LINGUISTIC PREMISES FOR DESIGNING A TEXT-TO-SPEECH SYNTHESIZER

LUCIANA PEEV, FELICIA ŞERBAN

Abstract. The approach of the project made possible the study and elaboration of the formalized model of the Romanian phonology. The study was performed at the theoretic level and a complete set of phonemes, characteristic for the Romanian language, has been established. The *Text-To-Speech* (TTS) synthesis system is based on a thorough linguistic analysis of the text, its phonetic transcription and the transformation of the phonemes into diphonemes through acoustic parameters. Phonetical transcription is necessary due to the fact that the form of the Romanian written texts differs considerably from the spoken form, which asks for a complex processing to be performed before the effective transformation of the phonemes (or diphonemes, respectively) into graphemes. The scientific results obtained are more than promising. Thus, an extremely important chain closes, in bringing classical linguistics close to the contemporary information technology: linguistics receives new and modern tools for study, and the new results obtained will contribute to the polishing of the informatics systems.

1. PRELIMINARIES

A Text-to-Speech synthesizer (TTS) is a computer based system, able to read a text. There are systems, known as Voice Response Systems, which can concatenate isolated words or parts of phrases. These systems can only be used with a reduced vocabulary (a few hundred words) and when the phrases to be uttered observe a very strict structure, like in the case of the vocal announcements made in railway stations, regarding the arrival of trains. When it comes to the synthesis (TTS) meant to vocally reproduce a certain text, the possibility exists to store all the words in a language, but the outcome would not justify the considerable effort. The ideal TTS synthesis system is the one that can dynamically model, based on a rigorous mathematical description, the way the voice sounds are produced. But, these systems are in their beginning and the results are still to be seen. Another method is the one to define a system able to create the pronunciation based on the phonetical transcription of texts. This last method has been chosen in order to design the TTS system within the project **Text Based Synthesis of Speech in the Romanian Language, Based on an Expert Linguistical System for Multisensory Interfaces (INFOSOC-2004).** We considered that this method would provide the better chances in the modelling of a pronunciation, as close to the natural one as possible, in order to vocally reproduce a written text, due to the

RRL, **LI**, *3–4*, p. 441–456, Bucureşti, 2006

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

fact that the method takes into consideration the changes the sounds undergo in different phonological contexts. Cf. Dutoit (1997), Gaudinat (1998), Ribeiro (2003), Sproat (1997).

At first sight, this task doesn't seem too difficult to accomplish, but, despite the present state of our knowledge and technology, as well as the progress made in the field of artificial intelligence, we have to express a few reservations. Man learns in his childhood to pronounce in his mother tongue and he can, at any time, speak easily phrases not known before. He has learned the rules unconsciously when he learned to speak, then he has learned them in a simpler form during grammar school and improved them by the year. The moment when we can say that artificial intelligence can equal human intelligence is still far away. The TTS system approached in the above mentioned project looks, in broad lines, like this:

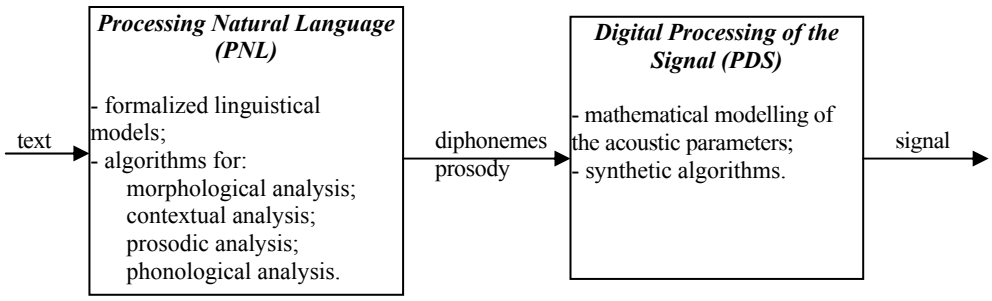

Fig. 1

The synthesizer is based on two important modules, namely: the Natural Language Processing Module (PNL) and the one for the Digital Processing of the Signal (PDS), as can be seen in Fig. 1. Next, we shall present the principles on which the formalization of the Romanian language phonology is based, in order to realize the phonological transcription of a written text, a transcription capable, in the end, to reproduce the sound units (diphonemes) in a word, along with prosodic information.

2. PRINCIPLES OF PHONOLOGY

When we approached the formalization of the Romanian phonology we took into consideration the extant studies on the Romanian phonology and we tried, as much as we could, to respect a widely accepted linguistical theory. Cf. Avram (1956, 1958), Vasiliu (1965), Sfîrlea (1970), Stan (1996), also Dascălu-Jinga *et al.* (1988). We also found a real help in the thesaurus of the linguistical knowledge gathered in the work elaborated by the **RoLingua** team, which includes the morphological and syllabication patterns of the Romanian language, the stress as well as a rich lexicon of over 80,000 words covering more than two million flexional forms. Cf. Peev *et al.* (1996, 1997), Şerban *et al.* (1996, 1998, 2000). The

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

lexicon of the knowledge thesaurus is based on the *Dicţionarul explicativ al limbii române* (DEX 1975/1996), *Dicţionarul ortografic, ortoepic şi morfologic al limbii române* (DOOM 1982/2005), as well as on words taken from *Dicţionarul de neologisme* (1978).

Based on the analysis and the phonological interpretation performed for the flexional forms and syllabication, as well as on the analysis of the influence the stress has over the flexion, we could formalize the phonological rules. The TTS system in the above mentioned project intends to reproduce an utterance as close to the natural one as possible in the vocal reproduction of the written text, which implies respecting the changes the sounds undergo in different phonological contexts. Taking into consideration the differences, not in the least negligible, between written and spoken Romanian, (spelling is not totally "phonetical", as it is commonly asserted), phonological transcription for the Romanian language is justified. In order to verify the way the phonological formalization was accomplished we used the international phonetic system IPA. Cf. also Tătaru (1999).

Generally, phonetical transcription is based on a set of rules and on a phonetically transcribed vocabulary. The Romanian language has a phonological system based on rules that can be formalized. We considered that the set of rules has the greatest importance in the phonetical transcription and that using the totally transcribed vocabulary is not necessary, except from keeping some phonological attributes as grapheme equivalences. There are not many exceptions and they represent especially the foreign words that entered the language and were not assimilated. Of course, there are also other situations, such as dealing with the graphemes having double phonological representation (e.g. the grapheme *x*). Resolving such situations is possible by using the *grapheme-to-grapheme equivalence method,* method developed in this project.

3. PHONETICAL CONSIDERATIONS

In order to realize the phonological formalization of the Romanian language all the extant phonological situations in the actual Romanian language have been highlighted and two sets were established:

– **The complete set** comprising over 200 phonological situations identified through distinct units (phonemes and allophones) that have been formalized. They are not yet going to be implemented in the TTS system within this project, as they require extended human and time resources.

– The reduced set comprising 51 phonological situations taken from the complete set, identified through distinct units (phonemes and allophones). We considered this set illustrative in order to demonstrate the performances of the TTS system implemented in this project.

In the present paper, we shall use the term ***phoneme*** for both phonemes and allophones. The right term would be ***pronunciation unit*** or ***phonic unit.***

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

The phonetic rules formalized cannot be reduced to the simple *grapheme-phoneme* equivalence, because in the Romanian language there are phonemes for which there is no grapheme in spelling, and the vowels may have the value of semivowels within diphthongs and triphthongs. We also considered that the complete or incomplete occlusion of the consonants *n* and *m* is noticeable.

The exceptions represent words in which the phonemes cannot be identified by grapheme equivalence. Some of them belong to the inherited vocabulary of the Romanian language: *eu, el, ea, ele, ei, este, eşti* (as compared to the rest of the words starting with *e,* these develop an initial unsyllabic element [ʲ]); others are foreign words which entered the Romanian language, but keep the pronunciation from the original language: *bleu, bleumarin, business, joker, joule, röntgen* etc.

3.1. Rules corresponding to the complete set of phonetical units

In the complete set a thorough classification of the sound nuances in the Romanian language has been performed, also taking into consideration the interaction between the phonemes. The designed TTS system is based on the concatenation of diphonemes. Diphonemes are two sound groups formed of two phonemes, half a sound being cut out of each phoneme. When the words are reconstructed by concatenating the diphonemes, it is evident that the result will be better if the sounds to recompose belong to the same sound group. We are not going to reproduce entirely and systematically the transcription of the words in the international phonetic system, as we don't consider it necessary.

3.2. Phonetical rules specific to consonants

Complete formalization of the complete phonological situations of the consonants show that the consonants are influenced especially by vowels, but there also exist situations when consonants can influence each other. This situation is very obvious in the extended set, not shown here. In the present paragraph we present a few elements of the reduced set of phonological situations, referring to consonants (phonemes and allophones).

• *Neutral* consonants:

[b, k, d, f, g, h, j, l, m, n, p, r, s, ʃ, t, ts, v, z]

• *Palatalized* consonants do not have a special graphical correspondence in the Romanian language and are represented by several graphemes (*c*, *g, ch, gh* followed by *e* or *i*)

[k', g', tʃ, dʒ]: ***ce**rc* [tʃerc], ***ge**r* [dʒer̥]

• *Nasal* consonants or consonants with *incomplete occlusion* [m̥, n̥]: *î**n**vă̱t̥* [ɨm̥vˈəts]

• *Nasal velar* consonant [ŋ]: *ba̱ncă* [bˈaŋkə]

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

Some situations in the extended set are dealt with indirectly through the diphonemes set, performed on the bases of the presented phonemes. Thus, the aspiration and deafening of the consonants is revealed in the recording of the diphonemes that occur at the beginning of the words.

3.3. Phonetical rules corresponding to vowels

Vowels contract distinct relationships as compared to consonants. Also, vowels are the nucleus of the syllable, and the stress depends on them. Vowels can influence each other reciprocally: they can become more open or closed, palatalized, labialized, can become longer or shorter, semivowels or unsyllabic elements accompanying the vowels. Even though, as compared to other languages, the series that characterize a vowel do not determine the realization of a sound different from the vowel, the sound being the same, the acoustic parameters being different. This fact is clearly exposed in the extended set, which is not being presented here. In the reduced set the phonemes presenting palatalization, labialization and nasalization were left aside and only the allophones corresponding to semivowels and the deafened *i* were kept. In order to represent the unsyllabic elements for which there is no graphic correspondent we have used some special signs.

3.3.1. *Vowels*

• *Neutral* vowels: [a e i o u ə ɨ]
• *Deafened* or *devocalized* vowels: [ɨ̥]

3.3.1. *Semivowels*

These elements can never occur under *stress* and, compared with the vowels, they are *shorter*. Semivowels can be part of a syllable only if they accompany a vowel. Alone they cannot form a syllable. A vowel cannot be accompanied by more than two semivowels and in this case they form diphthongs and triphthongs. The vowels in the central series *a, â* (*î*) can never become semivowels.

• Semivowels: [e̯ , j , o̯ , w]
• Vowels [e, i, o u] can be semivowels [e̯ , j , o̯ , w̯] in different phonological contexts: *aiuri̯t* [ajur'it].

3.3.3. *Unsyllabic elements*

• In *initial position* (after pause), the vocalic phonemes can be accompanied by unsyllabic elements, with a timbre resembling that of the respective vowel, such as: the vowels in the anterior series [e, i] can be accompanied by an unsyllabic element [e̯] or [ʲ], those in the central series, [a, ə, ɨ] by – [ə̯], and those in the posterior series

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

[o, u] by – [ᵒ̯] or [ʷ]. In their turn, vowels preceded by such an unsyllabic element are more closed as compared to others that are not in the same phonological situation. In the literary language the unsyllabic elements developed before pause (in initial position) they rarely occur: there are a few exceptions, namely those related with the personal pronoun (*eu, el, ea, ei, ele*) and the flexional forms of the verb *a fi* (*este, eşti*). Due to this situation, in this stage we shall not consider these elements, safe from the already mentioned exceptions.

• When a *hiatus* is formed (in the case of the syllabic pause) unsyllabic elements may also occur, depending on the nature of the vowels between which they develop and which help in uttering of the hiatus. The vowels [e, i, o, u] behave as they do in initial position and are accompanied by unsyllabic elements with a resembling timbre, as above specified. But the vowels [a, ə, ɨ] are accompanied by an unsyllabic element having a timbre resembling the precedent vowel: the vowels [e] or [i] will be followed by the element [ᵉ̯] or [ʲ], and [o] or [u] will be followed by [ᵒ̯] or [ʷ]. When the anterior vowels [e, i] and the posterior vowels [o, u] are concerned, the accompanying unsyllabic elements can be optional, but in the case of the vowel [o] followed by the central vowels [a, ə, ɨ] this element is compulsory.

The unsyllabic elements developed in the hiatus [ᵉ̯, ʲ, ᵒ̯, ʷ] resemble the semivowels, but they are shorter and they don't have a graphic correspondent: *adăuga* [adəʷugˈa]. The cases when such elements may occur will be presented with the presentation of the hiatus.

3.3.4. *Special TTS phonic units*

The designed TTS system is based on the concatenation of diphonemes, which consist of half phonemes. We considered that half an unsyllabic element is very difficult to perceive and thus, difficult to cut up. For this reason, we preferred to introduce the *phoneme + unsyllabic element* as a *distinct unit,* in those cases where there are preceded by unsyllabic elements vowels. Such groups are the following:

ᵉ̯a	*creat* [kreᵉ̯ˈat]
ʲa	*patriarh* [patriʲˈarh]
ʲe	*familiei* [famˈiliʲej]
ʲi	*sfîi* [sfiʲˈi]
ᵒ̯a	*coagulat* [koᵒ̯agulˈat]
ʷa	*accentua* [aktʃentuʷˈa]
ʷo	*virtuos* [virtuʷˈos]
ʷu	*perpetuu* [perpˈetuʷu]
ʷə	*ingenuă* [indʒˈenuʷə]
ʷɨ	*luând* [luʷˈind]

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

In the present paper we didn't note the changes in timbre the vowels undergo as well as other pronunciation particularities, which do not refer to the phenomena dealt with here, in order to avoid possible mistakes while the subsequent processing of the text and because the list is too extended to be reproduced here. In the complete set all these situations are formalized. Phonetic experiments can decide if additional allophones, others than the ones in the reduced set, are necessary.

3.3.5. *Diphthongs*

Diphthongs – combinations of two phonemic elements, one of them a vowel and the other a semivowel, both of them being always part of the same syllable – may occur in any position in the word: initial, middle or final.

Ascending diphthongs (the first element is a semivowel)**:**

eo	[e̯o]	*pleosc*	[pleˈo̯sk]
ea	[e̯a]	*se<u>a</u>ră*	[seˈo̯arə]
ia	[ja]	*nui<u>a</u>*	[nujˈa]
ie	[je]	*b<u>a</u>ie*	[bˈaje]
io	[jo]	*fui<u>o</u>r*	[fujˈor]
oa	[o̯a]	*bo<u>a</u>lă*	[boˈo̯alə]
ua	[wa]	*acuar<u>e</u>lă*	[akwarˈelə]
uă	[wə]	*plou<u>ă</u>*	[plˈowə]
uâ	[wɨ]	*plou<u>â</u>nd*	[plowˈɨnd]
ue	[we]	*înşeu<u>e</u>z*	[ɨnʃewˈez]
uu	[uw]	*as<u>i</u>duu*	[asˈiduʷu]

Descending diphthongs (the first element is a vowel):

au	[aw]	*dau*	[daw]
ai	[aj]	*cai*	[kaj]
ău	[əw]	*dul<u>ă</u>u*	[dulˈəw]
ăi	[əj]	*dul<u>ă</u>i*	[dulˈəj]
âu	[ɨw]	*brâu*	[brɨw]
âi	[ɨj]	*b<u>â</u>jbâi*	[bˈɨʒbɨj]
ei	[ej]	*lei*	[lej]
eu	[ew]	*greu*	[grew]
iu	[iw]	*scriu*	[skriw]
iu	[ju]	*am<u>o</u>niu*	[amˈonju]
ii	[ij]	*c<u>o</u>pii*	[kˈopij]
ii	[ji]	*l<u>e</u>ii*	[lˈeji]
oi	[oj]	*d<u>o</u>ină*	[dˈojnə]
ou	[ow]	*cad<u>o</u>u*	[kadˈow]
ui	[uj]	*cui*	[kuj]

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

3.3.6. *Triphthongs*

Thriphtongs (groups of sounds consisting of three phonemic elements, one of them being a vowel and two semiwovels) always start with a *semivowel* and can also be described as consisting of a *semivowel + diphthong*.

Ascending triphthongs (the vowel is preceded by semivowels):

eoa	[e̯o̯a]	*leoa̲rcă*	[le̯o̯ˈarkə]
ioa	[jo̯a]	*creioa̲ne*	[kreje̯o̯ˈane]

Ascending-descending triphthongs (the vowel is between the semivowels):

eai	[e̯aj]	*vreai*	[vre̯ˈaj]
eau	[e̯aw]	*avea̲u*	[ave̯ˈaw]
ieu	[jew]	*maie̲u*	[majˈew]
iai	[jaj]	*suia̲i*	[sujˈaj]
iau	[jaw]	*suia̲u*	[sujˈaw]
iei	[jej]	*miei*	[mjˈej]
oai	[o̯aj]	*leoa̲ică*	[leo̯ˈajkə]
uai	[waj]	*înşeua̲i*	[ɨnʃewˈaj] perf. s. I
uau	[waw]	*înşeua̲u*	[ɨnʃewˈaw] imperf. VI
uăi	[wəj]	*ro̲uăi*	[rˈowəj]
uei	[wəj]	*ca̲uei* (G-D)	[kˈawej]

The diphthongs and triphthongs are presented in the reduced set without mentioning the closing, palatalization or labialization of the vowel; and the hiatus is expressed only by means of the neutral vowels or of the special TTS phonic units.

3.3.7. *Hiatus*

Two adjoining vowels are generally pronounced with a certain pause between them, but there also exist situations when an *unsyllabic element* is produced which makes the pronunciation easier. Based on the extant research hiatuses were classified according to the possible or impossible occurrence of an unsyllabic element. We must also observe the recommendations to correctly pronounce the neologisms in the literary language, according to the norms, even though the hiatus is, generally, very difficult to utter for the Romanian speaker, without introducing a transition element between the vowels.

The unsyllabic element accompanying the vowels [e, i, o, u] has a timbre resembling the one of the preceded vowel: [^{j}e,^{j}i, ^{w}o, ^{w}o].

The vowels [a, ə, ɨ] can also be preceded by an unsyllabic element having a timbre close to that of the precedent vowel: [e^{e}a, $ɨ^{j}$a, o^{o}a].

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

In the reduced set, the hiatus is expressed only by means of neutral vowels or the special TTS phonic units. We shall give examples of hiatuses in which an unsyllabic element is created**:**

• *e-e* if the second *e* is part of an alternation *e+e/ea* the vowel [e] will be preceded by an unsyllabic element [ʲ̃], *ah**ee**ni* [aheʲˈeni] / *ah**eea**n* [aheę̂ˈan].

– if *ee* occurs at the end of a feminine noun, and this has a determiner, a change similar, in form, to the alternation will be produced *e/ea: id**ee*** [idˈeʲe]/ *id**eea*** [idˈeę̂a];

– if *ee* occurs in the middle of the word, we must distinguish between the words *formed on Romanian soil* and *neological loans.* In the case of the Romanian derivatives, the latter of the two *e* sound becomes semivowel, while in the case of the neological loans it doesn't: *coreean* [koreˠˈan] (Romanian derivative), *gheenă* [geˈenə̯] (neologism). The difference will be done using the equivalence method;

– if the two *e* resulted from the addition to a word of an element of composition, between the two [ee] no unsyllabic element will develop: *neechitabil* [neek'itˈabil];

– the situation when the two *e* come from a foreign word and there is no other pronunciation is an exception: *beetovenian* [betoveniʲˈan]; *jeep* [ʒip].

• [i-i], [u-u] between vowels an unsyllabic element develops [ʲ] respectively [ʷ]:

ii → [iʲi] *sf**ii*** [sfiʲˈi]

uu → [uʷu] *perpetuu* [perpˈetuʷu];

– the situations when the two *i* come from the composition, by means of neological prefixes, are exceptions (*anti-*, *semi-);* in this case no unsyllabic element is developed: *anti**i**mperialist* [antiimperiʲalˈist]; *semi**i**ndustrializa* [sˈemiindustriʲalizˈa]

• [a-e] between the vowels an unsyllabic element is developed [ʲ] only in old words: *aer* [ˈaer] or [ˈaʲer] etc.

• [i-o] an unsyllabic element is developed [ʷ]

io → [iʷo] *audio* [aʷˈudiʷo]

• [i-u] an unsyllabic element is developed [ʷ]

iu → [iʷu] *ştiutor* [ʃtiʷutˈor];

– exception: if *u* is under stress the unsyllabic element is not developed:

iu → [iu̯] *naț**iu**ne* [natsiˈune]

• [i-e], [i-a] an unsyllabic element develops [ʲ]

ie → [iʲe] *famil**ie*** [famˈiliʲe]

ia → [iʲe] *patr**ia**rh* [patriʲˈarh]

– exception: if the hiatus *ia* comes from words formed with composition elements *anti-* and *semi-* no unsyllabic elements develops:

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

*anti**a**eri**a**n* [ˌantiaeriʲˈan], *semi**a**nalizabil* [sˌemianalizˈabil]

• [u-a], [u-ă], [u-o], [u-â] an unsyllabic element develops [ʷ]

ua → [uʷa] *accentu**a*** [aktʃentuʷˈa]
uă → [uʷə] *ingen**uă*** [indʒˈenuʷə]
uo → [uʷo] *virtu**o**s* [virtuʷˈos]
uâ → [uʷɨ] *l**uâ**nd* [luʷˈɨnd]

• [e-a], [o-a] an unsyllabic element develops [ᵉ̯] respectiv [ᵒ̯]

ea → [eᵉ̯a] *cre**a**t* [kreᵉ̯ˈat]
oa → [oᵒ̯a] *co**a**gula**t*** [coᵒ̯agulˈat]

• [a-i], [e-i], [o-i], [u-i] an unsyllabic element develops [ʲ]

ai → [aʲi] *în**ai**nte* [ɨnaʲˈinte]
ei → [eʲi] *pe**i**saj* [peʲisˈaʒ]
oi → [oʲi] ov**oi**d [ovoʲˈid]
ui → [uʲi] *pu**i**şor* [puʲiʃˈor]

– exception: in the *ai* hiatus in the words composed with the prefixes *contra-, extra-, intra-, supra-* no unsyllabic element develops: *contr**aa**tac.*

• [a-u], [e-u], [o-u], [â-u], [ă-u] an unsyllabic element develops [ʷ]

au → [aʷu] ***au**rit* [aʷurˈit]
eu → [eʷu] *gr**eu**tate* [greʷutˈate]
ou → [oʷu] *r**ou**ra* [roʷurˈa]
âu → [ɨʷu] *r**âu**şor* [rɨʷuʃˈor]
ău → [əʷu] *gr**ău**nte* [grəʷˈunte]

– exception: in the *ou* hiatus which occurs in words composed with neological pseudo prefixes *auto- , micro-, hipo-, amino-, mono-, pseudo-, termo-, bio-, agro-* etc. no unsyllabic element develops.

• [ă-i], [â-i] an unsyllabic element is introduced [ʲ]

ăi → [əʲi] *c**ăi**na* [kəʲinˈa]
âi → [ɨʲi] *mâr**âi*** [mɨrɨʲˈi]

In conclusion: alternations are specific to the Romanian language, and when a hiatus is formed in their position, a linking unsyllabic element always develops between the vowels or the semivowel that will be altered. The same situation occurs between the root and the ending.

4. MODELLING OF THE PHONETICAL RULES

4.1. The linguistic resources used

Designing of the phonetical transcription module is based on the extant thesaurus of linguistical knowledge, TEZAROM. This, alongside the rich

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

vocabulary, also contains formalized morphology and syllabication, as well as useful information on stress and its migration in flexion. From the linguistical resource thesaurus we shall extract only the necessary information for the designing of the linguistical analysis module (PLN – Natural Language Processing).

4.2. The equivalence method

The phonetical rules presented cover 80% of the phonetical transcription, but as already shown, there are situations when the rules are insufficient as no univocal relation between the grapheme and the phoneme can be established. In order to solve this inconvenience an *equivalence method* has been designed. This method has two components: one implicit, where a known transformation of the graphemes takes place, and one explicit, where the transformation of the graphemes must be indicated through a phonetic attribute associated with the word.

We remind that spelling in Romanian language also presents aspects derived from loans, such as double consonants, double vowels, consonants that can be equaled with a group of phonemes (*k, x, q*). The *grapheme – group of graphemes* correspondence is always the same for the majority of the graphemes (the *implicit equivalence* applies here), but there are situations when this correspondence cannot be unequivocally established (the *explicit equivalence* at word level applies here). It is the case of the foreign words (*background* [bˈækgrawnd]), words that didn't fully adapt to the phonetic spelling of the Romanian language (*foehn* [fœn], *bleu* [blœ]) and of the words which can be spelt by means of graphemes that can be associated with different phonemes when transcribed (*e**x**clus* [e**ks**klˈus], *e**x**amen* [e**gz**ˈamen]).

4.3. The phonetical transcription

An important step in the realization of the project is the *phonetical transcription* at word level, and that was our primary goal in this stage. Transcription can be considered the foundation stone for every TTS system. Of course, in order to perform a natural reading of the texts, a linguistic analyzer must know how to perform several analyses, starting with the text processing, removing the ambiguity of homophones, contextual analysis, syntax etc. to determining the intonation of the text to be read. Without an accurate transcription all this will fail.

Observing classical grammar in formalized models up to the day (morphology, syllabication, stress) proved the indubitable relation between the different chapters of grammar. The formalization of the phonological rules, presented in the previous chapter, is based on this relation. We demonstrated that with the hiatus the possibility exists that the vowels are accompanied by an element unsyllabic which has no written correspondent, but playing an important role in

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

pronunciation. Morphological information often helps us recognize this fact. When we presented the hiatus, we made an important observation: the alternations and the endings corresponding to flexion are specific for the Romanian language and, because of this, when two vowels meet in such a situation, an unsyllabic linking sound will be produced (*ah**ee**ni* [ahe^{j}$^{|}$eni̯], *idee* [ide^{j}e]).

In Romanian, vowels can be found in different phonetic situations in the cases when they can determine the way words are pronounced. If the signs represent semivowels these will unconditionally accompany a vowel and together they will form diphthongs and triphthongs. In case they represent vowels they may or may not be accompanied by unsyllabic elements.

Recognition of these groups can easily be done based on the way *syllabication* is done and knowing the *position of the stress.* Only one vowel can exist in a syllable. If there are several vowels that means that the syllable contains diphthongs or triphthongs, and we have to determine the vowel in them. If the simple recognition of the diphthong or triphthong is insufficient in order to determine the vowel, the stress will then help us (the case of the *ii* diphthongs [ij] / [ji̯] or *iu* [iw] / [ju] at the end of the word).

4.4. Phonetical rules at syllable level

A syllable can be composed of graphemes that represent consonants and vowels, and among vowels only one can really be a syllabic vowel. The analysis of the syllabized words shows that no more than 4 segments can exist in one syllable.



Fig. 2

where: SegC1 = {C}…
SegC2 = {C}…
SegV1 = {V}…
SegV2 = {V}…

One consonant segment cannot consist of more than three consonants and the vocalic one cannot have more than three vowels.

The fourth vocalic element in the first case and the third vocalic element in the second case only occur at the end of the word. In this situation, if the vocalic element is noted with the letter *i,* this can represent a deafened *i* [i̯] (*frați* [fratsi̯])

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

and, in this situation, it generally indicates the plural with an article or may indicate the character of the palatal consonants [tʃ, dʒ, k', g']: *arici* [arˈitʃ].

If the consonantic segments are formed of the palatal consonants [tʃ, dʒ, k', g'.] followed by *i* or by semivocalic *e*, they only participate in indicating the nature of the consonants: *ceai* [tʃaj].

The diphthong *iu* [ju, iw] presents an ambiguity that can be lessened through knowing the position of the stress within the word. If the syllable is the last one in a word and the stress occurs on the vowel *i* in the diphthong, then *u* is a semivowel. In other situations, *u* is a vowel and *i* a semivowel: *abag̲i̲u* [abadʒˈiw]; *am̲o̲niu* [amˈonju].

The diphthong *ii* [ij, ji] presents an ambiguity that can be lessened if we take into consideration the proximity of the precedent syllable. If the precedent syllable ends with a vowel, then we have the diphthong [ji] (semivowel + syllabic vowel). If not, we have [ij] (syllabic vowel + semivowel or, by a different interpretation, syllabic *i*): id̲e̲ii [idˈeji], copi̲i̲ [kopˈij].

4.5. Phonetical rules at hiatus level

The hiatus consists of two vowels between which the syllabic cut up occurs. In order to pronounce them, between the two vowels, under certain conditions, an unsyllabic element can develop, and it makes the connection between the two of them and has no graphical correspondent. The analysis of a hiatus means the analysis of the last vowel in the syllable on the left as related to the present syllable. Like we showed above, there are hiatuses always demanding the development of an unsyllabic element and hiatuses that develop such an element in precise situations. In general, if the last vowel in the syllable on the left is one and the same with the last vowel of a prefix or composition element, between the two vowels an unsyllabic element will never develop. Unsyllabic elements can occur if the second vowel is part of a vocalic alternation or if it is part of the ending.
Examples:
– vocalic alternation: *academicia̲n* [akademitʃiʲˈan]
academicie̲ni [akademitʃiʲˈeni̯]
– junction of the root with the flexion: *cree̲z* [creʲˈez]

4.6. Applying the theoretical phonological model

The theoretical phonological model has been successfully applied upon a lexical basis of approximately 80,000 words covering over two million words having flexion. The comparison of the theoretical phonological model with a

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

complex lexical bases resulted in finding some new phonetic situations, which didn't undergo the study and abstract modelling of the phonology of the Romanian language. Thus, groups of vowels (diphthongs and triphthongs) have been detected, which are to be found only in the flexed forms (*înşe**uau***) and which have not been found in the classical documentation. Verifying the regime of the hiatuses in the phonological model led to the improvement of the established rules. Also, some phonemes have been added [*œ,y*]. They are not characteristic to the Romanian language, but they are still found in some words belonging to the usual word stock: *b**leu*** [blœ], *al**u**ră* [alˈyrə]. Foreign words that entered the Romanian language often keep their original spelling and in order to recognize such situations we can use the letter equivalence method (grapheme-grapheme). The infrastructure of this method has been prepared and is translated by assigning some phonetic attributes to the words in discussion and for which an extra equivalence grapheme-grapheme is given. The words that need this method are not many but finding them implies browsing and word-by-word verifying of the vocabulary. Applying the phonological model to the vocabulary indirectly led to verifying the extant morphological and syllabication model's consistency. We consider that the verifying of the phonological model has been rigorously done and the results obtained are edifying: *the phonological model totally answers the phonological transcription of the Romanian language.* As a result of the verification we present a few paradigms phonetically transcribed by the phonological module implemented in TEZAROM.

aripioară **noun, feminine**

	Without article	
	Sg.	**Pl.**
NA	aripjo̯ˈarə	aripjo̯ˈare
GD	aripjo̯ˈare	aripjo̯ˈare
	With article	
NA	aripjo̯ˈara	aripjo̯ˈarele
GD	aripjo̯ˈarej	aripjo̯ˈarelor

familie **noun, feminine**

	Without article	
	Sg.	**Pl.**
NA	famˈiliʲe	famˈilij
GD	famˈilij	famˈilij
	With article	
NA	famˈiliʲa	famˈiliʲile
GD	famˈiliʲej	famˈiliʲilor

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

5. THE METHODOLOGY FOR DRAWING UP THE DIPHONEMES LISTS

The diphonemes lists have been created in order to supply the vocal signal database with the support for the recording of the sound groups (diphonemes) that are found in different phonetic situations. The sound groups are cut up from representative words.

The methodology for drawing up the diphonemes lists observes the position of the diphoneme within the word and the fact that the phonemes that make it up can be part of the same syllable or of different syllables. Thus, the syllabic pause is implicitly observed, as well as the aspiration of the occlusive consonants, the lack of explosion or implosion of the same consonants at the beginning or end of the word.

The lists of diphonemes were grouped on phonemes in order to easily watch and verify the recordings entered in the vocal signal database. The list for one phoneme includes all its combinations with the 51 phonemes from the reduced set, and for each combination one can find the situations connected with the position of the diphoneme in the word (initial, medial, final) and the state of the phonemes to each other (the same syllable or different syllables). The fact that not all the combinations are found in the Romanian language is mentioned. For those that exist the corresponding phonetical situation has been exemplified by a word taken from the extant vocabulary. In the initial stage, taking into consideration that the project must demonstrate the viability of the chosen solutions, we extracted only the diphonemes from known words. Of course, the diphonemes specific to the Romanian language can also be found in some other situations, that differ according to the word, namely those formed at the meeting of two words when they are uttered together, within syntactic phonetics. The name of the diphoneme is formed by the names of the two phonemes separated by underscore (e.g. ***m_a_***). Information is added to this name, such as the position of the diphoneme within the word (**b** – beginning of the word, **m** – middle of the word, **f** – end of the word) and whether the phonemes are part of the same syllable or not (**s** – same syllable; **d** – different syllables).

Of course, there are not diphonemes for all the positions. Some are impossible, others have not been yet determined, namely the majority of those are formed by the union of words while they are pronounced and which do not have a correspondent in isolated words. Taking into consideration the large number of diphonemes (over 12,000), leaving them aside, for now, is justified. Identification of the phonemes, cutting them up and verifying them implies a large time span. Determination and verification of the diphonemes was performed by using instruments into the medium of the thesaurus implemented – TEZAROM. The vocal base established from the recording of the diphonemes is to be implemented.

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)

6. CONCLUSIONS

The research performed can be used for both setting up a vocal base standing on phonemes and for the realization of a vocal base consisting of syllables with the corresponding phonological analysis modules. Of course, the setting up of a good TTS system also implies a contextual analysis, syntax and prosody. In this paper we presented only the phonological model implemented in the project ***Text Based Synthesis of Speech in the Romanian Language Based on an Expert Linguistical System for Multisensory Interfaces*** **(INFOSOC 2004).**

REFERENCES

Avram, A., 1956, "Contribuţii la studiul fonologiei limbii române", *Studii şi cercetări lingvistice*, 193–204.

Avram, A., 1958, "Semivocalele româneşti din punct de vedere fonologic", *Studii şi cercetări lingvistice*, 7–14.

Dascălu-Jinga, L., M. Teodorescu, A. Ulivi, 1988, *Antologia fonetică a limbii române,* Bucureşti, Institutul de Cercetări Etnologice şi Dialectologice.

Dicţionarul ortografic, ortoepic şi morfologic al limbii române, 1982/2005, ediţia a 2-a, Bucureşti, Univers Enciclopedic.

Dicţionarul explicativ al limbii române, 1975/1996, ed. a 2-a, Bucureşti, Univers Enciclopedic.

Dutoit, T., 1997, "High-Quality Text-to-Speech Synthesis: an Overview", *Journal of Electrical & Electronics Engineering: Special Issue on Speech Recognition and Synthesis,* 17/1, 25–37.

Gaudinat, A., J. P. Goldman, E. Wehrli, 1998, "Le système de synthèse FIPSVox: syntaxe, phonétisation et prosodie", XXIIèmes JEP, Martigny.

Marcu, F., C. Maneca, 1978, *Dicţionar de neologisme,* ediţia a 3-a, Bucureşti, Editura Academiei Române.

Peev, L., L. Bibolar, E. Jodal, 1996, "Un model de formalizare a morfologiei limbii române", Simpozionul internaţional „Limbaj şi tehnologie", Bucureşti.

Peev, L., L. Bibolar, E. Jodal, 1997, "A Formalization Model of the Romanian Morphology", International Workshop "Speech and Computer" SPECOM'97, Cluj-Napoca.

Ribeiro, R., L. Oliveira, I. Trancoso, 2003, "Using Morphosyntactic Information TTS Systems: Comparing Strategies for European Portuguese Spoken Language Systems Lab" (electronic edition).

Sfîrlea, L., 1970, *Pronunţia românească literară. Stilul scenic,* Bucureşti, Editura Academiei.

Sproat, R. (ed.), 1997, *Multilingual text analysis for Text-to-Speech Synthesis: The Bell Labs Approach,* Boston, MA, Kluwer Academic Publishers.

Stan, I. T., 1996, *Fonetica,* Cluj-Napoca, Editura Presa Universitară Clujeană.

Şerban, F., L. Peev, L. Bibolar, D. Bucerzan, 1996, "Banca de date a limbii române – fonetică şi fonologie", in: D. Tufiş (ed.), *Limbaj şi tehnologie,* Bucureşti, Editura Academiei Române, 157–160.

Şerban, F., D. Bucerzan, L. Peev, L. Bibolar, 1998, "Database of the Romanian Language"*, Românìa orientale,* 11, 99–112.

Şerban, F., L. Peev, L. Bibolar, 2000, "La base de données de la Langue roumaine", *Terminometro: La terminologie en Roumanie et en République de Moldova,* hors-série, 4, Atlas-Clusium, 40–42.

Tătaru, A., 1999, *Dicţionar de pronunţare a limbii române,* ediţia a 2-a, Cluj-Napoca, Clusium.

Vasiliu, E., 1965, *Fonologia limbii române,* Bucureşti, Editura Stiinţifică.

BDD-A243 © 2006 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.31 (2026-11-27 17:12:27 UTC)