ABOUT PROSODIC PHRASING OF THE NOUN PHRASES IN SPEECH

Vasile APOPEI

Institute of Computer Science of the Romanian Academy Iasi Branch
E-mail: vasile.apopei@iit.academiaromana-is.ro

This paper aims to highlight the fundamental elements of syntactic structure that underlies prosodic phrasing of utterances in Romanian. The present study investigates the relation between the syntactic structure of the text, especially the noun phrases structure, and the prosodic hierarchy of the uttered text. The analysis covers aspects of prosodic phrasing for short and long noun phrases, with different structures and type of modifiers, in different syntactic functions (subject or complement).

Key words: prosodic phrasing, noun phrases, prosodic domains, functional labels.

1. PROSODIC DOMAIN HIERARCHY

Analysis and modeling of prosodic elements in this research is based on the autosegmental metrical framework using the ToBI system (Pierrehumbert [18, 19]; Ladd [15]; Gussenhoven [9]; Beckman et al. [5]) and hierarchical organization of prosodic domains (Selkirk [20], [23]; Nespor & Vogel [17]; Ladd [16]; Truckenbrodt [26]; Wagner [27]; Ito & Mester’s [10, 11]). The prosodic hierarchy framework states that the intonation contour can be represented by a hierarchy of prosodic units, which are realizations of some types of phonological domains. The F0 contour of particularly higher order prosodic domains consists of a sequence of melodic elements that are within the lower level domains from this hierarchy. Consequently, different types of phonological domains (prosodic) are defined and presented with their ranking. According to the Strictly Layered Hypothesis (SLH) in a tree of utterance, any level cannot be skipped or repeated (Selkirk [21]). In the further developments of the SLH was accepted that the constraints imposed of this hypothesis may be violate. Level skipping was introduced in the model intonation by Ito & Mester’s [10] and Selkirk’s [22]. Level repetition by recursion of prosodic domains was demonstrated by Ladd [16], Wagner [27], Ito & Mester’s [11].

Prosodic modeling used in our research group [12, 13, 14] is based on a hierarchy of prosodic domains similar to that used by Selkirk [23] for phrasing intonation in English, and Ito & Mester’s [11] for phrasing intonation in Japanese (Fig. 1)

```
Utterance
| Intonational Phrase (IP)
| Major Phrase (MaP) - also known as intermediate phrase (ip)
| Minor Phrase (MiP) – also known as Accentual Phrase
| Prosodic Word (PW)
```

Fig. 1 – The prosodic hierarchy used by Selkirk in [23] for English.

To annotate the tonal events from the F0 contour we use the ToBI labels system [5, 12, 14, 15]. Sun-Ah Jun [24, 25] said that “prosodic structure is distinct from the syntactic structure, but it is projected from syntactic structure, conforming to the properties of prosodic structure”. Even if there are utterances for the
prosodic phrasing that do not correspond with the projection of syntactic structures, in the neutral speech of the narrative texts noun phrases (NP), verbal phrases (VP), adverbal phrases (AdvP), adjectival phrases (AP), and prepositional phrases can be related to intonational MiP/MaP.

To describe an intonational contour, at each level of the prosodic hierarchy, we used functional labels, whose significance is given below [12, 13, 14]:

- PH (PUSH) and PO (POP) annotate the delimitative prosodic domains. In neutral intonation, a PUSH domain is more prominent than a POP one.
- F (FOCUS) is assigned to a prosodic domain whose target tone reaches the highest pitch level in an affirmative statement. It is used for describing the F0 contour at PW and MiP level;
- L (LINK) annotates a prosodic domain between PH and PO. It is used for describing the F0 contour into MiP and MaP level;
- PU (POP-UP) is assigned to a particular type of POP domain that indicates the continuation of utterance. It is used to described the intonational contour of the PW and corresponds to the descendent-ascendent final contour (Dascalu-Jinga [7], [8]).

For description in the text mode of the intonational contours with this functional labels, we use the following delimiters: / / for PW; (…) for MiP; [… ] for MaP; {…} for IP. For clarity of description we use the separator ’/ ’ only when it is not overlaped by a higher order separator, such as [{Acestph+sportivpu}

The present study investigates the relation between the syntactic structure of the text, especially the noun phrases structure, and the prosodic structure of the uttered text. This analysis falls within our research group interest [4, 6, 12, 13, 14] to improve the syntactic-prosodic interface for Romanian text-to-speech systems.

2. THE STRUCTURE OF ROMANIAN NOUN PHRASES

The Romanian NP [1], [2], [3] has as head a noun in nominative or accusative and his optional adjuncts can form the following structure <determiner> <quantifier> head <modifiers> <possessor>. At text level, the NP can have a syntactic function of subject, complement (object) and/or attribute. The angled bracket < … > is used to indicate the optional adjuncts from the NP structure.

The determinant has the most important position in the structure of a NP and assigns a defined or undefined interpretation for the noun. The determinant is achieved by: the definite or indefinite article; demonstrative adjectives and possessive ones (acest om ‘this man’); indefinite pronouns adjectives (orice copil ‘any child’); cardinal or ordinal numerals (doi copii ‘two children’, a doilea premiu ‘the second prize’); negative pronominal adjectives; possessive clitic adjectives (soră-mea ‘my sister’). In [1, 2, 3], the postnominale demonstratives (N + acesta / celălalt / acela ‘it / other / one’) and the semi-independent demonstratives (N + cel/cea ‘at / the ’), which always appears after the noun with the definite article are called emphatic determiners. Passed in [2] as "a label for components of determiner type that doubles the primary determiner of the nominal phrase ", the emphatic determiner reveals at utterance level an emphatic focus (Copilul celălalt este cuminte ‘the other child is sapient ’.

The quantifier constituent is distinguished only in the determinant presence (aceşti doi copii ‘these two children’, niciun alt copil ‘no other child’), by the words/sintagms with quantitatively understanding (cardinal numerals, pronouns and pronominal adjectives, undefined adjectives with quantitative value). In absence of the determinant, the word with quantitatively understanding has the determinant function (doi copii ‘two children’). The presence of the following words [3]: tot, toată, toţi, toate, tuturor, amândoi, amândouă, amânduror, rustrei, rustrele, întreg, întreaga, întregului, întregii, before the determinant or at the final NP, are named external quantifiers (toţi aceşti copii ‘all these children’). Uttering the NP with external quantifier before determiner the intonational focus is on the external quantifier. Uttering the NP with an external quantifier at the end of a NP (Fig. 5), reveals the phonological separation of the external quantifier. The external quantifiers make a phonological phrase with the VP that follows ([Aceştii studenţi foarte bine pregătiţi] [toţi merg la concurs.] ‘These students very well prepared all go to the contest’).

The modifier is an optional constituent of the NP which semantically characterize the nominal head (om bun ‘good man’). Modifiers are achieved by: <degree adverbs> + adjectives/participle (o casă <foarte> mare ‘a big house’); reflexive pronouns (băiatul insuşi ‘boy himself’); ordinals (premiul al doilea ‘second
prize’); prepositional groups (băiatul din dreapta ‘the boy from right’); nominal group denominatives (profesorul Popescu ‘the Professor Popescu’); verbal phrases (VP) with the nonpredicative verb head (copii alergând pe nisip ‘children running on the sand’); relative clauses (băieții care au venit ‘boys who came’). Uttering the NP with a modifier achieved by an adjective/participle with the superlative or comparison degree put the focus on the degree (Era un pianist extraordinar de talentat. ‘He was an extraordinarily gifted pianist’, Este elevul cel mai silitor din clasa ‘He is the most studious student in the class.’).

The possessor is an optional constituent of the NP, which establishes a relationship of possession of the head (cartea Mariei ‘Maria’s book’, caietul studentului ‘student’s notebook’). The possessor constituent is achieved through: a noun/pronoun in the genitive (cartea fetei ‘girl’s book’, cartea ei ‘her book’); the prepositional group introduced by the preposition a (mașina a doi oameni ‘car of two people’); the dative clitics of (cartea-i ‘the book’s’); possessive adjective (cartea mea ‘my book’) or relative sentence (cui mi-ai spus ‘whom you told me’).

3. AN ANALYSIS OF THE PROSODIC PHRASING OF THE NOUN PHRASES IN SPEECH

This analysis aims to highlight the fundamental elements of syntactic structure that underlies the prosodic phrasing of utterances in Romanian. In this section, I outline how the NP structure and the syntactic structure of sentences influence prosodic phrasing of the utterances. In the first part of this section I will highlight aspects of prosodic phrasing for short noun phrases (NP composed from <determiner> <quantifier> head <single-word modifiers>) and in the second part there will be highlighted aspects of prosodic phrasing for long noun phrases (NP composed from <determiner> <quantifier> head and multiple-word modifier/multiple modifiers). In this paper, the sintagm ‘single-word modifiers’ will be used to refer a modifier achieved by a single word.

For this analysis, I selected a set of 100 sentences with different structures for NP. Also, each NP was included in narrative sentences with syntactic functions of subject and complement (the object). These sentences were uttered “neutral” by two native Romanian speakers. All recordings were monophonic and were obtained using Canyon CNR-MIC1 microphones and the Praat program (www.fon.hum.uva.nl/praat), used for sound recording and speech processing, at a sampling rate of 11 KHz and a bitrate of 176 Kbps. The same program was used for handmade labelling of the pitch accents and the constituents of the prosodic domains.

3.1. Analysis of short noun phrases

The analysis of the aspects of prosodic phrasing for the short NP was performed on sentences with NP composed from determiner, quantifier, head and single-word modifiers. These structures of NP’s have been included into narrative SVO sentences, with syntactic functions of subject or object.

The first two analysed utterances refer to the sentences which used the determiner + head NP structure (acest sportiv ‘this sportsman’), with subject syntactic function in the first case and complement in the second case. In the first case the utterance of the text Acest sportiv este bine pregătit. ‘This sportsman is well trained.’, presented in Fig. 2, has the intonational contour composed of two MaPs.

![Waveform, spectrogram, F0 contour and prosodic phrasing of the sentence Acest sportiv este bine pregătit (‘This sportsman is well trained’).](image-url)
The first MaP corresponds to the subject achieved by the NP *acest sportiv* ‘this sportsman’ and the second MaP corresponds to the nominal predicate *este bine pregătit* ‘is well trained’. The adjectival phrase (AP) *bine pregătit*, from the second MaP determines at prosodic level a MiP, that is subordinated to the verb *este*. The boundary tone of MaP before the predicate is an H- one, and indicates continuation of the utterance. The intonational contour of utterance presented in Fig. 2, can be described with functional labels like: 
\{[Acest PH+F /sportiv PU] PH [este PH (bine PH+F /pregătit PO) PO] PO}\.

In the second case, the utterance of the text: *Am antrenat acest sportiv.* ‘I coached this sportsman.’, presented in Fig. 3, has also two MaPs. The first MaP corresponds to the predicate *am antrenat* ‘I coached’ and the second MaP corresponds to the object achieved by NP *acest sportiv* ‘this sportsman’. The determiner *acesta* carries the focus of this sentence and his F0 contour has the same tonal level at the beginning and at the ending of the determiner.

The intonational contour of utterance presented in Fig. 3 can be described with functional labels like: 
\{[Am antrenat PH [acest PH+F /sportiv PO] PO}\.

The next analysed utterance correspond to the text which used the determiner + quantifier+ head NP structure *oricare trei puncte* ‘any three points’ with subject syntactic. The prosodic phrasing of the sentence *Oricare trei puncte pot forma un triunghi.* ‘Any three points can form a triangle’ presented in Fig. 4, is composed of two MaPs. The first MaP corresponds to the subject achieved by the NP *oricare trei puncte* ‘any three points’ and the second MaP corresponds to the predicate *pot forma* ‘can form’ and the object *un triunghi* ‘a triangle’.

The quantifier presence in the NP structure determines at intonational level the separation of the noun with the quantifier into a MiP. The predicate, achieved by the modal verb *pot* and the adjunct *forma* reveals at prosodic level a MiP. The boundary tone of MaP before the predicate is an H- one, and indicates continuation of the utterance. The intonational contour of utterance presented in Fig. 4, can be described with functional labels as: 
\{[Oricare PH+F [trei/ puncte PU] PO} \{[pot PH+F /forma PO] PH+F un triunghi PO\}. 

Fig. 3 – Waveform, spectrogram, F0 contour and prosodic phrasing of the sentence *Am antrenat acest sportiv* (‘I coached this sportsman’).

Fig. 4 – Waveform, spectrogram, F0 contour and prosodic phrasing of the sentence *Oricare trei puncte pot forma un triunghi* (‘Any three points can form a triangle’).
The analysis of intonation for utterances with short NP, reveals that prosodic phrasing for NP depends on the syntactic function of the head. If the NP is a subject, these groups form at prosodic level a MaP and have the boundary tones H-. When the NP is a complement, this is grouped at prosodic level into a MaP and have the boundary tones sequences L-L%.

3.2. Analysis of the long noun phrases

The analysis of the aspects of prosodic phrasing for the longer NP was performed on sentences with NP composed from <determiner><quantifier> head and longer modifier/multiple modifier. These structures of NP’s have been included into narrative SVO sentences, with syntactic functions of subject or object.

The utterance of the sentence with SVO structure *Aceşti elevi foarte bine pregătiţi toţi merg la concurs*. ‘These students very well prepared all go in the competition’, presented in Fig. 5, has the intonational contour composed of two IPs. The first IP corresponds to the subject achieved by a part of NP *aceşti elevi foarte bine pregătiţi* ‘these students very well prepared’. The second IP corresponds to the external quantifier *toţi* ‘all’ and VO sequence. The intonational contour corresponding to the first IP is composed by two MaPs. The first MaP corresponds to the *determiner & head* from the NP structure. The second MaP corresponds to the *longer modifier* (degree adverbs & adjective) and is composed by two MiP corresponding to the degree adverbs (*foarte bine* ‘very well’) and adjective (*pregătiţi* ‘prepared’).

The boundary tone of IP before the predicate is an H- one and indicates continuation of the utterance. The intonational contour of the utterance presented in Fig. 5, can be described with functional labels as: \{[Aceşti\textsubscript{IP}/elevi\textsubscript{PO}] [(foarte\textsubscript{IP}/F/bine\textsubscript{PO})\textsubscript{PU} pregătiţi\textsubscript{PU}]\} {toţi\textsubscript{IP}/F/ merg\textsubscript{L}/ la concurs.\textsubscript{PO}}.

The next two analysed utterances refer to the sentences which used the *determiner + head + two modifiers* NP structure (*camera cea mai îngrijită şi mai spaţioasă* ‘the most neat and most spacious room’), with subject syntactic function in the first case and complement in the second case. In the first case the utterance of the sentence with SVO structure *Camera cea mai îngrijită şi mai spaţioasă a fost închiriată*. presented in Fig. 6, has the intonational contour composed of two MaPs.

The boundary tone of IP before the predicate is an H- one and indicates continuation of the utterance. The intonational contour of the utterance presented in Fig. 6, can be described with functional labels as: \{[Camera\textsubscript{IP}/cea\textsubscript{PO}] [(mai îngrijită\textsubscript{IP}/F/si mai spaţioasă\textsubscript{IP})\textsubscript{PU}]\} {a\textsubscript{PO} fost\textsubscript{IP}/F/ închiriată\textsubscript{PU}}.
The first MaP corresponds to the subject achieved by the NP and has the continuation H- boundary tone. The second MaP corresponds to the nominal predicate a fost închiriată. The intonational contour corresponding to the first MaP is composed by three MiPs. The first MiP corresponds to the determiner & head from the NP structure (camera ‘the room’). The second MiP corresponds to the first modifier (cea mai îngrijită ‘most neat’) and the third MiP corresponds to the second modifier (mai spaţioasă) including the conjunction şi ‘and’. The intonational contour of the utterance presented in Fig. 6, can be described with functional labels as: \{CameraPH+F (cea mai PH/ îngrijităPO)L(şi mai/ spaţioasă) PO\[a fost/ închiriată\]PO\}. As it results from the analysis of intonational contours in Fig. 6, the modifiers achieved by the adjectival phrases (AP) are separated at phonological level to MiP.

The next analysed utterance corresponds to a sentence which has the NP as complement. The utterance of the sentence with SVO structure Am închiriat camera cea mai îngrijită şi mai spaţioasă. ‘We rented the most neat and most spacious room.’, presented in Fig. 7, has the intonational contour composed of three MaPs. The first MaP corresponds to the verb and head of the NP Am închiriat camera ‘we rented the room’. The second MaP corresponds to the first modifier (cea mai îngrijită ‘most neat’) and the third MaP corresponds to the second modifier (mai spaţioasă) including the conjunction şi ‘and’.

![Fig. 7 – Waveform, spectrogram, F0 contour and prosodic phrasing of the sentence Am închiriat camera cea mai îngrijită şi mai spaţioasă (‘We rented the most neat and most spacious room’).](image)

The boundary tones for the first and second MaP are of type H-, tones, which indicate continuation of the utterance. The intonational contour of the utterance presented in Fig. 7, can be described with functional labels as: \{(Am /închiriat)PH cameraPU cea maiPH+F/ îngrijităPU şi maiPH/spaţioasă.PO\}.

The next two analysed utterances correspond to the sentences which used the NP with one modifier achieved by a relative clause Turiştii care au venit aseara în staţiune ‘The tourists who came last night in the resort’ with subject syntactic function in the first case and complement in the second case.

The utterance of the sentence with SVO structure Turiştii care au venit aseară în staţiune sunt din Italia ‘The tourists who came last night in the resort are from Italy.’, presented in Fig. 10, has the intonational contour composed of three MaPs: the first MaP corresponds to the head of NP; the second MaP corresponds to the relative clause with VOO structure; the third MaP corresponds to the sentence VO group.

![Fig. 10 – Waveform, spectrogram, F0 contour and prosodic phrasing of the sentence Turiştii care au venit aseară în staţiune sunt din Italia (‘The tourists who came last night in the resort are from Italy’).](image)
The boundary tone of MaP before the predicate is H-one, and indicates continuation of the utterance. Prosodic phrasing of the modifier achieved by a relative clause is correlated with its syntactic structure. In this case, the clause has the structure VOO and intonation contours highlight the achievement of two MiP, the first MiP is made of the verb and the first object, the second MiP is made of a recursive structure and corresponds to the second object. The intonational contour of utterance presented in Fig. 10, can be described with functional labels as: \(\{\text{Tuриптii}_{\text{PH}} \left(\text{care}_{\text{L}} \left(\text{au venit}_{\text{PH}}/\text{asearä}_{{\text{PO}}_{}} \text{in stațiune}_{\text{PH}}\right)\right)_{\text{PH}} \left(\text{sunt}_{\text{PH}}/\text{din Italia}_{{\text{PO}}_{}}\right)_{\text{F}}\}\).

The utterance of the sentence with VO structure \(\text{Am văzut} \text{ turiştii care au venit aseară în staţiune.} \)‘I saw the tourists who came last night in the resort.’, presented in Fig. 11, has the intonational contour composed of two MaPs.

The first MaP corresponds to the verb and the first part of NP (\textit{determiner+head}). The second MaP corresponds to a relative clause with VOO syntactic structures. The boundary tone of MaP before the relative clause is an M-one and indicates continuation of the utterance with an explanation. The intonational contour of utterance presented in Fig. 11, can be described with functional labels as: \(\{\text{Am văzut}_{\text{PH}} \left(\text{turiştii}_{\text{PH}}\right)_{\text{PH}} \left(\text{care}_{\text{L}} \left(\text{au venit}_{\text{PH}}/\text{asearä}_{{\text{PO}}_{}} \text{in stațiune}_{\text{PO}}\right)\right)_{\text{PH}} + \text{F}_{\text{PO}}\}\).

The analysis of intonation for utterances with the longer NP reveals that prosodic phrasing for NP depends on the syntactic function of the head. If the NP is a subject, these groups form at prosodic level a MaP/IP and have the boundary tones H-\%H. When the NP is a complement, the head of the NP (\textit{determiner+head}) and predicate are grouped at prosodic level into a MaP and have the boundary tones H-. The prosodic phrasing of the modifier depends of its structures (eg. multiple modifier, relative clause).

4. CONCLUSIONS

The description of intonational contours with functional labels and delimiters allow an easier analysis of the degree to which syntactic groups can be used in the syntax-prosody interface, for the extraction of prosodic prediction rules in text-to-speech synthesis systems. Based on the observations made in this paper and previous results [4, 6, 12, 13, 14, 15] I can highlight some important processing steps for the prosody syntax interface:

(i) the first step is to establish short syntactic groups and sintagms consisting of two, three words (adverbial phrases, adjectival phrases);
  (ii) establishing intermediate syntactic groups (relative clauses, longer noun phrases, verbal phrases) which includes word groups made in the first step, and establishing dependency relations between their components;
  (iii) setting words/sintagms that cause the focus on the intonational level (functional words);
  (iv) mapping the word groups established in the previous steps into a hierarchy of prosodic domains.

The longer syntactic constituents determine prosodic phrasing problems. To solve the prediction problem of intonation for such constituents, the algorithm proposed by Watson [28] can be used for determining the prosodic boundaries.
ACKNOWLEDGMENTS

The research presented in this paper was performed in the Institute of Computer Science of the Romanian Academy Iasi Branch research program. The prosodic hierarchy and functional labels used for the description of the F0 contour have been developed in last years in collaboration with my colleague Doina Jitcă.

REFERENCES