SYNTAX-PROSODY INTERFACE FOR ROMANIAN WITHIN INFORMATION STRUCTURE THEORIES

Neculai CURTEANU, Diana TRANDABĂŢ, Mihai Alex MORUZ

Institute of Computer Science, Romanian Academy, Iaşi Branch
{curteanu, dtrandabat, mmoruz}@iit.tuiasi.ro

Corresponding author: Neculai CURTEANU

The following main ideas have been pointed out and put to work within our paper: (a) *Information Structure* (IS) theories (topic-focus, theme-rheme, background-contrast, informational and contrastive focus, focus projection rules, etc.) on text are shown to behave currently as a consistent linguistic tool that can stand behind a correct, language and contextual-depending, mapping of text into speech discourse, on the basis of local-global, textual-intonational structures and markers. (b) Functional and logical-semantics (including pragmatics) approaches for IS modelling of the local and global syntax and discourse, coming from the text side, and reliable intonational units within coherent theories on spoken discourse, coming from the prosody side, are needed. We outline here the SCD (Segmentation-Cohesion-Dependency) and FDG (Functional Dependency Grammar) parsing strategies and linguistic theories. (c) Concerning the IS semantics, this paper takes a first (experimental) step to design an adequate syntax-prosody interface for Romanian, by adapting and applying the Prague School's TFA (Topic-Focus Articulation) algorithm. The contributions we consider worth to be mentioned are: (d1) theoretical and computational local-global parsing frameworks of Romanian, emphasizing the functional and hierarchical organization of linguistic categories and markers: SCD and FDG. (d2) outline of the relationship between textual local-global syntactic-discourse structures and prosodic intonational units; (d3) implementation, for the first time (to our knowledge), of the TFA algorithm for Romanian.

Key words: Syntax-prosody interface, Topic-Focus Articulation (TFA) algorithm, SCD markers, FDG tree, topic-focus, theme-rheme, Information Structure Theory.

1. A TRIBUTE TO EVA HAJIČOVÁ'S SPEECH

Our basic ideas on developing a sound prediction for (Romanian) prosody received (indirectly) a strong, both theoretical and computational support from Eva Hajičová's recent speech at her ACL Lifetime Achievement Award [15]. The survey performed in section 2 of this paper is intended to reveal, from the transparency of certain papers or from the misleading technicalities of others, the golden issues of functional and hierarchical marking and markers for linguistic categories, and the (often hidden) way which their boundaries are settled by. The mappings that linguistically project the categories, together with their boundaries and markers on the language interfaces, particularly on the syntax-phonology interface, represent another major goal to be followed in the present approach.

Our initial intuition was to use the RST (Rhetoric Structure Theory) discourse structures [22] for the general layers of global prosody structures of spoken discourse. Confirmations of our approach for the intensive use of RST discourse markers to design the main (global and sentence-level) lines of prosody are also coming from a series of recent papers having Chiu-yu Tseng as constant author [36], [37]. The use of text discourse theories (including those for Chinese), clause markers and RST discourse markers to predict the prosody at the global level could better suggest the interplay between prosodic intonational units, proven to be difficult to predict for Romanian [1].

As for the local, sentence-level prosody prediction, we grant the phonological approach derived from the intensive use of *information structure* (IS) and its intimate relationship to sentence-level discourse and

prosody. There are well-known studies (starting with R. Jakobson [18]) on the implications between IS (Topic-Focus Articulation – TFA in Prague School's terminology) in sentence-level syntactic structure and prosody. Furthermore, Steedman's CCG (Combinatory Categorial Grammar) [31], [32] pays the due importance to the close relationship between IS, sentence structure, and prosody. Valuable approaches for Romanian on this line come from [10], [11], [34].

It is also known, from *centering theory* [12], the crucial significance of sentence IS for the local-global discourse structure and coherence, in either written or spoken language (see also [4]). This is why we have chosen to model and experiment the local, sentence-level prosody with a first TFA (Topic-Focus Articulation) "approximation" for revealing IS and predicting sentence-level prosody in Romanian. Section 4 exposes the TFA algorithm in [16], adapted and applied to Romanian, and compares the results to an empirical phonological annotation [1] on a set of 150 sentences (see figure 1 in subsection 4.1. for the intended place of our implementation within a general scheme of Romanian prosody prediction).

2. SYNTAX-PROSODY INTERFACE AS AN *IS*-DRIVEN MAPPING

2.1. On the Relation between Focus, Prosody, and Word Order in Romanian

Göbbel [10] gives evidence on the relationship between *argument structure* of the Romanian sentence, word order, and the *focus* (F) *projection* into the prosody of the Romanian sentence, using the phonologic theories of S. Schmerling [26], C. Gussenhoven [13], [14], and E. Selkirk [27], [28] for English and German. Selkirk [28] and other researchers, including E. Göbbel [10], [11], developed an F-projection phonologic counterpart of the syntactic F-bar theory, derived from the IS sentence substructure meanings of topic-focus (theme-rheme, given-new, background-focus). This should be called Focus-*phonetic* (F-phon) *projection theory*, providing F-marking and F-projection rules, at the local, sentence (or Chomsky's *phase* [5]) level, and global, inter-clause (sentence or discourse-segment) level. The approach in [10] is, to our knowledge, a *first effective application* of a phonologic discourse theory (this subsuming a syntactic phonology) at sentence-level for Romanian. Other outstanding achievements on this line of research are the TFA (Topic-Focus Articulation) of the Prague School [30], [17], [16], [15], and the theory around Steedman's work on CCG (Combinatory Categorial Grammar), which has the TFA approach as baseline, but refining and extending it to a much more general framework that relates *directly* the phonologic (phonetic), syntactic (surface), and logical-form (deep level) structures of natural language [32]. We remark that *theoretical prosody* based on multi-dimensional contextual information [34] provides a constructive criticism on the last two discourse theories. Prosodic consequences of the logical-oriented (textual) discourse theory of Hans Kamp's DRT (Discourse Representation Theory) [21] on spoken discourse are also outlined. An interesting text preprocessing procedure for Romanian speech synthesis is proposed in [2].

Coming back to Göbbel's work in [10], [11] and to the research line in [13], [14], [28], the accentual patterns of this focus-theoretic approach bring evidence that, in double-object clause constructions, the *direct* (internal) *arguments* can not be represented as *syntactic adjuncts.* In *broad focus* context, *arguments* and *adjuncts* are prominent, while *predicates* may be deaccented if they have an internal argument [11].

2.2. Sentence Prosody: Intonation, Stress, and Phrasing

The current approaches to prosody modelling enclose IS categories and features such as *topic-focus* or *given-comment* (Prague School), *theme-rheme* and *background-focus* [32], [28], [29] etc., categories which are considered (almost unanimously) as also belonging to the *syntactic representation* of the sentence.

Selkirk [28] investigates the relationship between intonation, phrasal rhythmic patterning, and prosodic phrasing, with the purpose of correctly predicting and relating the sentence prosody to its syntactic and discursive meaning. Selkirk [28] proposes an intonation grammar which aims to establish the (informational and/or contrastive) *focus marking* and *projection* within an *argument-structure*-based account of the sentence and Chomsky's *phase* syntactic category [5] (see also [29]). Concerning the *focus projection theory*, we point out two research lines competing with Prague School's TFA: *restricted* [26], [13], [14] ***vs.*** *extended focus projection* [26], [28], [32]. Selkirk and Kratzer [29] give several interesting proposals concerning the computation of a phrase stress in the context of *informational focus* (F), *contrastive focus* (FOC), and

Chomsky's *phase*. It follows naturally that FOC, F, F-projection rules, and the organization of sentences into *phases,* all play a crucial role in defining sentence-level patterns of phrase stress and prosody. The same is true for the relationship between spoken, intonational cues and syntactic-rhetoric textual boundaries used in spoken and written discourse segmentation. The segmented units are based on the *attention-intention* focus and *given-new* (topic-focus in TFA terminology) notions, understood as syntactic and phonologic-prosodic markers of the language, either at local or global level.

2.3. Local and Global Intonational Units in Speech Discourse

In our approach to prosody modelling, the syntax-prosody language interface is concurrently "approximated" with textual and intonational units. When analyzing the dimensions and distinctions of phonological intonational units of spoken discourse, S. Calhoun [3] advocates that there are *three layers* affecting the intonational contour: *global extrinsic*, *local extrinsic*, and *intrinsic*, and that the theme-rheme (or topic-focus) structures may lie on the *local extrinsic* layer. We consider that our implementation of the TFA algorithm for Romanian (section 4) is situated on this local extrinsic level of intonational units (and textually, on the syntactic clause). At this level, Calhoun et al. [4] provides a remarkable pioneering work on corpus annotation with intonational units.

This *local* level of intonational annotation is extended by the work on *global* phonological discourse level. An interesting support for modelling the speech discourse based on prosodic-phonological entities and markers, relevant for textual discursive units, is coming (surprisingly, or rather not) from Chinese, with significant studies of Chiu-yu Tseng and collaborators [36], [37] (abbreviated hereafter as "Tseng-prosody").

Tseng-prosody specifies that spoken discourse prosody is constituted from *prosodic phrase groups*, whose phonologic boundaries are *prosodic fillers* represented by the prosodic word markers and prosodic phrase-discourse markers. The Tseng-prosody approach builds up the spoken discourse from the following defined *prosodic entities* (rather than phonetic and/or phonological units): syllables, prosodic words, prosodic phrases (or utterances), breath group, and prosodic phrase groups. These constituents are, respectively, associated with *five* break indices. Of special interest for the global-level prosody prediction is the analysis of the relationship between prosodic fillers (intonational markers) and the corresponding textual discourse markers. The value of the Tseng-prosody modelling is that, coming from such a structurally different language, it provides similarly basic intuitions on both the lower layers of prosody, when predicting prosodic words from lexical words, and on the higher layers of the prosody interface, when pointing out the strong correlation of prosodic unit boundaries and discursive phrase-structure markers. It is of special *significance* to mention that *Tseng-prosody* modelling relies and is influenced by the RST discourse parsing of English, intensively using the discourse markers, *e.g.* [23], [24], references closely related to SCD markers and our own approach to Romanian discourse parsing [8].

2.4. Steedman's Combinatory Categorial Grammar and Collaterals

This subsection outlines significant features of the research line represented by Steedman's CCG (the landmark paper [32]) framework and its collaterals. The abounding terminology involved by several *discourse theories* around *information structure* (IS), *e.g.* theme-rheme, topic-focus, topic-comment, given-new, background-focus, contextually bound - contextually non-bound (TFA approach) etc., needs accurate comparative and compatibilization studies.

Steedman's intuitions and reasons support an associated Montague-style compositional semantics, with intonation structures arising from discourse semantics and (more) primitive literal meanings distinguished within IS and speaker-hearer commitment. Steedman's paper [32] is important not only for the effective novelty of the proposed CCG framework, but equally for discussing and relating achievements and conjectures from syntax, IS, discourse semantics, phonology and prosody, logical form (LF) and, most importantly, their interfaces. The principle of *Sense Unit Condition* [27], defined in terms of head dependencies at the LF level, showed the freedom limits within which *intonation* can *diverge* from a traditional *constituent structure*. CCG [32] accomplishes the intuition that the *intonational structure* (of English) is directly *subsumed* by the surface *syntactic structure*, and covers a wide inventory of prosodic tunes and information categories. Here are several significant claims and observations in [31], [32], [33]:

(*a*) Surface structure and IS *coincide*, IS representing the *interpretation* associated to a constituent analysis of the sentence. **(*b*)** *Intonation* coincides with surface structure in the sense that all *intonational boundaries* coincide with *syntactic boundaries*. **(*c*)** The reverse *does not* hold: not all surface syntactic boundaries are *explicitly* marked by intonational boundaries. **(*d*)** The majority of IS boundaries go unmarked by explicit intonational boundaries; IS boundaries may be marked by other, *more subtle,* articulatory markers. **(*e*)** IS constituents and their semantic interpretation provide the logical forms that discourse semantic functions apply to, the boundaries of these IS constituents *lining up* with the intonational-structural boundaries, when these are present. **(*f*)** IS has to be referred from the partial specification implicit *in the tones* in *exactly* the same way as *predicate-argument* relations have to be implicitly derived in the *sequence of words*. **(*g*)** Phrasal intonation and intonational contours are defined in terms of pitch accent(s) tones and boundary tones. Pitch accents are realized as maxima or minima in the pitch contour and coincide with the perceived major emphasis (emphases) of the prosodic phrase. *Boundary tones* mark the *right-hand* side *boundary* of the *prosodic phrase*, differing therefore from the *syntactic* and *discourse markers*, which bound the *left-hand* side of a *textual phrase*. **(*h*)** An *intonational phrase* (or *tune*) is defined as a sequence of one or more pitch accents, ending with a boundary tone. [32] reminds that intonational phrases (phrasal tunes) have been variously associated with specific discourse meanings, social attitudes, illocutionary acts, propositional attitudes and maintenance of mutual belief, various IS notions. **(*i*)** Two independent dimensions of IS are revealed: the *theme* of an utterance and the *rheme* that the utterance contributes to. Certain specific, but not unique, *intonational phrasal tunes* (thus ToBI tags) can be associated (not uncontroversially) with a *theme*, and some *other tunes* with a *rheme*. **(*j*)** The pragmatic, context-dependence meaning of the basic IS pair *theme-rheme* is stated in the following terms of [32 :659]: theme tunes *presuppose* a rheme's *alternative set*, while rheme tunes *restrict* this rheme's alternative set. **(*k*)** [32] describes the technical aspects of CCG framework (principles, rules, examples), which relate directly the surface syntactic constituents to the associated LF*s* and, through IS constituents and a combinatory prosody, to the intonational phrases corresponding to intonational pitch and boundary marking of the theme and rheme. Moreover, [32] investigates the linguistic and semantic nature of theme and rheme, showing that it is the speaker who mainly determines what is the theme and rheme, and what contrasts they embody, and not necessarily the text.

Our present belief is that, for devising an adequate prosody prediction for Romanian, it is necessary to assimilate and adapt the Prague School's TFA and Steedman's CCG experience [32], [4] for the *local prosody* design, compatible with the local-level spoken discourse theories on IS. This should be coupled with the *global-level prosody* approach based on prosodic intonational units (such as Tseng's *prosodic phrase groups*), compatible with the global spoken discourse through, for instance, RST textual discourse [24] and *global linking* algorithm [38, Ch. 7].

3. MODELLING THE LOCAL AND GLOBAL SYNTAX AND DISCOURSE WITH FUNCTIONAL AND LOGICAL-SEMANTICS

This section briefly outlines elements of textual structures that the functional FX-bar theory [7], [9] and SCD (*Segmentation-Cohesion-Dependency*) parsing strategy [6], [8] deal with. Their essential ingredient is the graph-type *hierarchy of* (classes of) *linguistic markers* that are used to distinguish between the local (intra-clausal) and global (inter-clausal and rhetoric discourse) text structures. These markers are used, among other morphological-syntactical information, to obtain the FDG parsing tree of the written corpus.

3.1. The Segmentation-Cohesion-Dependency Linguistic Theory

The SCD (Segmentation-Cohesion-Dependency) linguistic strategy [8] pointed out that the natural language is constructed from *local* to *global structures*. We consider *local structures* to be those structures that build a single finite-clause or a single (finite or non-finite) lexical predication (including both), in sum, finite or non-finite sub-clause and clause-level structures. *Global structures* could be classified into (at least) two main categories [7]: **(1)** global structures composed of finite-clauses or lexical predications using logical operators, syntactic operators (e.g. for the relative clause), and second-order theta-relations (*i.e.* second-order

predicational relations, for the so-called subjective, predicative, direct-completive clauses etc.). **(2)** The usual clause-based global text structures are the sentence, paragraph, section, chapter, etc.

SCD considers four *major lexical categories* (and their functional projections within the FX-bar theory): *the Noun, the Verb, the Adjectival* and *Adverbial. The Noun* (N) and *the Verb* (V) are the only lexical categories that have their own lexical (non-referential) meaning, and they are also saturated (representing their own semantic heads). The other two lexical categories, the *Adjectival* and *Adverbial* (hereafter A) category modifiers, play a central role in the syntactic organization of the *functional* X-bar (FX-bar) general schemes [7], [9].

The SCD markers are functional and relational *lexical categories*, *e.g.* clause-level and discourse collocations (cue phrases, connectors, etc. called *clause* and *discourse markers*), but also lexically empty (covert) functional categories, such as *tense*, *inflection* or the intrinsic presence of the *predicational feature* [6], which can be ascribed, possibly and equally, to each of the major N, V, or A lexical category (and inherited by the XG phrase which that category is heading). The SCD *parsing strategy* defines several representation levels of textual marker classes as boundaries of the XG (X = N, V, A), clausal, inter-clausal, and discursive syntactic structures. These classes are:

- **M0:** the *word dictionary form*, represented by the functional role of morphological-grammatical inflection; it corresponds to the lexical level of each textual word.
- **M1**: *X*1-*level* markers, (X = N, V, A), *i.e.* markers to be applied to the phrase structure *level* (or XG).
- **M2:** markers that introduce a (finite or non-finite) *clause,* or a syntactic category phrase with the semantic head N, V, A.
- **M3:** *inter-clausal* (discourse) markers: functions or relations (when correlated), having as arguments two or more finite clauses.
- **M4:** *discourse* markers, which determine the rhetorical relations that can be established between discourse segments.

M*k* marker classes belong to an established graph-type hierarchy [8 :162, Fig. 1], whose ordering is essential in the parsing process.

3.2. The Functional Dependency Grammar Theory

The Functional Dependency Grammar (FDG) theory [19] attaches (functional) dependencies to a (heavy) annotated text using functional-syntactic connections between the words of a given sentence. The FDG framework was introduced and analysed by Järvinen and Tapanainen [19], [20]. An adaptation and implementation of the FDG formalism for Romanian, benefiting from the use of the SCD markers, is described in [25]. The FDG has two conceptually different components: syntactic-semantic based analysis and structural dependency graphs. The linguistic representation in structural dependency graphs closely follows the Structural Syntax of Tesnière [35]. The essential ideas of Tesnière's Structural Syntax are:

(*a*) The basic element of the syntactic structure is not a word, but a *nucleus* (elementary phrase). **(*b*)** Nuclei of the sentence have mutual directed dependencies, called *connections*. **(*c*)** Every nucleus has one and only one *syntactic head*. The connections thus form a tree where the head element of the main clause is the root of the sentence. **(*d*)** The variation in word-order of the sentence does not affect the structural analysis, since the syntactic functions of the words remain the same. **(*e*)** There is a close parallelism between syntax and semantics, *i.e.* the syntactic structure is motivated by the semantic interpretation rather than by word-order configurations, morphological markings or classical syntactic analysis.

This type of FDG parsed text allows for the extraction of N, V, A headed phrases, using the following method: every N, V or A, together with all its successors, forms a N-, V-, A-phrase with the parent being the head for that phrase. Clauses can be extracted from the FDG parsing tree in a similar manner, considering a finite verb as the clause head.

3.3. Intonational Unit Acquisition from FDG-Trees

Apart from being used as input for the TFA algorithm in Section 4, the FDG parsing is also used for determining prosodic entities for any given sentence, using the following empirically acquired rules:

- **Clause acquisition**: every subtree that has as its root a verb forms a clause, together with all of its descendents, with the exception of the verbal descendents which are themselves heads of verbal groups and their subtrees;
- **Main clause acquisition**: all verbs on the highest level in the FDG tree of a sentence are heads of the main clauses;
- **IP** (*Intonational Phrase*) **acquisition**: each main clause, together with all its subordinate clauses, forms an IP;
- **ip** (*intermediate phrase*) **acquisition**: each syntactic phrase is an *ip*. There are cases when *ip* structures are interrupted by speech break markers (punctuation marks, coordinative conjunctions, etc.)
- **RU** (*Rhythmical Unit*) **acquisition**: *(i)* a parent and its first successor form together a RU (it does not matter whether the successor stands at the left or the right of the parent). The RU also contains the first adjacent descendent of the successor, recursively, to the leaf; *(ii)* Non-adjacent successors are each in separate RUs, in a similar way as in *(i)*; *(iii)* If a parent has two or more adjacent successors, the first successor forms an RU with its parent and the other successors form an RU by itself.

Examples of FDG-based acquisition of intonational units:

(1) [[[O'Brien apucă]$_{RU}$ [sticla de gât]$_{RU}$]$_{ip}$]$_{IP}$ [[[şi turnă în pahare]$_{RU}$ [un lichid]$_{RU}$ [de un roşu aprins]$_{RU}$]$_{ip}$]$_{IP}$.

(2) [[[Lui Winston i se trezeşte]$_{RU}$ [nişte vagi amintiri]$_{RU}$]$_{ip}$, [[ceva ce văzuse demult]$_{RU}$ [pe un perete]$_{RU}$ [sau pe un panou]$_{RU}$]$_{ip}$ – [[o sticlă mare]$_{RU}$, [făcută din lumini electrice]$_{RU}$]$_{ip}$ [[care parcă se mişca]$_{RU}$ [în sus]$_{RU}$ [şi în jos]$_{RU}$]$_{ip}$, [[turnându-şi conţinutul]$_{RU}$ [într-un pahar]$_{RU}$] $_{ip}$]$_{IP}$.

4. TOPIC-FOCUS COMPUTING FOR ROMANIAN

This paper takes a first step to design an adequate syntax-prosody interface for Romanian, by adapting and applying the Prague School's TFA (Topic-Focus Articulation) algorithm [16]. The purpose of this algorithm is to compute the topic-focus (theme-rheme) for a given sentence to obtain, on the basis of sentence IS elements, a better acquisition of the sentence intonational units, and subsequently, an improved naturalness for the assignment of tone and tune phrases to the established intonational units of the sentence.

4.1. The Methodology

The starting point of our prosody prediction system (Figure 1) is the morphologically-syntactically tagged text, with SCD markers and FDG dependencies. From the FDG parsing tree, two directions are derived: **(1)** grouping the FDG branches into syntactical constituents, and **(2)** detecting intonational units from the FDG tree using the acquisition rules presented above.

The *Topic-Focus Articulation* (TFA) *Algorithm* receives as input the syntactic units of a sentence, SCD markers (such as the definite or indefinite article), and certain semantic features. The outcome consists of IS-based units, which are actually syntactic constituents with topic-focus values. The obtained syntactic IS is used to re-arrange the intonational units according to their relevance for the speech dynamics, leading to IS-based intonational units. Applying transformation rules as those proposed by Steedman's CCG for the local level or those drafted by Tseng-Prosody (see Subsection 2.3) for global level, the corresponding ToBI labels could be automatically established with better adequacy.

The right-hand side of Fig. 1 presents the empirical annotations of the spoken text. The manually determined intonational units and the ToBI labels will be compared to the automatically acquired ones in order to evaluate and adjust the TFA procedure against the (currently empirical) annotated gold corpus.

4.2. The Corpus

We selected for recording approx. 150 sentences from George Orwell's novel "1984". This novel was selected because we already have an important number of morphological-syntactical annotations for that corpus (morphology, POS, NP-VP groups, SCD markers, etc.). At syntactic level, we added for each

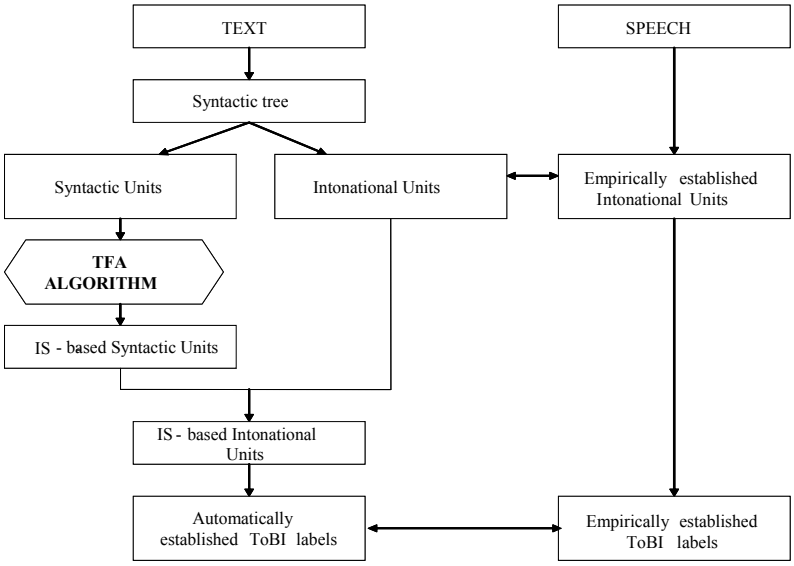


Figure 1. The architecture of the prosody prediction system

sentence an FDG-tree. At phonological level, each sentence was listened to by at least two annotators who added ToBI tone labels and phonological entity boundaries to the spoken texts. This manual annotation of the spoken corpus will be considered a gold standard for the validation of the rules developed to automatically identify the intonational units and for the assessment of the topic-focus within the TFA algorithm output. This is how the annotators identified the intonational units in the two sentences above:

(1) [[[O'Brien]$_{RU}$ [apucă sticla de gât]$_{RU}$]$_{ip}$]$_{IP}$ [[şi turnă în pahare]$_{RU}$]$_{ip}$ [[un lichid]$_{RU}$ [de un roşu aprins]$_{RU}$]$_{ip}$]$_{IP}$.

(2) [[[Lui Winston]$_{RU}$]$_{ip}$ [[i se trezură]$_{RU}$ [nişte vagi amintiri]$_{RU}$]$_{ip}$]$_{IP}$, [[[ceva]$_{RU}$ [ce văzuse demult]$_{RU}$ [pe un perete]$_{RU}$ [sau pe un panou]$_{RU}$]$_{ip}$]$_{IP}$ – [[[o sticlă mare]$_{RU}$]$_{ip}$, [[făcută]$_{RU}$ [din lumini electrice]$_{RU}$]$_{ip}$ [[care parcă se mişca]$_{RU}$]$_{ip}$ [[în sus şi în jos]$_{RU}$]$_{ip}$]$_{IP}$, [[[turnându-şi conţinutul]$_{RU}$]$_{ip}$ [[într-un pahar]$_{RU}$] $_{ip}$]$_{IP}$.

One can easily see that there are several discrepancies, such as the splitting of the subject and the verbal group in separate rhythmic units, or the detection of several IP*s* within the same main clause. This contrastive analysis helps us to improve the intonational units' detection (*e.g.* we study now the possibility of adding preposition or length–based exception cases), keeping in mind however that the manual annotation has been empirically performed.

4.3. The TFA Algorithm

The input for the TFA algorithm consists of FDG trees for Romanian sentences. Besides the morphological annotation of each word, the *semantic features* of constituents are also classified according to their *specificity degrees*: (*i*) *general* – low specificity, contextually non-bound; (*ii*) *specific* – high specificity, contextually non-bound; (*iii*) *indexical* – mid-specificity, contextually-bound. Examples of specificity degrees for *temporal complements* are: general ("*niciodată*", "*mereu*", etc), indexical ("*astăzi*", "*anul acesta*"), specific ("*22 iunie*", "*într-o frumoasă zi de mai*"). The TFA algorithm utilizes the specificity degrees only for verbs and *temporal* / *locative* complements. The position of the verbal direct and indirect arguments within the *systemic order* (SO) is important when computing the sentence *topic* and *focus*. According to [16], the SO for the main kind of complementations in English is:

Time – Actor – Addressee – Objective – Origin – Effect – Manner – Directional.*from* – Means – Directional.*to* – Locative

The SO that we shall use for Romanian within the TFA algorithm is similar, being supported by the observation that, in clauses with 3-valenced verbs, the indirect object usually stands *before* the direct object

for the purpose of avoiding the possessive ambiguity (`I-am dat cartea Ioanei` may be interpreted as `Ioana` being the receiver or the possessor of the book).

Figure 2 presents an example of the algorithm input, where the parsing tree is described in a LISP-style syntax and the output of the procedure consists of the topic-focus assessment for the verb and all its complementations.

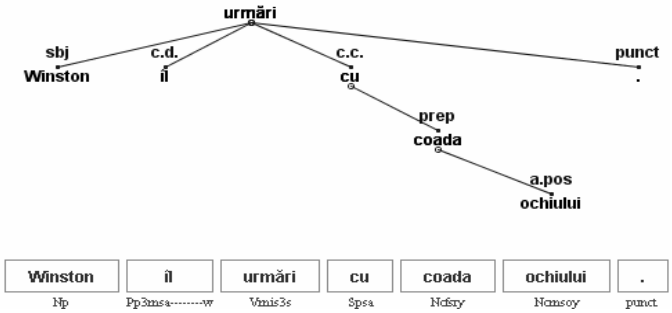


Figure 2a. FDG parsing tree for the sentence "`Winston îl urmări cu coada ochiului.`"

```
verb (topic(f), sem(interm), label(il urmari), ltree(act(topic(f), det(1),
so(1), surf(np), label (Winston))), rtree(mod(topic(f), det(1), so(8),
surf(np), label (coada), ltree(prep(cu)), rtree(apos(ochiului)))))
```

Figure 2b. TFA algorithm input for the sentence "`Winston îl urmări cu coada ochiului.`"

([[[Winston(t), îl urmări(t/f)]$_{RU}$, [cu coada ochiului (t(f))] $_{RU}$] $_{ip}$] $_{IP}$)

Figure 2c. TFA algorithm output for the sentence "`Winston îl urmări cu coada ochiului.`"

Let **S** be the set containing the Verbal Group (VG) [6], [9] and all its complements from a sentence **s**. Let **f** denote the *focus*, **t -** the *topic*, and **t/f** denote ambiguous topic-focus elements; **t(f)** is a special kind of ambiguity (this element is **f** only when there is no other **f** in the reading of the sentence). **S(i)** denotes the *i*-th complementation in the surface order, **verb(S)** is the VG, and **last(S)** is the last element in the surface word order. The **TFA Algorithm** is exposed below:

```
1.  f = S;
    /*all elements are presumed to be focus*/
2.  if(last(S)=verb(S))
    2.1. if(subject(S) is definite)
              f=f-{subject(S)}; t=t∪{subject(S)};
             f=f∪{verb(S)};
              goto 3.2;
    2.2. else if(subject(S) is indefinite)
              f=f-{verb(S)}; t=t∪{verb(S)};
              f=f∪{subject(S)};
              goto 3.2;
    2.3. else
              f=f∪first(S);
              f=f-{verb(S)}; t=t∪verb(S);
              goto 3.2.
3.  else
    3.1.   if(verb(S) is general)
              f=f-{verb(S)}; t=t∪{verb(S)};
           else if(verb(S) is specific)
              f=f∪{verb(S)};
           else
              f=f-{verb(S)}; t/f=t/f∪{verb(S)};
    3.2.   for(i=0 to pos(verb(S)))
    /*for all complements in front of the verb*/
              if(S(i) is touched by 2)
                 continue;
              if((S(i) is indefinite subject) or (S(i) is specific temporal complement))
    /*specific means it is neither general, nor indexical*/
                 f=f-{S(i)}; t/f=t/f∪{S(i)};
              else
                 f=f-{S(i)}; t=t∪{S(i)};
```

```
3.3.    /*to the right of the verb*/
        i. if(pos(last(S))=pos(verb(S))+1 and (last(S) is definite noun or personal pronoun))
/*the verb has only one complement to the right, and that complement is an indefinite noun or a personal pronoun*/
                f=f-{last(S)}; t/f=t/f∪{last(S)};
        ii.if((last(S) is temp or loc) and last(S) is specific)
                f=f∪{last(S)};
           else
                f=f-{last(S)}; t=t∪{last(S)};
                last(S) is moved to the left of the verb in the CD
        iii.let (S(i),S(j)) be the rightmost pair that fails to follow SO
            for(a=pos(verb(S))+1 to i)
                f=f-{S(i)}; t=t∪{ S(i)};
                for(a=i+1 to last(S)-1)
                        f=f-{S(i)}; t/f=t/f∪{S(i)};
                        f=f∪{last(S)};
        iv.if(ii not met and iii not met and last(S) is indefinite)
                f=f∪{last(S)};
        v.for all S(i) where i>pos(verb(S)) and S(i) not touched by i, ii, iii, iv
                f=f-{S(i)}; t/f=t/f∪{S(i)};
4. if all x in S-{verb(S)} are either t or t/f
   4.1. if((verb(S) is t/f) and (last(S) is definite or indexical or personal pronoun))
       f=f-{last(S)}; t(f)=last(S);
   4.2. else
            f=f-{verb(S), last(S)}; t/f=t/f∪{verb(S), last(S)};
5. if at this step, the sentence still contains no focus, then it does not bring any new
        information (it is only a comment on old information).
```

4.4. Examples of Parsing with the TFA Algorithm

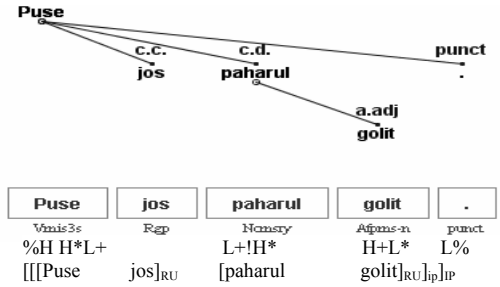
In order to emphasize the importance of the TFA algorithm for the prosody prediction, several examples of the application of the presented procedure are discussed. The figures format used in this section is the following: the left-hand side presents the FDG tree of the sentence, the morphological information of each word (under the word corresponding boxes), the manually marked ToBI labels, and the manually determined intonational units; the right-hand side of the figures presents *the input* of the TFA algorithm, containing all syntactic and semantic information needed for topic-focus computing, and *the output* of the algorithm, combined with the automatically discovered intonational units.

Let's consider thus the sentence "*Puse jos paharul golit.*", with the input information in Fig. 3. The TFA processing steps can be summarized as:

Step 1. All words are marked as being in the focus (*puse*(f), *jos*(f), *paharul golit*(f));

Step 2. After applying 3.1, the verb becomes ambiguous (*puse*(t/f), *jos*(f), *paharul golit*(f));

Step 3. After applying 3.3.**iii**, "*jos*" becomes topic and "*paharul golit*" becomes focus (*puse*(t/f), *jos*(t), *paharul golit*(f)); since no other step can be applied, this is the output of the algorithm.

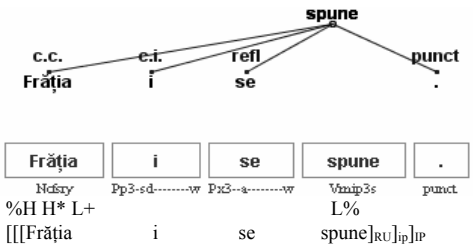


Input *(LISP-style functional syntax)*:
verb (topic(f), sem(interm), label (puse), rtree (loc(topic(f), sem(gen), so(10), surf(advp), label (jos)), obj(topic(f), det(1), so(3), surf(np), label (paharul), rtree(attr(golit)))))

Output *(IS-based units (TFA algorithm output), combined with automatically detected Intonational Units cf. the procedure in § 3.3)*:
([[[puse(t/f), jos(t)]$_{RU}$, [paharul golit(f)] $_{RU}$] $_{ip}$] $_{IP}$)

Figure 3. Corpus information and TFA input for the sentence "`Puse jos paharul golit`".

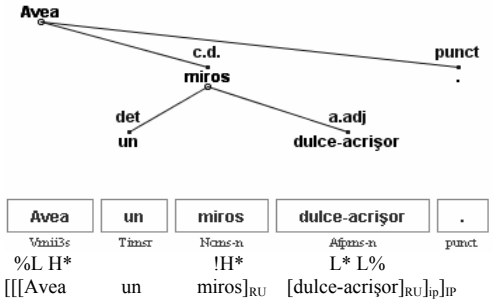
There is a perfect match between the manually extracted intonational units (from the FDG parsing tree) and the automatically acquired ones (presented in the algorithm output). Furthermore, the verb "*puse*", which was determined as ambiguous (t/f) by the TFA algorithm, is further refined as topic, since the adverb "*jos*", a topic, is situated in the same *rhythmical unit* (RU). The focused element in this example is the argument "*paharul*", as head of the RU "*paharul golit*".



Input:
verb (topic(f), sem(interm), label (i se spune), ltree (mod(topic(f), det(1), so(6), surf(np), label (Frăția))))
Output:
([[[frăția(f)]$_{RU}$, [i se spune(t)] $_{RU}$] $_{ip}$] $_{IP}$)

Figure 4. Corpus information and TFA input for the sentence "`Frăţia i se spune.`"

Figure 4 presents an example of mismatch between the empirically defined intonational units and the automatically extracted ones using the FGD rule-based algorithm. While the annotators have decided on a single rhythmic unit, intermediate phrase (*ip*) and intonational phrase (*IP*), the automatic acquisition algorithm has identified two RU*s*: the first one, containing the predicate nominal "*frăţia*", the sentence focus, and the second containing the rest of the sentence (the topic VG). This tendency is justified by the manually planted ToBI labels (the high tone on "*frăţia*"), and by the disrespect of the systemic order SO. We consider therefore that in this case, the automatically labelled units are correct.

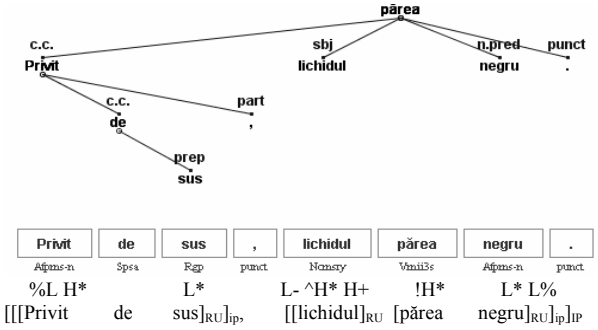


Input:
verb (topic(f), sem(gen), label(avea), rtree(obj(topic(f), det(0), so(3), surf(np), label (miros)), ltree(det(un)), rtree(attr(dulce-acrisor))))

Output:
([[[avea(t), un miros] $_{RU}$ [dulce-acrisor(f)] $_{RU}$] $_{ip}$] $_{IP}$)

Figure 5. Corpus information and TFA input for the sentence "`Avea un miros dulce-acrişor.`"

Figure 5 gives an example of an intonational unit that breaks the topic-focus distribution. The topic-focus computing of the intonational units revealed that the syntactic constituent "*un miros dulce-acrişor*" is a focused structure. When looking at the manual annotation, we observe that this constituent is split into two RU*s* ("*avea un miros*" and "*dulce-acrişor*"). The ambiguity arises since, according to the acquisition rules, the focus of a syntactic structure is its head, while, in this particular case, the head is in a different RU than the rest of the focus. The solution comes with the focus transfer to the second RU, since the argument "*un miros*" is part of a topic RU, developed by a general verb, and also because the right-hand side of the sentence is preferred to bear the focus.

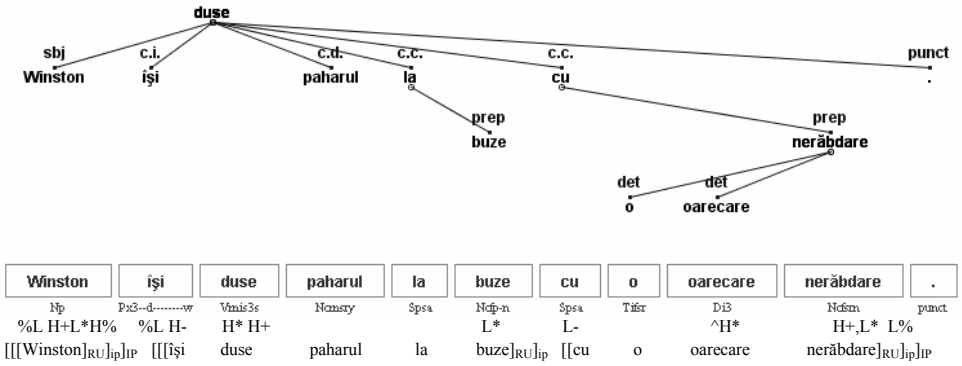


Input:
verb (topic(f), sem(gen), label(părea negru), ltree(mod(topic(f), so(6), surf(ap), label (privit), rtree(mod(topic(f), so(6), surf(advp), label (sus), ltree(prep(de))))), act(topic(f), det(1), so(1), surf(np), label (lichidul))))
Output:
([[[privit de sus(t)]$_{RU}$, [lichidul(t), părea negru(f)] $_{RU}$] $_{ip}$] $_{IP}$)

Figure 6. Corpus information and TFA input for the sentence "`Privit de sus, lichidul părea negru.`"

In the exemple in Fig. 6, the intonational units no longer break the topic-focus separation, but we have another type of contradiction. Within the same RU, namely "*lichidul părea negru*", there are two topic-focus structures: "*lichidul*", which is topic, and "*părea negru*", which is focus. In such a case, we consider "*lichidul*" as focus, since the copula verb behaves similarly to general verbs. This rule is confirmed by the

gold ToBI annotation and by the break that the annotators have felt in the utterance, break indicated by the closure of the intermediate phrase (*ip*) in the manual annotation of intonational units.



Input: verb (topic(f), sem(interm), label (îşi duse), ltree (act (topic(f), det(1), so(1), surf(np), label (Winston))), rtree (obj (topic(f), det(1), so(3), surf(np), label (paharul)), loc (topic(f), sem(gen), det(0), so(10), surf(np), label (buze), ltree(prep(la))), mod (topic(f), det(0), so(6), surf(np), label (răbdare), ltree(prep(cu), det(o), det(oarecare)))))
Output: ([[[Winston(t), îşi duse(t/f)]$_{RU}$, [paharul(t)] $_{RU}$, [la buze(t)] $_{RU}$] $_{ip}$, [[cu o oarecare nerăbdare(f)] $_{RU}$] $_{ip}$] $_{IP}$)

Figure 7. Corpus information and TFA input for "`Winston îşi duse paharul la buze cu o oarecare nerăbdare.`"

In figure 7, the insertion of the VG "*îşi duse*", which is ambiguous within a topic dominated *ip*, justifies the difference between the automatic and manual annotation. In the spoken variant of the sentence, the speaker performed a break before the VG, in order to mark the verb special position within its arguments, break that led the annotators to consider the clause subject in a separate *ip* (and even *IP*). The focus of the sentence goes to its most right-hand side constituent, *viz.* the head "*nerăbdare*" of the modal complement.

The analyzed examples show that the comparison between the automatically and manually annotated intonational units demands reconsideration and refinement of the currently existing rules (in order to replace the empirical annotation with a *gold* one), based on improved syntactic, semantic, discursive, IS, and SO criteria.

5. CONCLUSION

TFA theory and algorithm is a procedure intervening exactly on the core of syntax-prosody interface, with semantic, discursive, and pragmatic surgery tools, bringing clearer rules for establishing the accented-deaccented IS components on the prosodic intonational units of a sentence, for assigning proper intonational stress (ToBI labels). Finding the set(s) of rules by which one could evaluate *the focused* elements of a spoken utterance is *one* of the keys of the *prosody prediction*. This is the main challenge of this paper.

The *second* key of a natural prosody design for Romanian is to establish improved policies (involving variability, contextuality) that assign adequate sequences of *tones* and *tunes* (*i.e.* sequences of ToBI tags) on the *global* and *mid-level* IS of the text [38]. Theoretical (Steedman's CCG [32]) and experimental (Calhoun's framework on text-to-speech annotation [4]) experience are the referring points we are going to pursue.

REFERENCES

1. APOPEI V., JITCA D., *A Set of Intonational Category for Romanian Speech and Text Annotation*. In Proceedings ECIT2006 – 4th European Conference on Intelligent Systems and Technologies, Iasi, Romania, p.117-124, September 2006.
2. BURILEANU D., *Prelucrarea inițială a textului de intrare în cadrul unui sistem de sinteză a vorbirii pornind de la text în limba română*, in "Limba Română în Societatea Informațională – Societatea Cunoaşterii", D. Tufiş, F. Filip, (Eds.), Expert Press, Bucharest, p. 359-379, 2002, ISBN 973-8177-83-9.
3. CALHOUN S., *Phonetic Dimensions of Intonational Categories – the case L + H* and H*. Speech Prosody 2004*, International Conference, Bernard Bel, Isabelle Marlien. (Eds), ISCA Archive, Nara, Japan, March 2004.
4. CALHOUN S., NISSIM M., STEEDMAN M., BRENIER J., *A Framework for Annotating Information Structure in Discourse*. Proceedings of the Workshop on Frontiers in Corpus Annotation II: Pie in the Sky, Ann Arbor, Michigan, p.45–52, June 2005.

5. CHOMSKY N., *Derivation by phase.* In M. Kenstowicz (Ed.), Ken Hale: A life in language. MIT Press, Cambridge, MA, p. 1-52, 2001.
6. CURTEANU N., *Contrastive Meanings of the Terms* "Predicative" *and* "Predicational" *in Various Linguistic Theories* (I, II). Computer Science Journal of Moldova (R. Moldova), **Vol. 11**, No. 4, 2003 (I); **Vol. 12**, No. 1, 2004 (II).
7. CURTEANU N., *Functional FX-bar Theory Extended to Discourse (Rhetorical) Structures.* In 'Intelligent Systems' Conference Volume, H.-N. Teodorescu *et al.* (Eds.), Performantica Press, Iaşi (Romania), p. 169-182, 2005.
8. CURTEANU N., ZLAVOG E., BOLEA C. *Sentence-Based and Discourse Segmentation / Parsing with SCD Linguistic Strategy*. In 'Intelligent Systems' Conference Volume, H.-N. Teodorescu *et al.* (Eds.), Performantica Press, Iaşi (Romania), p. 153-168, 2005.
9. CURTEANU N., TRANDABĂȚ D. *Functional (F)X-bar Projections for Local and Global Text Structures. The Anatomy of Predication.* Revue Roumaine de Linguistique, Romanian Academy Editorial House, Bucharest (to appear).
10. GÖBBEL E., *On the Relation between Focus, Prosody, and Word Order in Romanian*, Romance Languages and Linguistic Theory 2001, Josep Quer, Jan Schroten, Mauro Scorretti, Petra Sleeman, Els Verheugd. (Eds.), p. 75-92, December 2003.
11. GÖBBEL E., *Focus in double object constructions*. Linguistics: an interdisciplinary journal of the language sciences, **Vol. 43**, Issue 2, p. 237-274, March, 2005.
12. GROSZ B. J., JOSHI A. K., WEINSTEN S., *Centering: a Framework for Modelling the Local Coherence of Discourse,* Computational Linguistics, 21 (2), 1995.
13. GUSSENHOVEN C., *Focus, Mode and the Nucleus*. In: On the Grammar and Semantics of Sentence Accents. Dordrecht: Foris. p. 11-62, 1984.
14. GUSSENHOVEN C., *A semantic analysis of the Nuclear Tones of English*. In: On the Grammar and Semantics of Sentence Accents. Dordrecht: Foris. p. 193-265, 1984.
15. HAJICOVA E., *Old Linguists Never Die, They Only Get Obligatorily Delete*, In: Computational Linguistic, **Vol. 43**, No. 4, p. 457-469, November 2006.
16. HAJICOVA E., SKOUMALOVA H., SGALL P., *An Automatic Procedure for Topic-Focus Identification*, Computational Linguistics, 21 (1), p. 81-94, 1995.
17. HAJICOVA E., PARTEE B.-H., SGALL P., *Topic-focus Articulation, Tripartite Structures, and Semantic Content*. Dordrecht: Kluwer Publishers, 1998.
18. JAKOBSON R., *Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves*, TCLP 2, Prague, 1929.
19. JÄRVINEN T., TAPANAINEN P., *A dependency parser for English,* Technical Report TR-1. Helsinki: University of Helsinki, Department of General Linguistics, 1997.
20. JÄRVINEN T., TAPANAINEN P., *Dependency Concordances*. International Journal of Lexicography, 11(3): p. 187–203, September 1998.
21. KAMP H., REYLE U., *From Discourse to Logic: Introduction to the Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. **Vol. 1 & 2**, Kluwer Academic Publisher, Dordrecht, The Netherlands, 1993.
22. MANN W., THOMPSON S. *Rhetorical Structure Theory: A Theory of Text Organization*. Research Report RS-87-190, Information Sciences Institute, Univ. of Southern California, Marina del Rey, California, 80 p., 1988.
23. MARCU D., *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*, Ph.D. Thesis, Univ. of Toronto, Canada, pp. 331, 1997.
24. MARCU D., *The Theory and Practice of Discourse Parsing and Summarization.* The MIT Press, 2000.
25. MORUZ M.-A., *Adnotator FDG pentru limba română* , Bachelor Paper, Faculty of Computer Science, Iaşi, 2006.
26. SCHMERLING S., *Aspects of English Stress.* Austin: University of Texas Press, 1976.
27. SELKIRK E., *Phonology and Syntax*, MIT Press: London/Cambridge, 1984.
28. SELKIRK E., *Sentence Prosody: Intonation, Stress, and Phrasing*. In John A. Goldsmith, ed., The Handbook of Phonological Theory, p. 550-569. London: Blackwell, 1995.
29. SELKIRK E., KRATZER A., *Focuses, Phases and Phrase Stress*. Ms., University of Massachusetts at Amherst, 2005.
30. SGALL P., HAJICOVA E., PANEVOVA J., *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*, eds. J. L. Mey, Dordrecht:Reidel - Prague: Academia, 1986.
31. STEEDMAN M., *Surface Structure and Interpretation*. Cambridge, MA: MIT Press. 1996.
32. STEEDMAN M., *Information structure and the Syntax-Phonology Interface.* In Linguistic Inquiry, Eds. Massachusetts Institute of Technology, **Vol. 34**, MIT Press, Cambridge, p. 649-689, 2000
33. STEEDMAN M., *Information-Structural Semantics for English Intonation*., In LSA Summer Institute Workshop on Topic and Focus, Santa Barbara, 2001.
34. TEODORESCU H.-N., *A Proposed Theory in Prosody Generation and Perception: The Multi-dimensional Contextual Integration Principle of Prosody*. In Trends in Speech Technology, Romanian Academy Publishing House, p. 109-118, 2005.
35. TESNIÈRE L., *Elements de Syntaxe Structurale, Klincksiek*, Paris, 1959.
36. TSENG C.-Y, PIN S.-H., LEE Y., WANG H.-M., CHEN Y.-C., *Fluent speech prosody: Framework and modeling*. Journal Speech Communication, **Vol. 46**, (Eds. Elsevier), p. 284-309, 2005.
37. TSENG C.-Y., SAN Z.-Y., CHANG C.-H., TAI C.-H., *Prosodic Fillers and Discourse Markers-Discourse Prosody and Text Prediction*. *The Second International Symposium on Tonal Aspects of Languages (TAL 2006*)*,* La Rochelle, France, p.109-114, April 2006.
38. Van VALIN Robert D., Jr., *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press, 2005.