the Romanian Language

Toma Ştefan-Adrian, Munteanu Doru-Petru
Telecommunication Department
Military Technical Academy
Bucharest, Romania
e-mail: toma.stefan.adrian@gmail.com
munteanud@mta.ro

Abstract — **In this paper, we report on an implemented rule-based tool for grapheme to phoneme conversion for the Romanian language and experiments run on two different sets of words. A 4779 words database consisting of the most frequent words in a text corpus collected from 93 books representing genuine Romanian literature, foreign literary works translated into Romanian and scientific texts was transcribed obtaining more than 95% accurate phonetic transcriptions, using a set of 102 letter-to-sound rules. Then, using the same rules, a larger set consisting of 15599 words was transcribed generating 91.46% accurate transcriptions. Although the rules weren't written by phoneticians and can be further improved, it is clear that even for mostly phonetic languages like Romanian, rule-based letter-to sound systems with manually introduced rules are limited by the human inability to spot important patterns in the pronunciation.**

Keywords**- phonetics; letter-to-sound; XML; rules**

I. INTRODUCTION

There are only a few languages that have important linguistic resources - large phonetic dictionaries organized in a ready to use form for speech and natural language processing. Unfortunately, at this time, the Romanian language is still not among them.

In Romanian, words have many inflected forms. Therefore their number is an order of magnitude greater than the root words. The Romanian language online dictionary has now more than 360.000 entries (http://dexonline.ro) so over a few million distinct inflected forms are expected to exist. At this moment, neither basic entries nor the inflected forms have phonetic transcriptions. Hence, a tool for automatic phonetic converter is useful for the research community and commercial applications alike.

There are two solutions proposed to address this problem. The first one, described in [1], is part of a text to speech (TTS) synthesis system and is based on a parallel architecture of neural networks. In this approach, the letter-to-sound system receives sequences of five letters (the target is the central letter) and outputs the allophone needed for further processing in the TTS system. Experimental results report a 3% error rate on a 1000 words testing set, using 4000 words for training. The second method [2] uses a mixed approach: rule-based and data-oriented. First, a raw phonetic transcription is generated using a set of rules. Then, a neural network refines the conversion process. The system was trained on 1000 words and tested on 400 words. The reported context error rate is around 5%. The main drawback for the data-oriented approach [3][4] is that it needs a lot of confident data for training (thousands of correct pronunciation examples).

Our goal is to generate a phonetic dictionary for the Romanian language in a faster manner and as confident as possible. For this, a rule-based converter was built consisting of two parts: a rules database defined using the extended mark-up language (XML) and a parser. The main advantages of this approach are that the context-based phonetic rules are easy to define and maintain, it needs low computational resources (time and memory) and that the system is easy adaptable to any language having a pronunciation based orthographic system. This method is the simplest way to obtain phonetic dictionaries for under-resourced languages because error analysis and rules extension are generally easier than data-driven methods such as neural networks, classification and regression trees (CART), etc. By using the XML formalism, the rules written by the authors of this paper can be easily refined by phoneticians that will discover a friendly way to describe them.

Expert systems have been used for some time for phonetic transcriptions. Examples of such systems for other languages are given in [5][6][7], etc. For morphemic languages the results obtained are contradictory. For example, for English, the system described in [5] produces correct pronunciations for approximately 90% of the words. The same system was tested again in [8], obtaining a 25.7% error rate. A rule-based LTS system for French was tested on a 60000 words dictionary and the obtained error rate was less than 1% [6]. In [7], is reported on a phonetic converter for Galician for which the obtained error rate was 1.5%. Galician, a variety of Portuguese is a language with a pronunciation based orthographic system.

The paper is organized as follows. In Section 2, the phonetic characteristics of the Romanian language are shortly presented. The letter-to-sound system is described in Section 3 followed by the experimental results in Section 4. The last section concludes the paper with final remarks.

II. ROMANIAN LANGUAGE PHONETICS

There are several types of orthographic representations but the most important are the morpheme-base system (e.g.,

978-0-7695-3862-4/09 $26.00 © 2009 IEEE
DOI 10.1109/ComputationWorld.2009.59

English, French) and the pronunciation-based system (Spanish, Italian, Romanian, etc) [9]. Starting with 1880, Romanian became mostly phonetic rather than etymologic, as it was considered before [10]. This was consolidated in the following years by the Romanian Academy.

TABLE I. ROMANIAN PHONEMES

Phonem	Graphic	Phonetic	Phoneme	Graphic	Phonetic
Vowels (7)			g_j	**G**HEM	**g_j**em
i	V**I**N	v**i**n	g	**G**ÂT	**g**1t
e	F**E**L	f**e**l	ts	**Ț**ARĂ	**ts**ar@
a	C**A**P	k**a**p	tS	**C**EAS	**tS**as
@	M**Ă**R	m**@**r	dz	**G**EAM	**dz**am
o	L**O**C	l**o**k	f	**F**ATĂ	**f**at@
u	S**U**R	s**u**r	v	**V**ALE	**v**ale
1	F**Â**N	f**1**n	s	**S**AC	**s**ak
Semivowels (4)			z	**Z**I	**z**i
j	DO**I**	do**j**	S	ŞA	**S**a
e_X	N**E**A	n**e_X**a	Z	**J**OI	**Z**oj
w	SA**U**	sa**w**	h	**H**AM	**h**am
o_X	**O**ARE	**o_X**are	m	**M**ARE	**m**are
Consonants (22)			n	**N**AS	**n**as
p	**P**ĂR	**p**@r	l	**L**APTE	**l**apte
b	**B**AR	**b**ar	r	**R**ÂS	**r**1s
t	**T**UN	**t**un	k_j	**CH**EL	**k_j**el
d	**D**AR	**d**ar	**Non-syllabic vowel (1)**		
k	**C**AL	**k**al	i_0	LUP**I**	lup**i_0**

Our work is focused on the pronunciation-based languages. Therefore, the system presented in this paper can be easily adapted to new languages if appropriate rules were defined.

In Romanian, any word can be graphically represented using 31 letters: A, Ă, Â, B, C, D, E, F, G, H, I, Î, J, K, L, M, N, O, P, Q, R, S, Ş, T, Ţ, U, V, W, X, Y, and Z. These letters are mapped to 34 phonemes: 22 consonants, 7 vowels, 4 semivowels and 1 non-syllabic vowel [11]. Pronunciation examples are given in table 1 in SAMPA [12] based format. Besides the Romanian SAMPA consonants inventory, the palatalized allophones of the consonants [k] and [g] in the *CHE*, *GHE*, *CHI* and *GHI* letter contexts are represented with [k_j] and [g_j] as distinct phonemes (according to [13] [14]).

In Romanian, phonetic transcription seems to be simple, as is the case of the word "speranță" (*eng. "hope"*):

Graphic representation	Phonetic representation
S P E R A N Ț Ă	s p e r a n ts @

However, we have identified only *18 letters* that are mapped to a single phoneme (especially consonants) with no exception (e.g., *D*, *F*, *J*, *L*, etc.). Hence, one single rule is needed those situations. All other letters require multiple rules in order to generate correct phonetic transcriptions.

The rules developed for Romanian by the authors of this paper are based on both phonetic principles enounced by Romanian phoneticians [10][11][13][14] and observations on a dictionary previously created in [15].

TABLE II. PHONETIC VALUES OF THE LETTER *E*

Case index	Pho-neme	Word example	
		Graphic	Phonetic
1	[e]	REAL	r e a l
2	[e_X]	MEA	m e_X a
3	[e e_X]	ACEEA	a tS e e_X a
4	[j]	EA	j a
5	[j e]	ESTE	j e s t e
6	[=]	CEAS	tS a s

The most difficult problem in defining phonetic rules for the Romanian language is to identify the correct transcription of the following letters: *E*, *I*, *O* and *U*. They are mapped to vowels, semivowels, phonetic zero (denoted by "=") or even a group of vowels. In our experiments, the transcription of the letter *I* generated most of the errors.

It is very difficult to identify the situations when these phonemes generate a vowel or a semivowel. For instance, when two consecutive vowel letters are uttered in the same syllable they will form a diphthong (vowel-semivowel or semivowel - vowel). When they are uttered in two different syllables they form a hiatus (vowel-vowel).

For example, *E* could be in one of the six cases described in table 2. One can observe that *E* followed by *A* generates the phoneme [e] in case 1 (hiatus) and the phoneme [e_X] in case 2 (diphthong). Phoneticians give no rule for such hiatus/diphthong case identification, except syllable decomposition [13], which is usually unknown. Still, we identified certain cases for which rules may be defined. For example, the letter *E* preceded by a consonant letter and followed by *A* and the end of the word, it is always transcribed as phoneme [e_X], as shown in table 2, second row.

III. SYSTEM DESCRIPTION

The letter to sound system architecture is given in Figure 1. The input word is passed to the Phonetic converter block or the Word transcription lookup block depending on the state of the "D" switch. If "D" is 2 the system searches for the input word in the Phonetic dictionary. If the word is found in the dictionary the system outputs the phonetic transcription. Else, the word is passed to the Phonetic converter where it is converted letter by letter to its phonetic representation.

A. Context based XML rules definition

The phonetic rules are essential for the performance of the system. Based on phonetic works [10][11][13][14] but also on observations made on a small phonetic vocabulary (3176 words) we have identified two classes of letters:

- the single rule letters class
- the multi-rule letters class

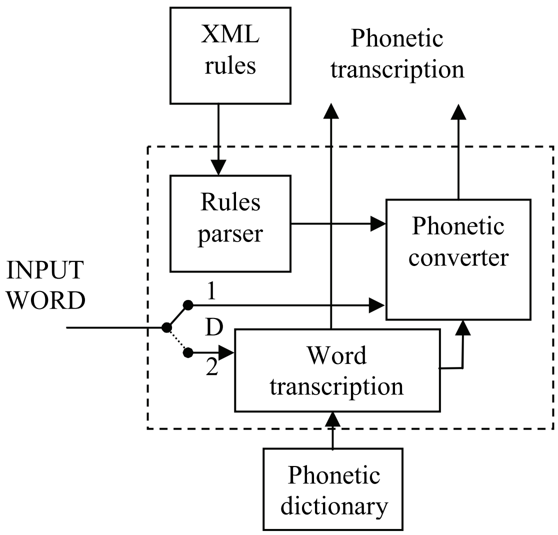

Figure 1. LTS system architecture.

The first class refers to 18 letters that are always mapped to the same phonemes (or group of phonemes) and, normally, do not generate any transcription error. The second class refers to the 13 remaining letters and generates most of the transcription errors.

For the rule definition we have defined a simple and easy way to write and read the rules, using the XML syntax. For a given vocabulary or input text, the rule file is parsed and the appropriate phonetic transcription is generated.

The rule definition file is presented below:

```
<rules>
    <define_groups>
   <cons>B,C,D,F,G,H,J,K,L,M,N,P,R,S,S',T,T',V,X
,Z</cons>
       ...
   <gr_AOU> A,O,U </gr_AOU>
    </define_groups>

   <phonemes>
       <p>a</p>
         ....
       <p>v</p>
       </phonemes>

   <letters>
       <l>A</l>
        .....
       <l>Z</l>
       </letters>

       <for_A>
      <r>
        <t>a</t>
      </r>
       </for_A>

    <for_E>
      .....
      <r>
        <left>cons</left>
        <right>cons|=</right>
        <t>e</t>
      </r>
      ...
      <r>
        <right>A</right>
        <right>=</right>
        <t>e_X</t>
      </r>
      .....
   </for_E>
      .......
</rules>
```

The rule file contains:

- the root tag (`<rules>`);
- definitions of the letter groups (the `<define_groups>` tag;
- the phoneme inventory (`<phonemes>` and `<p>` tags);
- the letter inventory (`<letters>` and `<l>` tags);
- identification tags for the set of rules of each letter (`<for_A>`, `<for_B>`, …, `<for_Z>`);
- rule definition tags (`<r>`, `<right>`, `<left>` and `<t>`).

The XML tags form a hierarchy presented in figure 2.

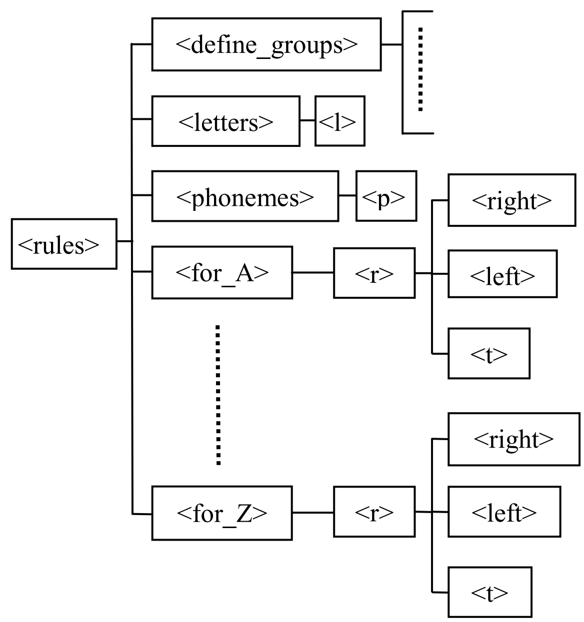

Figure 2. The XML tag hierarchy.

The groups of letters can be any subset of the letter inventory. For example, the group defined by <gr_AOU> tag refers to any of the letters *A*, *O* or *U*. Each time this tag will be encountered within a specific rule, every group element will be checked. In this way a user can adapt the system to any type of letter groups in any language.

The tags <for_A>, <for_E>, etc. mark the beginning of a set of rules for one letter, the tag <r> refers to a specific rule and the tags <left> and <right> refer to the left and right context of the letter. In the above example, the letter *E* has two rules listed.

The first rule means that an *E* after any letter from the consonants group (<cons> tag) and followed by any letter from the same group or by the end of the word denoted with <=> (<right>cons|=</right>) is mapped to the phoneme [e]. The second rule means that an *E* followed by an *A* and then by the end of the word is mapped to the phoneme [e_X].

There are three important characteristics of the rule definition:

a. the letter context is indicated by the <left> and <right> tags from left to right in a top-down representation;
b. the verification of a specific context is done either by checking a single letter or a group of letters; within a group, the rule is selected if **any** letter from that group meets the condition;
c. for a specific letter, rules are verified in a top-down fashion; if one rule is accomplished, then the others are skipped.

For experiments, a total number of 102 rules was derived.

IV. Experimental Results

In order to generate good rules, a phonetic converter has to collect the minimum rules that cover broad linguistic phenomena while avoiding side-effect. In [15], we have manually build a phonetic dictionary of 3176 words covering 23777 letter occurrences. (The term "word" refers to inflected forms of words, not lemmas). Using this dictionary and [10][11][13][14] we developed a basic set of rules. The rules were generated according to a frequency criterion: each rule must be applied at least 5 times within the dictionary. This threshold was empirically optimized in order to obtain a minimum ratio between the number of rules and their generality.

TABLE III. Confusion Matrix for the Transcription of the Letter *E*

	[e]	[e_X]	[=]	[j]	[j e]
[e]	-	13	-	-	-
[e_X]	3	-	-	-	-
[=]	-	-	-	-	-
[j]	-	-	-	-	-
[j e]	-	-	-	-	-
[e e_X]	-	8	-	-	-

TABLE IV. Confusion Matrix for the Transcription of the Letter *I*

	[i]	[j]	[i j]	[=]	[i_0]
[i]	-	19	2	25	31
[j]	19	-	69	-	-
[i j]	-	-	-	-	-
[=]	1	1	-	-	-
[i_0]	9	-	-	-	-

Then, the rules were refined in an iterative fashion. First, the dictionary words were transcribed by our system and compared to the original transcriptions. New observations were made on the wrong transcriptions, so that new rules were added to the first set. The process continued until no significant improvement was observed. The most important parameter that was monitored during this process was the overall word error rate (WER) that was decreased from 21% to 4.94% in 11 iterations.

We tested the system on a different dictionary of 4779 words. This database was developed in [16] and consists of the most 5000 frequent words from a text corpus made of 93 books representing genuine Romanian literature, foreign literary works translated into Romanian and scientific texts. Foreign words were discarded. To evaluate the performance of the system various statistics were generated. In tables 3 and 4 are presented the confusion matrices for the letters that generated most of the errors. The confusion matrices provide useful information regarding the types of transcription errors. For example, for the letter E (table 3), instead of [e] the system wrongly transcribed [e_X] thirteen times. We can see from table 4 that most errors are generated by the erroneous transcription of the letter I - 176 errors. In 69 cases the system transcribed [i j] instead of [j]. This information can be used to further improve the transcription accuracy.

Useful information is also provided by the letter error rate (LER) computed for a specific letter as the ratio between its incorrect transcriptions and the total number of its occurrences (table 5).

TABLE V. Letter Error Rate

Letter	LER [%]	Letter	LER [%]	Letter	LER [%]
A	0.00	*Î, Â*	0.00	*S*	0.00
Ă	0.00	*J*	0.00	*Ş*	0.00
B	0.00	*K*	0.00	*T*	0.00
C	0.00	*L*	0.00	*Ţ*	0.00
D	0.00	*M*	0.00	*U*	0.73
E	0.66	*N*	0.00	*V*	0.00
F	0.00	*O*	0.00	*X*	25.42
G	0.00	*P*	0.00	*Y*	0.00
H	0.00	*Q*	0.00	*Z*	0.00
I	5.34	*R*	0.00	*W*	0.00

In Romanian, the letters *Î* and *Â* represent the same sound, therefore having the same phonetic transcription.

For the entire test vocabulary, the overall letter error rate is 0.72 %, while the overall word error rate is 4.79 %. Note that the letter *X* has the highest LER (25, 42%). This high value is caused by the low number of occurrences of this letter (59 times). Compared to the letter *X*, for the other letters that generated error, the number of occurrences is in the number of hundreds or thousands. The LERs for the test database are consistent with the results for the 3176 words "training" database.

In future work, we intend to use the phonetic converter in a TTS system for embedded applications. Therefore we explored the possibility to run the LTS system with a lower number of rules so as to reduce processing time. The error rate and the relative computation time (normalized to its maximum value) are presented in Figure 3. We tested the system with 102, 92, 82, 72 and 62 rules. The selection criterion was based on the rules usage rate: 10 rules with the lowest usage rate were discarded at every step.

The results show that the error rate has a bigger increase rate than the decrease rate of the relative time. For example, for 82 rules we obtain an error rate of 10.5% - more that 100% increase, and a relative time of 0.88 – a 12% decrease.

A larger phonetic dictionary was used to further test the rules. This dictionary (15599 words) was developed under the SpeechDat specifications as a lexicon in a speech data-base using the LTS system described here and corrected manually. The results show an increase in the WER – 9.46%.

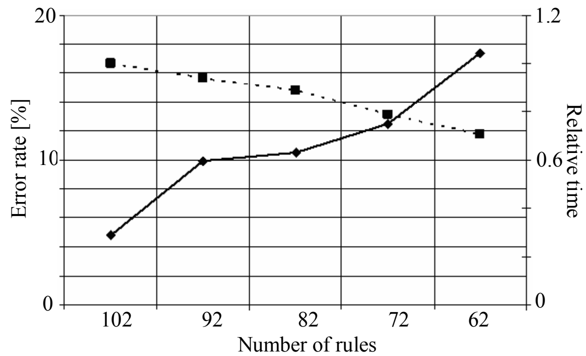

Figure 3. Error rate (continuous) and normalized time (dotted) versus the number of rules.

Although the rules were not written by phoneticians and can be further improved, it is clear that even for mostly phonetic languages like Romanian, rule-based letter-to sound systems with manually introduced rules are limited by the human inability to spot important patterns in the pronunciation.

The results presented above were obtained without the use of the *Phonetic dictionary* of the system (the "D" switch in position 1 in figure 1).

V. CONCLUSION

The work described in this paper demonstrates the possibility to translate letter to sound in Romanian language using a rule-based system. A set of 102 letter-to-sound rules was developed to translate a Romanian text into its corresponding phonetic transcription (using SAMPA representation) with good accuracy. Experiments were run on two sets of words. A 4779 words database was transcribed obtaining more than 95% (4.79% word error rate and 0.72% letter error rate) accurate phonetic transcriptions. Then, a larger set consisting of 15599 words was transcribed generating 91.46 % accurate transcriptions. In both experiments, most of the incorrect transcribed words generated single letter errors.

A comparison with other letter to sound systems for Romanian can be made only by comparing the results obtained on the smaller set of words since the letter-to-sound systems described in [1] and [2] were tested on 1000 words and 400 words. Although the error rate is slightly higher for the system described in this paper, the rule-based approach is easier to maintain and more flexible in developing large phonetic dictionaries for under-resourced languages.

We consider that with the help of expert phoneticians, the phonetic rules can be improved in order to obtain more accurate phonetic transcriptions. Because the rules were written within the XML framework they are portable, intuitive, easy to modify and extend. The version of the rules used in the experiments presented in this paper can be obtained by e-mail, from the authors.

We also analyzed the effect of discarding some of the rules on the processing time. The results show that if a small number of rules are ignored the system will still provide a good performance, while moderately lowering the processing time.

In future work, we will try to adapt the system to other under-resourced pronunciation based languages and we will use the phonetic converter in an embedded TTS system for the Romanian language. The 102 rules LTS converter can provide lower computational cost and power consumption comparing to other methods based on either large phonetic dictionaries or data-driven techniques.

ACKNOWLEDGMENT

This research was supported, in part, by the "Politehnica" University of Bucharest under research contract no. I100902, financed by Telisma.

REFERENCES

[1] D. Burileanu, M. Sima and A. Neagu, "A Phonetic Converter for Speech Synthesis in Romanian", in Proc. of the XIV the International Congress on Phonetic Sciences, San Francisco, USA, Vol. 1, pp. 503-506, Aug. 1999.

[2] D. Jitca, H.N. Teodorescu, V. Apopei and F. Grigoras, "An ANN-based method to improve the phonetic transcription and prosody modules of a TTS system for the Romanian language", in Proc. of Speech Technology and Human-Computer Dialogue (SpED), Bucharest, 2003, pp. 43-50.

[3] M. Bisani and H. Ney, "Investigations on joint multigram models for grapheme-to-phoneme conversion," in Proc. of the 7th International Conference on Spoken Language Processing (ICSLP), Denver, CO, Sep. 2002, vol. 1, pp. 105 -- 108.

[4] G. Bouma, "A Finite-State and Data-Oriented Method for Grapheme to Phoneme Conversion," in 1st Meeting of the North American Chapter of the Association for Computational Linguistics, Seattle, USA, 2000

[5] H. Elovitz, R. Johnson, A. McHugh and J. Shore, "Letter-to-sound rules for automatic translation of english text to phonetics", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-24, No. 6, 1976

[6] T. Dutoit, High quality text to speech synthesis of the french language, Ph.D. dissertation, Faculte Polytechnique de Mons, 1993

[7] D. Braga, L. Coelho, "Letter-to-sound conversion for galician TTS systems", Jornadas en Tecnologia del Habla, Zaragoza, pp. 171-176, 2006

[8] R. I. Damper, Y. Marchand, M. J. Adamson, and K. Gustafson, "A comparison of letter-to-sound conversion techniques for English text-to-speech synthesis", Proceedings of the Institute of Acoustics, 20 (6). pp. 245-254, 1998

[9] M. Tatham and K. Morton, Developments in Speech Synthesys, John Wiley & Sons, 2005

[10] I. Calotă, Small Encyclopaedia of Correct Romanian Language, Niculescu Press, Bucureşti, 2001

[11] Academia Româna (Romanian Academy), Dictionarul ortografic, ortoepic si morfologic al limbii române, Editia a doua (Romanian language orthographic, orthoepic and morfologic dictionary, the second edition), Univers Enciclopedic, Bucuresti, 2005

[12] http://www.phon.ucl.ac.uk/home/sampa/rom-uni.htm [22.08.2009]

[13] G. Beldescu, Actual Orthography of the Romanian Language, Scientific and Encyclopaedic Press, 1984.

[14] F. Şuteu and E. Şoşa, Orthographic Dictionary of the Romanian Language, ATOS & VESTALA Press, 1993

[15] D. Munteanu, Romanian Language Speech Recognition Methods, PhD thesis, MTA, 2006

[16] A. Vlad, A. Mitrea and M. Mitrea, Limba română scrisă ca sursă de informație, Ed. Paideia, 2003