

Recognition and Prediction for Implicit Contrastive Focus in Romanian

Neculai Curteanu, Alex Moruz,
Cecilia Bolea, Andrei Scutelnicu

Abstract

This paper is made up of two parts: **the first part** continues the theoretical investigations on Information Structure (IS), linguistic, and computational approaches suitable to provide solutions to the prosody prediction problem of Implicit Contrastive Focus (ICF) concept, introduced in our previous papers. ICF is meant to be the particular case but also the counterpart of the classical category of contrastive Focus at the finite clause level, as the second item in the Background-Focus pair of IS dimension. The classical contrastive Focus, which we called Explicit Contrastive Focus (ECF), is the intonationally F marked entity introduced by overt lexical contrastive markers. ICF labels the situations where contrastive intonational focusation occurs but without the lexical presence of the contrastive Focus markers! The only and main device to introduce the contrastive focusation on certain constituents is the syntactic dislocation from their specific positions in the Systemic Ordering (SO) of syntactic-semantic roles for the Romanian finite clause. The ICF problem means to obtain reliable algorithms and procedures on the Discourse-Prosody interface in order to accurately predict the contrastive Focus distribution within the Romanian ICF-type affirmative finite clause. **The second, applicative part** of the paper describes algorithms for solving the ICF problem for Romanian, trying to exploit the typically dislocated constituents in the finite clause and to predict their Prosodic Prominence (PP). Procedures for the development of intonational-prosodic patterns assigned to the ICF distribution by certain ICF estimation schemes are developed and tested for a balanced set of Romanian ICF-type affirmative finite clauses.

Keywords: prosody prediction for the Romanian clause, contrastive focus hierarchies; Communicative Dynamism degrees; Systemic Ordering, Implicit Contrastive Focus in Romanian; Implicit Contrastive Focus estimation scheme; Discourse-Prosody interface.

1 Theoretical Framework for the Prosody Prediction of the Romanian Contrastive / Non-Contrastive Focus

Aiming to present this paper as consistent and self-contained as possible, we outline in this section the theoretical framework and some results related to Romanian prosody prediction and intonational (particularly, contrastive) Focus modelling for Romanian [7], [8], [9], [10]. We also introduce three main dimensions of the Information Structure (IS) theories, in their updated terminology, referring the valuable results on (either classical or recent) IS modelling of the discourse-prosody interface. Information Structure (IS) and Intonation are seen as two autonomous and independent components of the grammar, closely related to each other: intonational phrasing and patterns express informational structuring, while a great part of IS is linguistically conveyed by prosody. The interaction between IS and intonation is studied on the following grammar interfaces: "intonation and phonology, Focus and phonological phrasing, intonation (focus) and syntax, IS and discourse analysis...". [23].

The communicative model of *Common Ground* (CG) is essential for the semantic theories on Information Structure and is defined as the amount of shared knowledge between the speaker and hearer, at a certain moment (e.g. [29]). The following *three dimensions* are considered to be important in the current IS theory and terminology [10]: The **first IS dimension**, *Givenness* or *Givenness-Newness*, refers to the feature of an entity that is present in the CG (Common Ground) database, at a certain time, within the discourse of the two interlocutors (speaker-hearer). We notice that not all the elements in the complementary set

of Discourse-Given (+*Givenness*) entities are Discourse-New (bearing the *Newness* feature) ones: some of them are contrastively focused in intonation. Actually, there are two types of *Newness* categories for the CG communication model: (a) an *informational Newness* (introducing in intonation the *informational Focus*), showing that the entity is a Discourse-New entity, not existing yet in the CG and being added currently to the CG, once introduced in the discourse. (b) A *contrastive Newness* (associated in intonation to *contrastive Focus*), refers to the contrastive quality of the entity knowledge in the CG model. Thus, two types of intonational Focus correspond in the IS theory to the two notions of Newness: *informational* (Discourse-New) *Focus* and *contrastive Focus*. The informational Focus is also denoted as Non-contrastive Focus (abbreviated as NF in [6]) to emphasize its distinctiveness to the Contrastive Focus (CF in [6]). For contrastive Focus, the corresponding entity in CG may be either Discourse-New or Discourse-Given, intonationally stressing, in any case, the existence of a semantic set of alternatives in the CG. We also remark that, for the free word order languages, the constituents bearing the Discourse-Given feature are usually preceding the Discourse-New ones in the Systemic Ordering (SO) order [40], [20] of the syntactic constituents / semantic roles of a finite clause. (For more details, see Section 2).

The **second IS dimension** consists of the pair *Background-Focus* (also referred as *Topic-Focus* in the Prague School's IS theories [20], [21]). The *Background-Focus* pair marks the difference between the entities belonging to CG (being Discourse-Given, i.e. G marked in the Selkirk's approach [39]), which form the *Background*, as opposed to the *Focus* marked entities. Focus constituents consist in Discourse-New (*informational Focus*) or rather intonationally contrasted (*contrastive Focus*), representing the entity which is informationally or contrastively emphasized, supported also through grammatical means besides intonation.

The following subtle relationship between the first two IS dimensions is worth to be highlighted: *Background* may have the feature *Givenness* (Discourse-Given) but not *Newness*, while *Focus* has usually the feature *Newness* (Discourse-New) but, when contrastive, may also

be *Givenness* (Discourse-Given). In a clause, the constituents being not Focus should bear the *Givenness* feature. The following examples illustrate the first two IS dimensions (see more in [10]).

1.a. **Adina_F** i-a fost prezentată lui Alex.

En: Adina was presented to Alex.

1.b. Și toată viața, el a iubit-o numai **pe Adina_{F-Contrast, Given}**.

En: And all his life, he loved only Adina.

1.c. Și toată viața, el a iubit-o **pe Ana_{F-New}**.

En: And all his life, he loved Ana.

In Ex.1.a, *Adina* is a constituent marked with Focus and opposed to Background entities. In Ex.1.b, *Adina* is a typical example of a constituent bearing the *Givenness* feature in CG, yet acting as **contrastive Focus**, being marked with the typical contrastive adverb *numai* (En: *only*). The Ex.1.c contains the constituent *Ana* which acts as **informational Focus**, since it is a Discourse-New entity. In [7] we applied the Prague School's TFA (Topic-Focus Articulation) algorithm to Romanian and extended it to inter-clause level [8], thus dealt with the Background-Focus and *Givenness*-Newness IS dimensions trying to use the results for Romanian prosody prediction.

This modal behavior of the Background-Focus IS dimension, compared to the *Givenness* dimension meaning, is important for the hierarchical evaluation utilized in [10] and in the present paper for the special category of *Implicit Contrastive Focus* (ICF), derived from standard *Explicit Contrastive Focus* (ECF) entity, which is marked by intonationally contrastive lexical markers.

The **third important dimension** of the IS theory relying on the CG communicative model is concerned with *Topic-Comment* (referred also as classical *Theme-Rheme*) clause-level structures. *Topic* (= *Theme*) should be understood as the semantic subject introduced in the communication act (predication-level), while *Comment* = *Rheme* refers to what is said or informationally emphasized about the *Topic* of the clause. *Topic* is also interpreted in the CG model as the *pointer address* where the *Comment* is stored [29]. Although Discourse-Given information is often linked to the *Topic* role and the prosodic promi-

nent category of a clause *Focus* is naturally to be found within the *Comment*, this is not always the situation because the *Topic* and the IS of a clause are generally distinct.

- 2.a.** [The lecturer]^{Topic} delivered [his speech today_F]^{Comment}.
2.b. [The lecturer_F]^{Comment} delivered [his speech today]^{Topic}.

In example 2, A. Leong [32] shows correctly that *The lecturer* is either Topic or Comment according to its Prosodic Prominence (PP), in each case (see also [16]). Alternative changing of *The lecturer* in either Topic or Comment, intonationally focused or not, is the exclusive result of distinctive discursive contexts, since the other two factors (from the four ones, specified by J. Firbas [6], which can modify the functional perspective of the clause) in ex. 2 remains untouched, namely the linear ordering of words and their semantics. It is obvious that (2.a) and (2.b) must evolve from different discursive contexts since they have distinct intonations on their Topic-Comment categories. In [10] we investigated the Topic-Comment algorithms developed in [31], inspired by J. Firbas [16] and extending the classical Hallidayan approach [22] to Theme-Rheme, for integrating them within an intonational-prosodic discourse phrasing and prediction on the Romanian finite clause.

2 Systemic Ordering and Communicative Dynamism for Prosody Prediction

The *Systemic Ordering* (SO) refers to a pre-established linear order of the clause constituents (syntactic-semantic roles) in a finite clause. SO is the statistical result of the most frequent linear ordering of the semantic roles in a finite clause, for all the predicates in the active diathesis of a certain language. For a specific predicate (or predication) p , the SO for p is denoted with SO_p . Thus SO can be understood as the corresponding p -free, statistical ordering, of all the SO_p -orders. SO and SO_p notions are language-dependent, possibly with strong similarities for sibling natural languages. For the Romanian language there are not reliable results on the SO and SO_p orderings, supported by computational linguistic consistent studies.

When referring to SO_p (statistically p -depending) or SO (statistically p free) order of the semantic roles in a specific clause, within a certain language, we can have the SO_p -disorder, respectively SO -disorder that may occur on a particular clause. The well-known Prague School in Computational Linguistics (CL) defined and worked with the concepts of SO [40], [21], [20], and Firbas' Communicative Dynamism (CD) [15], [16], as computational instruments for dealing with the classical TFA (Topic- Focus Articulation) and Theme-Rheme IS algorithms. For instance, the SO for Czech and English is considered to be the following (statistical) sequence of semantic roles [40], [20]:

$$Time \Rightarrow Actant \Rightarrow Predicate \Rightarrow Patient \Rightarrow Object \Rightarrow Origin \\ \Rightarrow Effect \Rightarrow etc.$$

Closely related to the SO -order (or SO -disorder) of the semantic roles in a finite clause, Jan Firbas introduced the notion of *Communicative Dynamism* (CD) associated (eventually) to a predicate p that is heading the distribution of the clause constituents [15], [16]. CD is understood as "a quality displayed by communication in its development (unfolding) of the information to be conveyed and consisting in advancing this development." [15], p. 30. Actually, each constituent of a clause carries a certain CD degree, defined by J. Firbas as "the extent to which the sentence element contributes to the further development of the communication." Furthermore, the "context dependent elements carry a lower degree of CD than context independent elements." [15], p. 31.

Background-Focus as TFA entities computing [20], [7], [9], and Topic-Comment algorithms can be seen as a component of the evaluation process for the Implicit Contrastive Focus (ICF), which represents one of the basic solutions to intonational focus assignment and prosody prediction for the Romanian clause. Section 5 contains an effective proposal for CD computing algorithm used to the recognition and prediction of ICF for the Romanian clause, derived from a more general algorithm for the hierarchical intonational Focus and Break prediction on the Romanian discourse-prosody interface in [10].

There exists also a close relationship between Topic-Comment algorithms, computation of CD degrees, and Prosodic Prominence (PP) of intonationally (either contrastive or non-contrastive) focused constituents in the clause. The main elements of this relationship can be outlined as follows [10]: (a) Topic Comment (Theme-Rheme) computing is important either by itself or for IS textual processing. (b) As significant approaches for Topic Comment (Theme Rheme) computing can be mentioned, e.g., the Prague School [15], [40], [21], or the A. Leong's Inference Boundary algorithm [31], [8]. (c) The effective utility of Topic-Comment IS dimension for prosody prediction can be only partial, as J. Firbas specify the four main factors contributing to Topic-Comment computing: CD (Communicative Dynamism) vs. SO (Systemic Ordering) of the clause semantic roles, their semantics, the discursive context, and intonation. Naturally, only the first three textual factors can be utilized on the discourse-prosody interface, thus prosody prediction. (d) When performing a similar analysis for the TFA algorithms in the IS framework, one may notice that TFA, with its contextually-bound and non-bound entities, is a decent competitor for clause-level Discourse-Given (G marked) vs. Discourse-New (informational Focus) category tagging [20], [7], [8].

In an updated terminology, context-dependent entities are subsumed by the Discourse-Given ones (also labeled as G-marked constituents), while the context-independent entities are retrieved as either Discourse-New or Contrastive Focus (F-marked constituents) [39]. Prague School and its Topic-Focus Articulation (TFA) algorithms work with contextually-bound and contextually-nonbound constituents in a clause, for trying to acquire the intonationally focused (vs. deaccented) entities [20], [21]. The relationship between contextually bound vs. G marked and F marked constituents is underlined clearly in [10]: "TFA and Topic-Comment [31] algorithms are just partial steps toward clause-level foci computing, in a discursive setting: the contextually-bound entities are a superset of the Discourse-Given ones, while the contextually non-bound constituents are possible Discourse-New foci." The contrastive but still Discourse-Given ones (both F marked and G marked in terms of Selkirk [39]) seem to be ignored by the TFA focus

assignment procedure. Despite this weakness, TFA offers real advantages of a good linguistic modelling and efficient computing procedure, compared to the expensive anaphora resolution one, for obtaining the Discourse-Given constituents.

3 Explicit and Implicit Contrastive Foci. Facts and Arguments

Explicit Contrastive Focus (ECF) and *Implicit Contrastive Focus* (ICF) phrases should be used with the following meanings [10]: ECF describes those categories of contrastive Focus introduced by specific lexical markers, while ICF designates the situations where contrastive intonational focusation is covered by dislocation / disorder of the semantic roles within the finite clause, but without the lexical presence of the contrastive Focus markers. The only device to introduce the contrastive focusation on certain constituents is the syntactic dislocation from their standard position in the Systemic Ordering (SO) of syntactic-semantic roles for the Romanian finite clause. The ICF problem consists in obtaining reliable algorithms and procedures on the Discourse-Prosody interface in order to predict realistically the contrastive Focus distribution in the Romanian ICF-type affirmative finite clause.

We denoted ICF (Implicit Contrastive Focus) in a clause since it should look very similar to ECF, but the overt, lexical focusation markers for contrastive Focus are missing. The same concept but not under this label, ICF is particularly investigated for pre- and post-verbal situations in [34], being of major interest for at least two reasons: (i) The close relationship with Topic-Comment (Theme-Rheme) algorithms; (ii) The behavior forms for Romanian, including covert markers, textual discursive contexts, syntactic dislocations / disorders, discursive-intonational patterns.

Lexical, overt markers of contrastive intonation introduce the category of Explicit Contrastive Focus (ECF), denoted also as primary contrastive Focus or, hereafter, *First Occurrence Focus* (FOF). FOF

is specified by its broad Focus (projection) domain, and its pattern of contrastive marker(s). For a second (or even third) contrastive marker, the existence of the second (or third) focusing entity is possible, called Second (or Third) Occurrence Focus (abbreviated SOF, respectively TOF) [14], with distinct focusing domains, embedded or not into each other and, most important, in a hierarchy of their Prosodic Prominence (PP), constrained by the intonational hierarchy pattern of the corresponding contrastive focusation markers. This problem means to assign an intonational hierarchy pattern to the contrastive Focus entities, when two or more contrastive phrase markers generate several contrastive Foci, with broad contrastive foci domains embedded into each other (or possibly disjointed). For instance, in [10]:

ECF Example:

3. Chiar [soțul ei]_{FOF} doar [o singură dată]_{SOF} a văzut tabloul]_F.

En.:* Even [her husband]_{FOF} only [once]_{SOF} saw the picture].

En.: Even her husband saw the picture only once.

The commencing * (asterisk) denote the word-by-word translations, intended to provide the reader with a hint about the Romanian word order.

The contrastive markers *chiar* (*even*) and *doar* (*only*) introduce their own contrastive Focus categories and domains (broad Focus), [soțul ei]_{FOF} (her husband) being the domain and FOF, while [o singură dată]_{SOF} a văzut tabloul]_{BroadF} is the subordinated domain with its SOF [o singură dată]_{SOF}. A comprehensive analysis is formulated in [39].

We consider that the presence of explicit contrastive markers that triggers ECF is a problem for which one can establish detailed and reliable prosodic patterns, hierarchically organized on classes of intonationally contrastive markers. Derived from ECF but also seen as its counterpart, we identified the ICF evaluation problem as the main difficulty issue to be solved inside the Romanian clause, whose complementary solution, added to the ECF computation of the hierarchical pattern of contrastive Foci at discourse and clause level, can offer a sound procedure to the intonational focus assignment for the prosody

prediction problem of the Romanian clause.

3.1 Prosodic Prominence of Contrastive vs. Non-Contrastive Foci

The problem of the intonational-phonetic relationship between Explicit Contrastive Focus (ECF), associated with focusation markers, and informational Focus, which corresponds to a Discourse-New entity, was solved by H. Truckenbrodt (1995) and Mats Rooth (1996). They attested the *Contrastive Focus Prominence Rule* (CFPR) as follows [39]:

CFPR: Within the scope of a focus interpretation marker, the corresponding contrastive focus constituent is the most prominent metrically.

The CFPR principle confirms, with remarkable robustness on phonetic data, that the Explicit (i.e. lexically-overt marker defined) Contrastive Focus (ECF) is more prominent against any other kind of intonational focus, in any context, be it either informational (Discourse-New) Focus or contrastive SOF (Second Occurrence Focus). The same result is recently reinforced for Catalan Spanish in [6]. For instance [10]:

ECF Examples:

4.a. Chiar_i [Maria_i]_{FOF} i-a anunțat imediat_j [lui Ion_j]_{SOF} accidentul.

En: *Maria herself announced immediately to Ion the accident.

4.b. Chiar_i [Maria_i]_{FOF} i-a anunțat lui Ion imediat_j [accidentul_j]_{SOF}

En.:*Maria herself announced to Ion immediately the accident.

For the classical question of the relationship between contrastive and (non-contrastive but) informational foci, the following solution is known to hold in speech analysis: the informational Focus is represented by a pitch accent, with a peak of intonational prominence, while the contrastive Focus corresponds to a constituent whose broad Focus is syntactically more complex, represented by an intonational contour with one or more pitch accents. These facts are parameterized

for a large class of natural languages for which the informational Focus is determined by pitch accent, e.g. Italian, Spanish, Portuguese, Romanian, English, German, etc. [18]. A special fact is that the distinction between informational and contrastive Focus does not hold for French, where the intonation is displayed (not only) by pitch accents but rather by boundary tones and tonal phrasing. This makes CFPR to be language-dependent, nevertheless. In subsection 4.3 we describe more recent arguments for the distinctiveness measure between Contrastive Focus (CF) and Non-contrastive Focus (NF) in Spanish [6].

Experimental analyses established that, although SOF does not typically bear an intonational pitch accent, there is a clear phonetic prominence by duration and intensity of the signal. The prominence accent may be different depending on whether the contrastive Focus marker is located at the beginning or the end of the clause. Example:

ECF Example:

5. [Maria_i]_F l-a anunțat pe Ion mult mai târziu_j , chiar ea_i.

En: *Maria announced Ion much later, herself.

E. Selkirk [39] performed the comparison between the prosody of SOF and Discourse-Given (G marked) constituents, which are not contrastive (thus F marked). Example (6.b) contains contrastive Foci, with non-embedded domains. A controversial issue in this case is that FOF does not seem to be genuinely contrastive, thus making the second Focus an authentic SOF. This fact raises again the necessity of establishing in advance hierarchies not only on focusation markers of the local modifier categories (such Adverbs, for instance), but also on the contrastive Focus markers of the clause-level semantics roles, either lexically marked or dislocated within the clause through specific ICF patterns. Intonational and prosodic hierarchies of focusation markers, on all the syntactic and prosodic textual structures, completed by compatibility and clash-solving procedures are the solution for the prosody prediction problem which we currently advocate.

ECF-ICF Examples:

6.a. Ni s-a spus să avem numai gânduri bune_F.

En: We were told to have only good thoughts.

- 6.b.** Dar chiar ne-am plictisit să avem numai gânduri bune_{F, G}.
 En: But we really got tired of having only good thoughts.
- 6.c.** Dar ne-am plictisit să avem gânduri bune_G.
 En: But we got tired of having good thoughts.

Combining CFPR and G-Marking Condition [39] entails important consequences [10]: the algorithms computing the informational Focus (Discourse-New, e.g. Prague School’s TFA [20], [7], [9]), the evaluation rules for non-contrastive, clause-level intonation NSR [39], SAAR [19] and SBAR [7], [8], [9], and the procedures for contrastive Focus calculus (FOF, SOF, narrow and broad Foci, Focus patterns for clause-level syntactic categories) can be applied in a hierarchic, independent, and separate manner.

3.2 Contrastive Focus Examples and Prosodic Patterns

For the prosody prediction of Romanian, we proposed to use in [10] a hierarchic system at prosodic discourse level for the Contrastive Focus (CF), Non-contrastive Focus (NF), and Discourse-Given (G marked) categories. [10] presents in detail a table with the intonational inequalities (of statistical nature) for these Focus categories oriented to the Romanian language. Since the system of statistical inequalities proposed in [10] is strongly language-dependent, the contained inequalities (actually, hierarchies) should be equated with specific weights established a priori by fundamental speech-oriented studies on the Romanian language. In [10] we developed a general procedure for hierarchical intonational Focus and break assignment on the discourse-prosody interface, whose derived algorithm is displayed and applied to ICF prediction in Section 5.

The Contrastive Focus Prominence Rule (CFPR) confirms that ECF (or, simply CF, as denoted in [6]) is more prominent against any other kind of intonational focus, in any context, be it either informational Focus or SOF [13] (denoted Non-contrastive Focus and abbreviated as NF in [6]).

However, the recent study [6] brings well-balanced conclusions (subsection 4.3) on the relationship between Contrastive Focus (CF) and

Non contrastive Focus (NF) entities (the latter comprising the informational Focus as Discourse-New ones), both concerning their relative intonational prominence as well as the influence of the syntax on the nature of CF and NF constituents. The investigated language is the Catalan Spanish and the results may be distinctive depending on the clause constituent (Verb, Predicate, Object, Subject etc.) and dimension (broad vs. narrow) of the focused structure.

We also notice that the G-marked (Discourse-Given) non-focused entities, resulted commonly from anaphora resolution, could be evaluated with the computationally less expensive but also less accurate contextually-bound constraints from the TFA classical algorithm [20], [8], [9].

Several examples of implicit contrastive Focus (suggesting, for some cases, the corresponding explicit variant) are presented. The * before an English sentence means that the sentence, even when grammatically correct, is only marginally used in this form, as in [10]:

ICF-ECF Example:

7. Doar Cerceii_{iFOF} Maria chiar ieri_{iSOF} și i-a cumpărat.

En: *Only the earrings_{iFOF} Maria even yesterday_{iSOF} bought for herself.

The curly brackets contain the contrastive Focus marker that is lexically missing. This example illustrates once more the use of labels for both the ECF and ICF concepts. In the next example, the clitic is doubling the direct complement if it is focused contrastively. Similar examples of prosodic contrastivity on double complements are analyzed in [34]:

ICF-ECF Examples:

8.a. Supa de legume, [bluza_i bucătăresei_F] a murdărit-o_i , nu fusta.

En: *The vegetable soup, the cook's blouse has stained , not her skirt.

8.b. [Bluza_i bucătăresei_{iFOF}], [cu supa de legume_{iSOF}] și-a murdărit.

En: The cook's blouse was stained with the vegetable soup.

The labels iFOF and iSOF in the above example abbreviate the expressions of implicit First Occurrence Focus and implicit Second Occurrence Focus for these types of Contrastive Foci [10].

ICF Example:

9.a. Maria și-a cumpărat cerceii ieri_F .

En: Maria bought her earrings yesterday.

9.b. Și-a cumpărat ieri_F , cerceii, Maria.

En: *Bought yesterday the earrings Maria.

Once again, we remark the Contrastive Focus (CF) Rule working on both Italian and Romanian: the CF constituent in the clause left-periphery is dislocated as the immediately pre-verbal one, providing an important sample of CF, actually ICF, prosodic pattern at clause level. This is a sample rule representing the principle(s) according to which the CD degrees of the semantic roles are to be computed within the finite clause. Other pattern examples of Romanian ICF variants derived from Italian are displayed in [37], e.g.:

ICF Examples:

10.a. Le_i-o_j prezintă pe soția_j mea, părinților_i lui_k, Ion_k .

En: He presents my wife to Ion's parents.

10.b. Le_i-o_j prezintă părinților_i lui_k, pe soția_j mea.

En: He presents my wife to his parents.

10.c. Le_i-o_j prezintă Ion_k, părinților_i lui_k, pe soția_j mea.

En: * Presents Ion to his parents my wife.

4 Contrastive Focus Distribution within Finite Clause

A natural question for the ICF problem is the following: for two finite, dependent clauses, each of them containing SO disorder, thus constituent dislocations, is it possible that a clause to contain each other dislocated constituents? This problem is discussed in [28] and the next subsection, the answer to this problem being (almost always) negative.

4.1 Constituent Dislocations in Adjacent Finite Clauses

We remark from the beginning the importance of the inter-clause phrase markers, thus of those markers that play an essential role in the task of finite clause segmentation and parsing (which means segmentation plus dependency establishing, on several levels of inter-clause relationship nature). Let us consider some instances of several dependent finite clauses. As one can see, the dislocations, i.e. movements of syntactic-semantic constituents from the Systemic Ordering (SO) of the clause at hand, cannot be performed from one clause to another but only within the same clause.

In fact, the same rule of consistency applies to complex nominal groups (NGs), dominated by predication semantic heads, as well as to the Verbal Groups (VGs) whose semantic heads are predication verbs standing in non-finite tense (gerund, past participle, or supine). Thus the general rule is that dislocations of a finite clause constituent(s) into another dependent, adjacent finite clause are forbidden. An equivalent statement of this rule is the following: dislocations of the syntactic-semantic roles of a predication in the scope of another predication are forbidden. We found some peculiar exceptions from this rule, but their syntactic and / or semantic form is situated at the limit of linguistic correctness [28], p.9.

11.a. [C₁(Cât despre) Ion, am plecat] [C₂înainte să-l examineze Popescu].

11.b. [C₁(Cât despre Ion, n-am întâlnit fata] [C₂-RELcare l-a văzut ultima dată].

11.c. *[C₁Pe Ion am plecat] [C₂înainte să-l examineze Popescu (pe Ion)].

11.d. *[C₁Pe Ion n-am întâlnit fata] [C₂-RELcare l-a văzut anul trecut (pe Ion)].

Syntactic-semantic role dislocation constructions, for Contrastive Focus (thus specific to ICF situations) or Non-contrastive Focus (for the occurrence of Discourse-New, informational Focus) within finite, dependent clauses, or non-finite dependent predications are illustrated as follows:

- 12.a.** [Am spus] [că [pe Victor] nimeni nu l-a văzut].
- 12.b.** [Aș dori] [[pe Ion] să-l chemați mâine].
- 12.c.** [Mă întreb] [,] [pe Petre_i cine-l_i mai crede].
- 12.d.** [Îl_i întreb pe Petrei] [cine-l_i mai crede].
- 12.e.** [Pe Petru_{F1} Maria nu l-ar ajuta,] [pe Gheorghe_{F2}, da].
- 12.f.** [Eu pe Popescu_F l-am văzut (nu pe Ionescu)].
- 12.g.** [Eu romanul ăsta_F l-am citit (nu altul)].

Some remarks are necessary concerning the use of clause level markers, particularly the complementizers introducing subordinate clauses and NGs (typically, the relative pronoun “care” (that, which) and its derivate phrases), and their role to introduce inter-clausal prosodic patterns. In [25], [26] there are analyzed the NG and clausal markers “dacă” (if, whether) and “oare” (somehow, rather, etc.), both in their individual but also correlate functionality. Other important complementizers such as “că”, “de”, “ca”, as well as the markers “să” and “a” for subjunctive and non-finite VG are examined. Studies like [25] provide really useful information for system of statistical inequalities and prosodic patterns to be applied as solutions for the ECF and especially ICF problems. The next subsection continues and deepens the investigation.

For making consistent predictions on the clause-level Contrastive Focus, mostly located in the left-periphery (pre-verbal side) but also right-periphery of the clause, the algorithm of solving the ECF, respectively ICF, problem attempts to find specific prosodic patterns for each of the two problems, directly useful both in recognition and prediction of the Focus categories.

As in [10], the general solution to the ICF problem should be based on reliable statistical inequalities and discursive contexts in which they are applied, including this case where both complementizers, as clause-level phrase markers, and the contrastive focusation markers, need accurate and appropriate statistical prosodic hierarchies. The statistical inequalities describing the intonational-prosodic behaviour of the Focus markers should be designed and tested for both affirmative / declarative clauses and the interrogative / indirect interrogative clauses, each one with its specific discursive contexts.

The analysis in [25] provides some explanations on the left periphery (pre verbal) organization of the finite / nonfinite clause in Romanian. The prosodic type of the clause is given by the lexical complementizer markers that introduce the clause, for the case of indirect interrogative, relative, and Wh interrogative clauses. The analysis of the markers “dacă” and “oare” proposes the location determining and the interrogative elements playing the syntactic boundary for the Background and Focus categories in the left-periphery of the clause.

We mention that the theoretical linguistic analysis provided in [25] relies on the X bar projection of the functional categories represented by inter-clause markers (such as complementizers, for instance) and intra-clause markers (for NGs, VGs, AGs). This approach is easily equated within our theory of FX bar functional projection developed in [11], this fact providing a particular advantage in computational and linguistic theoretical terms.

4.2 SO-Dislocated Categories and Constraints on Their Clitics

The relative pronouns or complementizers and the lexical material around them introduce, in general, the prosodic type of the clause and the syntactic-semantic dependency in relation to the other clauses, bearing the SO dislocations (i.e. dislocations from the Systemic Ordering, SO) of contrastive Focus entities, while the Background constituents may occur between the semantic heads of the clause. There are strong suppositions to believe in the presence of an intonational hierarchy of the adverbs for Romanian, similar to Italian, in which the movement of an adverb to the clause periphery is realized either to the left-periphery, or to a (contrastive) Focus position.

The role of (generalized) quantifiers as clause-level markers for finite clause delimitation should be taken into account, as well as certain aspects that are specific to the Romanian syntax, as it is the possible, lexical missing of the grammatical Subject of a finite clause, as opposed to the same situation in English. Dislocation of such a constituent can be just only traced.

The prosodic analysis of the intonationally focused categories within the clauses introduced by the special marker *oare* (usually classified as adverb but functioning actually as a complementizer marker), in interrogative or affirmative contexts, gives rise to the following interesting observation: in addition to the lexical-semantics, grammatical, discursive et al. criteria, important prosodic consequences may have the etymological classification of certain phrase markers. From the etymology, important linguistic affiliations can be inferred concerning the semantics of some textual and / or intonational-prosodic markers, hence the distribution of discursive and prosodic structures they involve. For instance, the etymology of *oare* can be used as argument for defining this marker as functional head. In [35], the adverb / marker *oare*, with its variant *vare*, has its etymological origin in the Latin verb *volet* (*vrea* En: want, *dorește* En: wish, desire). The verb *volet* is itself a category operating as predication head, which changes the syntactic role of *oare* from a classical adverb rather into a predication conjunction, requiring similar arguments to those derived from its etymological source. Another issue when etymological classification may be important for the analyzed phrase markers is the sequence recognition and meaning establishing for sequences like *dacă oare*, *oare dacă*, or *oare* *[+]* *dacă*, where *[+]* is the symbol for overt, explicit lexical elements.

In many works, including [35], [28], the distribution of the Focus categories ensued from dislocations, in the same context as that of clitic pronouns, which may double or not the focused category, within a single finite case (or predication) is investigated. Dislocations are discussed for the two important situations of the left and right periphery within a clause. In this framework, the clitic doubling special relationship is analyzed for a given syntactic constituent, being achieved necessary and / sufficient conditions for the occurrence of this phenomenon.

[28] considers Romanian, Bulgarian and Greek languages to follow two types of distribution and inter-dependence for the Background-Focus categories that are SO-dislocated at the clause-level: the SO-dislocated constituents (denoted CAT-Disloc) and the clitic pronouns (denoted Clitic[Db]l-Disloc) with the role of Background-Focus categories can be dislocated into the left (predominantly) or right periph-

ery.

The following remarks regarding the distribution of the dislocated categories CAT-Disloc, deployed within finite clauses of ICF or ECF kind, distribution of the clitic pronouns, doubled or not, and focused categories in the ICF clause type are useful for both Romanian but also Bulgarian and Greek [28]: (a) In Romanian, the Background-Focus categories can be doubled by clitics, under certain conditions. (b) The Romanian language is more permissive than the Greek and Bulgarian regarding the phenomenon of clitic doubling. (c) The following linear ordering for the CAT-Disloc, Clitic[Db]l-Disloc, and the Focus categories in Romanian, Bulgarian, and Greek Balkan languages is proposed in [28]:

(13) CAT-Disloc > Clitic[Db]l-Disloc > Focus.

This sequential precedence of the mentioned categories is questionable and disputable in the context of more general and recent analyses in the literature. For instance, the first sequence does not always hold (counter-example, 14.c), while the proposed precedence between CAT-Disloc and Focus can be broken in the situation of a discursive context where contrastive Focus occurs on the first position (as in Ex. 2.b) within the clause, either it is Background (Discourse-Given) or informational Focus (Discourse-New).

ICF Examples:

14.a. [_TMioareii] [_Tinelul] la nuntă i l-a dat Anghel.

14.b. [_TInelul] [_TMioarei_i] la nuntă i l-a dat Anghel.

14.c. Anghel i-a dat [_TMioarei_i] [_Tinelul] la nuntă.

14.d. *Anghel i l-a dat [_TMioarei_i] [inelul_i] la nuntă.

14.e. Anghel i l-a dat [_TMioareii] [inelul_i] la nuntă.

In the clitic doubling syntax, the Clitic[Db]l-Disloc may appear before the CAT-Disloc category in the ICF clause-type, provided that certain semantic features should hold, e.g. human or associated being condition for the CAT-Disloc category. Examples 14.d and 14.e support these constraints on the clitic-doubling event in Romanian.

Syntactic behavior of reflexive pronouns and expressions containing a reflexive pronoun shows that anaphoric-type expressions are bound

(in the sense of logical binding operation) of their antecedents. In all Balkan languages [28], reflexive expressions cannot be standard Background (G marked) SO-dislocated categories positioned (in SO-disorder) in the left-periphery of the clause, while the corresponding clitics Clitic[Db]l-Disloc can be located in the proper grammar left-periphery. The ex. 15 that follows brings again doubts on the general correctness of linear precedence proposed in (13).

ECF Example:

15.a. *Cât despre sine_i, Victor_i nu s_i-ar pune în pericol.

15.b. Pe sine_i, Victor_i nu s_i-ar pune în pericol [t_i].

In the grammar-accepted examples, the anaphora must be logically bound with its antecedent, which shares the same index with, further specified in the category trace, i.e. the syntactically corresponding empty category, with the same index.

4.3 Contrastive and Non-Contrastive Focus in Spanish. Clause-Level Prominence and Distribution

The main objective of the complex analysis in [6] is to highlight the intonational-prosodic features of the two most important types of Focus, Contrastive Focus (CF) and Noncontrastive Focus (NF), the latter being in general Discourse-New focused entities, in different discursive contexts and for certain application domains (i.e. scopes) represented as clause-level constituents. The result of CF and NF classification can be summarized as within the following statements concerning the relationship CF-NF.

Depending on the scope of the Focus category, [6] could demonstrate that for the application domains Subject, Predicate, and Broad scope, between CF_{Subject} and NF_{Subject}, CF_{Predicate} and NF_{Predicate}, respectively CF_{Broad} and NF_{Broad}, there exist small or not significant syntactic differences. In other words, between Contrastive and Non-contrastive (informational, as typical examples) Focus there are not significant differences on the syntactic analyses exhibited by the two types of Focus, when applied to the three intra-clause structures.

Regarding the differences between CF and NF when applied to the syntactic constituents Verb and Object of the clause, a different trend has been noticed as expressing of the two types of Focus. More precisely, the participant subjects to the intonational-prosodic experiment described in [6] preferred shorter and simpler clauses for NF by using pronouns or by excluding lexically that phrase structure part whose meaning is presupposed in the discourse.

The most important contribution of the prosodic experiments in [6] we have to take into account is the results that the distinction between the two most important types of focus, CF and NF, can be revealed just relying on the intonational-prosodic features, without the (powerful) syntactic tool of word order, contrary to the classical results represented mainly by [45]. Moreover, [6] could specify those prosodic features that play an important role in the differentiation process between CF and NF. Namely, only the duration and pitch features proved to be relevant for the contrastive Focus specification.

Since the relative duration proved to be not essential in CF-NF differentiation, the two details suggest that the main difference between CF and NF is of phonetic nature primarily. This conclusion is consistent with similar results found for English in terms of differentiating between various types of focus [39], [38], [19]. The observations in [6] should be analyzed carefully in the context of our special interest for computational tools relying on IS theories, syntax, and discourse context, aiming to achieve consistent linguistic models and reliable algorithms for the prosody prediction of the Romanian CF, including valuable solutions to the ICF problem.

Another really useful contribution of the work achieved in [6] is the proposed model of organizing the intonational-prosodic experiments, by a coherent group of considerations regarding the subjects, scenarios, discursive contexts, Focus types and their syntactic scopes within the (Catalan) Spanish clause, results that are easily transferable and portable to the Romanian.

5 Algorithms for ICF Prosody Prediction

In Section 2 we showed the Jan Firbas definitions for the concepts of Systemic Ordering (SO), depending or free of a certain predicate-predication p), Communicative Dynamism (CD) of a sentence (actually, finite clause), and the CD degrees of each of the clause constituents playing the syntactic-semantic roles [15], [16], [17], [42]. In this section we refer and speculate the well-known Prague and Brno Schools in Computational Linguistics and their results and intuitions on the Information Structure (IS) theory, proposing algorithms to solve the ICF problem for the Romanian finite clause, thus trying to design concrete solutions and reliable approaches for the prosody prediction of the Contrastive Focus in the Romanian clause and sentence.

5.1 Prague School's Functional Perspective of Spoken Sentence

As discussed in Section 2, the Prague and Brno CL Schools found that a context independent (contextually-unbound) element in the clause syntactic development is communicatively more important, thus carries a higher degree of CD than an element already present in the (scenario) context of the clause predication [40]. J. Firbas supports even earlier that "a context dependent element can never exceed a context independent element in CD." [15], p. 34. This is true but only for the entities which can be labeled with Discourse-Given (G marked) and Discourse-New features. However, one may find entities which are both G marked but also F marked, thus contrastively focused in the relationship Background-Focus. Such a constituent is intonationally always more prominent than any other constituent in the finite clause, according to CFPR (Contrastive Focus Prominence Rule) in subsection 3.1. Such a situation is illustrated in example 1, Section 1.

For the CD degrees computing process of the finite clause constituents, there are two closely related problems to be solved, in theory and practice: (a) As it follows from the Prague and Brno Schools' approaches to IS theory, the CD degrees are lower for the contextually-

bound categories (thus Discourse-Given, G marked entities), higher for the contextually-nonbound categories (Discourse-New entities), and should be the highest for the contrastively focused (F-marked) entities in the finite clause. But this last statement should hold whether the Prosodic Prominence (PP) of the semantic roles would rise monotonically increasingly with the CD degrees for the syntactic constituents of the semantic roles. This is the challenging second problem for the CD degrees computing: (b) Which is the relationship between the CD degree of a certain semantic role in a finite clause, and its PP within that clause?

As already expressed, assigning the CD degrees to the spoken sentence (clause) depends on four important factors (in the order of ascending priority): linear modification (i.e. word order), the semantic relations of the clause constituents, discursive context (called also retrievability) of the immediately preceding context, and the prosodic features of the clause semantic roles. For the written text remain, of course, only the first three factors do matter. J. Firbas [16] initiates the following legitimate question: Can CD degrees of the clause syntactic constituents be equated with PP? [16], p. 216. This is exactly the task we proposed to solve as an essential prerequisite step to the ICF problem.

J. Firbas shows credibly the strong relationship between the CD degrees of the clause constituents and the IS dimension of Topic-Comment in the finite clause. Furthermore, J. Firbas highlights the more complete structure of Topic-Transition-Comment (under the former notation Theme-Transition-Rheme) and Comment-Transition-Topic (or Rheme-Transition-Theme, respectively) [16]. An important result is that the SO-order or SO disorder of the clause constituents does not entail necessarily the Topic-Comment entities ordering. The example 2 in Section 1 illustrates this fact, proved also within the A. Leong's Inference Boundary (IB) algorithms [31] for computing the Topic-Comment structure of the clause.

Trying to evaluate the essential relationship between CD degrees and PP (Prosodic Prominence) of the constituents within a finite clause, J. Firbas enounces the following three fundamental (and still

provoking) questions [16], p. 218: "Is linear modification the consequence of degrees of CD? Are the semantic characters of an element (its semantic character and the character of the semantic relations into which it enters) the consequence of a degree of CD? Is the operation of the immediately relevant context the consequence of a degree of CD?"

The answers to all these questions are considered to be negative [16]. The Firbas' general solution is the "outcome of the interplay of the three factors mentioned that determines the degrees of CD and their distribution..." within the clause constituents. "The degrees of CD... are the consequence of the interplay" among the three mentioned factors (discursive context, constituent semantics, and constituent ordering) and, by consequence, of their PP in the clause.

An intriguing fact, worth to be revealed: pointing out the importance of the context-dependent vs. context-independent, or contextually-bound vs. contextually-nonbound in the Prague School's language [20], or retrievable vs. irretrievable from the immediate context [16], or discursive context discriminated entities [16], J. Firbas reminds an outstandingly valuable result of Keijsper (1985) [27], whose importance was rediscovered two decades later in [39], and used to propose a new IS tagging system in our work [10]: differentiating between Discourse-New and Discourse-Given entities serves no effective utility for IS computing since not only Newness but also Givenness IS dimension can be focused within the clause by the more phonetically powerful contrastive Focus intonation.

J. Firbas provides sound arguments, discussed also in [42], that the contextual factor plays the dominant role in comparison to both the linear ordering and the semantic factor of the finite clause constituents. Since prosody is sensitive to the immediately changing context, in the example that follows, the pronouns are naturally context dependent. Even they (or it) should be dislocated within clause, the status of context-dependency is preserved, leaving to the other constituent(s), correspondingly dislocated (ex. 16.c) or not (ex. 16.b), the PP quality, thus the (contrastively) Focus marking.

16.a. I saw him. I met him in the street.

16.b. L-am văzut. L-am întâlnit pe stradă.

16.c. L-am văzut. Pe stradă l-am întâlnit.

Analyzing in [17] the measure in which the functional perspective of a spoken clause (thus its prosodic behaviour) can be predicted from the written counterpart of that clause, J. Firbas establishes that the PP (Prosodic Prominence) of a certain clausal constituent is a vehicle for the speaker's Comment (in the Topic-Comment IS dimension) to the further development of the communication, rising its CD degrees. The important deduction is that, "...to a certain extent distribution of degrees of PP over the spoken sentence can be predicted from the distribution of degrees of CD as determined by the interplay of the non-prosodic factors..." [17], p. 48–49, thus of the discursive context, semantics, and constituent order. Our main tool, proposed and used in [10] and the present paper, is the last one, possibly and partially combined with the first two ones, e.g. the contextually bound-nonbound computation of some categories in the TFA algorithms [8], [9], and certain easily identified semantic features of the others.

5.2 Computing the Communicative Dynamism Degrees as a Progressive Solution to the ICF Problem

Our purpose is to estimate the Focus categories in a finite-clause of ICF affirmative type. Two positions of a syntactic-semantic role within a finite clause are important: (1) The position of the corresponding constituent in the order of Systemic Ordering (SO) of the clause constituents established for the enclosed predicate (SO_p) or, in general, in the SO (predicate p-free). The constituent first position is naturally denoted as its place into the SO-order. (2) The second position of a constituent as syntactic-semantic role within a finite affirmative clause corresponds to the effective, real order of the constituent within the text at-hand containing the clause to be analyzed.

Since the aim is to compute the CD (Communicative Dynamism) degree of the constituent corresponding to its current place within the ICF-type clause, and its PP (Prosodic Prominence) is considered to be an increasing monotonic function of its CD degree (this function-shape hypothesis is debatable and could be replaced with another functional

pattern), the constituent place in the real, current order of the clause constituents is denoted with CD-order. For computing the CD degree of a constituent we associate to it the measure of its SO-disorder in comparison to its position within the SO-order. The measure of SO-disorder or, equivalent, the measure of the constituent CD degree, positioned in its CD-order, is proposed to be the "distance" between the CD-order and SO-order positions, and computed as the number of permutations necessary to remove the dislocated constituent from its CD-order and to reposition to its location in the SO order. Conventionally, the predicate is considered to change not its position relative to the SO-order in the clause.

For the textual SO-order of an affirmative finite clause, we established experimentally an ordering for the PP (Prosodic Prominence) values of the SO-ordered constituents (Fig. 1). This order is represented as Focus weights taking values in the interval $[0, 1]$, assigned to each syntactic-semantic role in the clause, for the SO-ordered constituents. The graph in Fig. 1 should be refined both on statistical bases and prosodic experience for utterances of the SO-ordered constituents.

The SO-coefficients assigned to the finite clause constituents in the SO order are used within the following algorithm for the estimation of PP values on the basis of the CD-degrees computed for the constituents in CD order (thus SO-disorder) in an affirmative ICF-type finite clause. The algorithm is derived from the General Procedure for Hierarchical Intonational Focus and Break Assignment on the Discourse-Prosody Interface [10], p. 10.

ECF-ICF Prosody Prediction Algorithm of the Romanian Finite Clause

1. Recognition of finite and non-finite Verbal Groups (VGs), of the Noun Groups (NGs) whose semantic head is predicationally marked, followed by text segmentation at finite and non-finite clauses;
2. Recognition of all the constituents and their syntactic dependencies in each finite clause;

3. Extracting the syntactic roles of the intra-clausal constituents;
4. Determining of the phrase markers for explicit contrastive Foci (e.g. *chiar*, *evident*, *doar*, *nu*, *nici* etc.) the Focus marker database (many of these phrase markers are also textual boundaries for the intra-clause, inter-clause, and rhetorical discourse text structures parsed with the SCD method [11]);
5. Determining the CD-degrees of the clause constituents by computing, for each constituent, the measure of its dislocation in comparison to its place in the SO-order of the clause;
6. Determining the Focus weights for the clause constituents by multiplying the SO-coefficient assigned to that constituent when positioned in the SO-order, and the Focus weight representing the dislocation measure (or "distance") of that constituent in comparison to its place in the SO order of the finite clause at hand. This is the contents of the ICF Estimation Scheme, which represents the core of estimation modeling for the Romanian ICF. Experiments to obtain schemes, procedures, and patterns for a better ICF estimation modeling are imperatively necessary;
7. For the constituents modified by a lexical contrastive marker as explicit (contrastive) Focus, we establish the weight values for the Focus as the maximal value among the clause constituent weights, since the constituent marked by contrastive Focus (explicit) marker has always maximal PP (according to the CFPR principle – subsection 3.1).
 - (a) If the finite clause contains a single constituent marked at contrastive Focus by a single contrastive marker, either explicit or implicit, then the corresponding constituent receives maximal PP.
 - (b) If several explicit contrastive markers do occur in the clause, and several constituents are contrastively marked (FOF, SOF, TOF, i.e. First-Second-Third Occurrence Focus), then

the PP value order in the clause is given by the prosodic hierarchy that can be established among the contrastive Focus markers, correlated or not.

8. The result of the algorithm consists in a vector of weights associated to the clause constituents in the CD-order, these weights representing the CD degrees, thus PP value ordering of the clause constituents.

In several situations we confirmed the intonational inequality $\text{FOF} >_{\text{phon}} \text{SOF}$ to hold, but found also counter or ambiguous examples. This makes the intonational inequality to depend actually on (possibly contextual) hierarchies of lexical contrastive markers that are applied to the focused constituents. This is the general image of the prediction models and techniques we propose to approach the (Romanian) prosody prediction (derived from [10] mainly): contextually-driven hierarchies on ECF and ICF lexical and non-lexical focusation markers.

As already mentioned, this algorithm is an adaptation of the General Procedure for Hierarchical Intonational Focus and Break Assignment on the Discourse-Prosody Interface for the finite clause-level constituents, described in [10]. The following useful observations on the above proposed ECF-ICF Prosody Prediction Algorithm are appended:

1. Finite clause recognition (segmentation) is necessary because each clause, including the non-finite one or implied-predicate supporting, has its specific Focus structure.
2. When computing the Focus weights (CD degrees) of the constituents, we take into account only their semantic heads, to the rest of the component phrase structure being assigned the same PP as that assigned to the constituent lead.
3. We chose to use syntactic instead semantic roles for the clause constituents since the cost of semantic role computing is too much in comparison to the potential benefit when evaluating the Focus degrees (PP values).

4. ECF markers are taken from (an updated) database of Focus markers created initially by the intonational text annotators.
5. The purpose of the algorithm is to estimate the tones and pauses for the finite-clause constituents in order to obtain a reliable prediction of the Romanian clause prosody. As general rule, the tones are established depending on the constituent Focus degrees, while the pause patterns have to take into account the number of constituent syllables but also the PP value sequence of the clause constituents.

ICF Generation and Evaluation Procedure

Four phases are used to artificially generate and evaluate the F0 signal for an ICF-type affirmative finite clause.

Step1. The ICF-type clause is synthesized with the Site-Vox program [2].

Step2. For the clause constituents, the Focus degrees (PP values) are computed, the sequence of these values being represented by two graphs: in the first one, the Focus degree values of the clause constituents are linked as nodes by direct edges, while in the second representation graph, the edge slopes are weighted with values that follow the general shape of the F0 signal for an affirmative finite clause. Both these graphs (with better accuracy for the second one) represent the PP pattern of the ICF-type affirmative clause at-hand.

Step3. The F0 signal output in the Step1 is drawn manually, closely to the second graph in Step2 which represents the PP values assigned to the ICF-clause constituents. The PP values are computed with the ICF Estimation Scheme in the step 6 of the ECF-ICF Prosody Prediction Algorithm.

Step4. The obtained signal is compared both in prosodic and numerical terms with the gold-registration of the original ICF-type affirmative finite clause.

An example of the result of the algorithm is given in Figure 2 below.

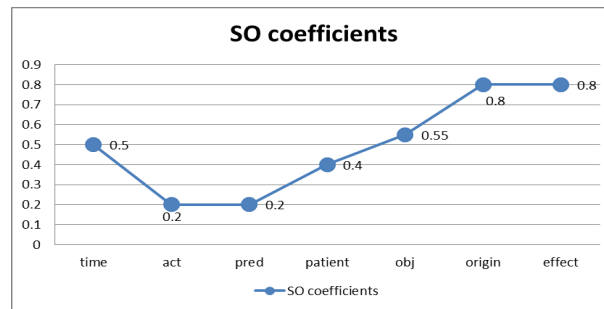


Figure 1. The Focus coefficients estimated for the SO affirmative clause

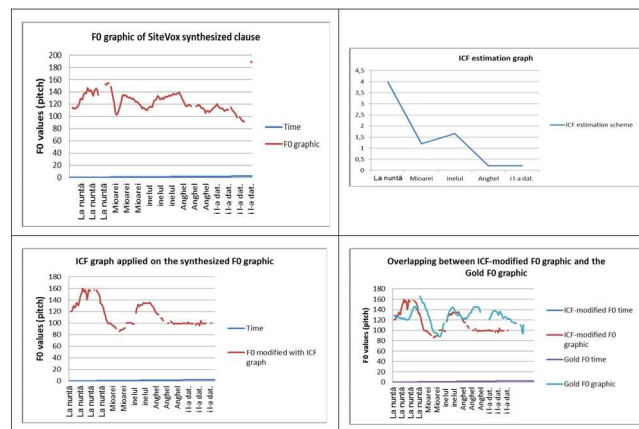


Figure 2. F0 graphic for the Ex.2-clause ICF-modified with Ex2Graph

6 Conclusions

In the algorithm for the estimation of PP values (i.e. Focus degrees) for the constituents of an ICF-type affirmative finite clause, the coefficients representing the Focus degrees are computed, for each constituent, by multiplying the constituent SO-weight with the dislocation measure (or "distance") of the same constituent, compared to its standard position in the clause SO order. For the basic operations enclosed in the ICF Generation and Evaluation Procedure, one can imagine and obtain semi-automatic or automatic treatments, but the key part of the Procedure is the computation of prosodic pattern for the ICF-type finite clause at-hand, namely the ICF Estimation Scheme in step 6 of the ECF-ICF Prosody Prediction Algorithm. The current approach we proposed is to compute the "distance" of a CD-dislocated constituent compared to the SO-position of the constituent, for the SO-order of the constituents in the ICF-type affirmative finite clause at-hand.

The main but partial result is that the ICF scheme applied for estimating the Focus distribution (or Prosodic Prominence) of the constituents within an ICF-type affirmative finite clause is not good enough. We need to design another ICF estimation scheme and to test it, in similar conditions, on the same test set of ICF finite clauses, thus to be applied on the same type of constituents and constituent dislocations from the clause SO. Regarding the inappropriateness of the current ICF estimation scheme, an important topic in the discussion is the 'correctness' of the Gold signal for the same analyzed clause, which the ICF-modified F0 graphic is compared to. Namely, the Gold registration for the ICF-type clause should be the most probable from the clause utterances for which the Focus distribution within the clause corresponds to the pattern / configuration of the dislocated constituents. If, and only if, the Gold signal of the ICF-clause at hand is 'correct', then its comparison to the ICF-modified clause by the ICF estimation scheme is meaningful.

The following two actions have to be taken in the near future: (a) to rediscuss / reestablish, if necessary, the Gold registrations for the ICF-type finite clauses that are analyzed, and (b) to propose another ICF

estimation scheme on the basis of which the F0 graphic of the SiteVox synthesized clause, modified on the basis of the new ICF estimation scheme, to match as close as possible the F0 Gold graphic shape of the clause.

Extensive experiments are needed to perform the cycle of the four steps of the ICF Generation and Evaluation Procedure, such that the outcome of the prosodic pattern to represent a consistent computing model and a reliable ICF estimation scheme for the prosody prediction of the Romanian ICF-type affirmative finite clause.

References

- [1] G. Alboiu, V. Motapanyane. *The generative approach to Romanian grammar: An overview*. In V. Motapanyane (Ed.), *Comparative Studies in Romanian Syntax*. Amsterdam; New York: Elsevier, 2000, pp. 1–49.
- [2] V. Apopei, D. Jitcă. *SiteVox Synthesizer Program*. Institute of Computer Science, Romanian Academy, Iași Branch, Version 2012.
- [3] C. Butnariu, N. Curteanu, C. Bolea. *Levels of Finite Clause Parsing: Implications and Applications*. *Memoriile Secțiilor științifice / Memoirs of the Scientific Sections, Series IV, Tome XXXIV*, 2011, ISSN: 1224-1407, pp. 41–54.
- [4] N. Chomsky. *The Minimalist Program*. The MIT Press, Cambridge, Massachusetts, 1995, 420 p.
- [5] N. Chomsky. *Derivation by Phase*. In M. Kenstowicz, editor, "Ken Hale: a life in language". MIT Press, Cambridge, MA, 2001, pp. 1–52.
- [6] Hye-Yoon Chung. *Two Types of Focus in Castilian Spanish*. Ph.D Thesis, The University of Texas at Austin, Austin, Texas, Dec. 2012, 216 p.

- [7] N. Curteanu, D. Trandabăț, M.A. Moruz. *Syntax-Prosody Interface for Romanian within Information Structure Theories*. In C. Burileanu and H.N. Teodorescu (Eds.), Proceedings of SpeD-2007 Conference (Speech Technology and Human-Computer Dialogue), Romanian Academy Editorial House, Bucharest, 2007, pp. 217–228.
- [8] N. Curteanu, D. Trandabăț, M.A. Moruz. *Discourse Theories vs. Topic-Focus Articulation Applied to Prosodic Focus Assignment in Romanian*. C. Burileanu, H.-N. Teodorescu (Eds.) SPED-2009 Conference: Advances in Spoken Language Technology, Constanța, Romania, The Publishing House of the Romanian Academy, Bucharest, 2009, pp. 127–140.
- [9] N. Curteanu, D. Trandabăț, A. Moruz. *Expanding Topic-Focus Articulation with Boundary and Accent Assignment Rules for Romanian Sentence*. Vaclav Matousek, Pavel Mautner et al.(Eds.), Proceedings of the 12th International Conference TSD-2009, Plzen, Czech Republic, Lecture Notes in Computer Science, No. 5729, 2009, pp. 226–233, <http://portal.acm.org/citation.cfm?id=1617166>.
- [10] N. Curteanu, D. Trandabăț, C. Bolea. *Integrating Contrastive Focus with Givenness and Topic-Comment: A Hierarchical Focus Architecture of the Romanian Discourse-Prosody Interface*, Proceedings of SPED-2011, 6-th Conference on Speech Technology and Human-Computer Dialogue, Brasov-Romania 2011, IEEE Conference Publications Program, ISBN: 978-1-4577-0440-6, 2011, pp. 146–158, IEEE Xplore Digital Library, <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5929232>.
- [11] N. Curteanu. *The Segmentation-Cohesion-Dependency Parsing Strategy and Linguistic Theory*, TehnoPress, Iași, România, xix + 429p., ISBN 987-973-702-928-7, 2012.

- [12] C. Dobrovie-Sorin. *The Syntax of Romanian*, Berlin, Mouton de Gruyter, 1994.
- [13] A. Dufter, D. Jacob (Eds.). *Focus and Background in Romance Languages*, Amsterdam / Philadelphia, John Benjamins Publishing Company, 2009.
- [14] C. Féry, S. Ishihara. *Interpreting Second Occurrence Focus*, Interdisciplinary Studies on Information Structure, S. Ishihara, M. Schmitz, and A. Schwarz (Eds.), Potsdam University, Germany, 2005.
- [15] J. Firbas. *A Functional View of 'Ordo Naturalis'*. Studia Minora Facultatis Philosophicae Universitatis Brunensis, K1 (1979), Brno Studies in English No. 13, Vol. K1, 1979, pp. 29–59.
- [16] J. Firbas. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press, 1992.
- [17] J. Firbas. *Can the Functional Perspective of a Spoken Sentence Be Predicted From That of Its Written Counterpart?* Studia Minora Facultatis Philosophicae Universitatis Brunensis, Brno Studies in English No. 20, Vol. K15, 1993, pp. 23–49.
- [18] E. Göbbel. *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.), 2003, p. 75–92.
- [19] C. Gussenhoven. *Types of Focus in English*, In Lee, C., Gordon, M., Büring, D. (Eds.): *Topic and Focus. Cross-Linguistic Perspective on Meaning and Intonation*. Springer Verlag, Heidelberg, 2007, pp. 83–100.
- [20] E. Hajičová, E. Skoumalova, P. Sgall. *An Automatic Procedure for Topic-Focus Identification*, Computational Linguistics, Vol. 21, No. 1, 1995, p. 81–94.

- [21] E. Hajičova, B. Partee, P. Sgall. *Topic-Focus Articulation, Tripartite Structures, and Semantic Content*. Dordrecht: Kluwer Academic Publishers, 1998.
- [22] M.A.K. Halliday, C.M.I.M. Matthiessen. *An Introduction to Functional Grammar*. (3rd. ed.), London: Edward Arnold, 2004.
- [23] K.v. Heusinger. *Intonation and Information Structure*. Habilitationsschrift, Faculty of Philosophy, University of Konstanz, June 1999, 236 p.
- [24] K.v. Heusinger. *Discourse Structure and Intonational Phrasing*, In: Chungmin Lee, M. Gordon, D. Büring (Eds.) *Topic and Focus. Cross-Linguistic Perspective on Meaning and Intonation*, Springer, Heidelberg, 2007, pp. 265–290.
- [25] V. Hill. *Complementizer Phrases (CP) in Romanian*, Italian Journal of Linguistics (Rivista di linguistica), Pisa, Italy, Vol. 14, No. 2, 2002, pp. 223–248.
- [26] V. Hill. *Discourse markers in interrogative clauses*. *Balkanistica*, Vol. 16, 2003, pp. 71–96.
- [27] C.E. Keijsper. *Information Structure*. Amsterdam: Rodopi, 1985.
- [28] I. Krapova. *Word order in Topic-Focus structures in the Balkan languages*, in Basciani A, Tarantino A, L'Europa D'Oltremare. *Contributi italiani al IX Congresso Internazionale dell'Associazione Internazionale d'Etudes du Sud-Est Européen*, ROMA, Bagatto Libri, Vol. 17, 2004, pp. 139–162.
- [29] M. Krifka. *Basic Notions of Information Structure*. In Féry, C., G. Fanselow, M. Krifka (Eds.) *The notions of information structure*, University of Potsdam, Germany, 2006, pp. 13–55.
- [30] C. Lee, M. Gordon, D. Büring (Eds.). *Topic and Focus. Cross-Linguistic Perspective on Meaning and Intonation*. Springer Verlag, Heidelberg, 2007.

- [31] P.A. Leong. *Delimiting the Theme of the English Clause – An Inference-Boundary Account*, SKY Journal of Linguistics, Vol. 17, 2004, pp. 167–187.
- [32] P.A. Leong. *Private Communication*. 2010.
- [33] A. Manolescu, D. Olson, M. Ortega-Llebaria. *Cues to Contrastive Focus in Romanian*, The University of Texas at Austin, Research Report, 2007, 16 p.
- [34] E. Onea, K.v. Heusinger. *Grammatical and Contextual Restrictions on Focal Alternatives*, In Dufter, A., D. Jacob (Eds.) *Focus and Background in Romance Languages*. Amsterdam / Philadelphia, John Benjamins Publ., 2009, pp. 281–308.
- [35] A. Rosetti. *Istoria limbii române*, Vol. 4-5-6, București, Editura științifică, 1966.
- [36] L. Rizzi. *The Fine Structure of the Left Periphery*, In L. Haegeman, (ed.) *Elements of grammar*, Dordrecht: Kluwer, 1997, pp.281–337.
- [37] V. Samek-Lodovici. *Topic, Focus, and Background in Italian Clauses*, In Dufter, A., D. Jacob (Eds.) *Focus and Background in Romance Languages*. Amsterdam / Philadelphia, John Benjamins Publishing Company, 2009, pp. 333–357.
- [38] E. Selkirk. *Contrastive FOCUS vs. presentational focus: Prosodic evidence from right node raising in English*. In B. Bel and I. Marlin (Eds.), *Speech Prosody 2002: Proceedings of the First International Speech Prosody Conference*, Laboratoire Parole et Langage, Université de Provence, Aix-en-Provence, 2002, pp.643–646.
- [39] E. Selkirk. *Contrastive Focus, Givenness and the Unmarked Status of Discourse-New*. In Féry, C., G. Fanselow, M. Krifka (Eds.) *The notions of information structure*, Univ. of Potsdam, 2006, pp. 124–146.

- [40] P. Sgall, E. Hajičová, J. Panevova. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Ed. by J. Mey. Dordrecht: Reidel and Prague: Academia, 1986.
- [41] M. Steedman. *Information-Structural Semantics for English Intonation*, In: Chungmin Lee, M. Gordon, D. Büring (Eds.): *Topic and Focus. Cross-Linguistic Perspective on Meaning and Intonation*, Springer, Heidelberg, 2007, pp. 245–264.
- [42] A. Svoboda. *An ABC of Functional Sentence Perspective (Part One)*, *An e-learning text for the students of English*, Opava, 2005, 15 p.
- [43] Tseng, Chiu-yu, Zhao-yu Su, Lin-shan Lee. *Prosodic Patterns of Information Structure in Spoken Discourse – a Preliminary Study of Mandarin Spontaneous Lecture vs. Read Speech*, *Speech Prosody 2010 Conference*, Chicago, P4.39, 2010, 4p.
- [44] M. Zimmermann, C. Féry (Eds.). *Information Structure. Theoretical, Typological, and Experimental Perspectives*. Oxford University Press Inc., Oxford, New York. M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 2010.
- [45] M.-L. Zubizarreta. *Prosody, Focus and Word Order*. MIT Press, Cambridge, Massachusetts, 1998.

Neculai Curteanu¹, Alex Moruz², Cecilia Bolea¹, Andrei Scutelnicu²

¹Institute of Computer Science, Romanian Academy, Iași Branch
E-mail: cecilia.bolea@iit.academiaromana-is.ro

²Institute of Computer Science, Romanian Academy, Iași Branch, and Faculty of Computer Science, "Alexandru Ioan Cuza" University of Iași
E-mail: mmoruz@info.uaic.ro
E-mail: andreis@info.uaic.ro