

Functional FX-bar Projections of the (Romanian) Verbal Group and Sub-Groups on the Syntactic-Semantic Interface*

Neculai Curteanu, Diana Trandabăț, Mihai Alex Moruz

Abstract

The aim of this paper is to investigate the syntactic / semantic substructures (called subgroups) of the Romanian verbal group (VG) [12], or verbal complex [25], starting with the achievements in the literature, and melted into the device of direct and inverse functional projection within FX-bar theory [7]. The paper examines several problems and their solutions for the syntactic-semantic theories of VG, as discussed in some fundamental papers, and we offer our explanation on the involved syntactic phenomena, the emphasis falling on the VG substructures (verbal subgroups, VSGs), VSG boundaries and composition within VG, direct and inverse FX-bar projections of VG, VG parsing, lexical semantics and intensional / extensional logic representations of the Romanian (verbal or nominal) predicate.

Keywords: functional FX-bar theory; verbal group; verbal subgroups; Romanian predicate and predication; intensional / extensional semantics; clause-level parsing.

1 Introduction

The aim of this paper is to investigate the syntactic / semantic substructures of the Romanian *verbal group* (VG), or *verbal complex* [1], [25], [16], starting from the instruments and current achievements in the literature, and melted into the device of (direct and inverse) FX-bar projections and theory [4], [5], [6], [7], [8], [12], [13]. Since [8]

we separated within the verbal group (VG) the so-called *default verbal group*, denoted then as *verbal group kernel* (VGK), that substructure of VG which may commute with its proper, ‘*outside*’ adverb, and does not enclose this modifier. The same VG / VGK substructure has been described in more detail as the *inside* part of the *verbal complex* [1], with its *inside* (special) adverbs (mai, cam, prea, și, tot) and *outside* negations and adverbs. In the present paper, the proper syntactic / semantic *substructures* of VG (together with VGK) shall be called *verbal subgroups* (VSGs).

Examples of VGs [10], [11]. VGK is represented in parentheses, included in VG; the unaccentuated pronouns (pronominal clitics) are in *italics*: “ nu că (nu *mi-l* va mai și plăti) greu; (nu-*i* cunoșteam); (*li* se cereau) ; (*își* mai recăpătase) ; (Ai consultat) ; (ar fi simțit) ; (*i* se așternea) ; (să se întâmples) ; (nu se putea abține); (n-*o* putea lua); (Nu *i*-ar fi trecut); (să poată afla); (să te intimideze); (să *vă* văd lucrând) ”.

As [1] remarks rightfully, VG provides both an *outside* (**nu**₁) *negation* and an *inside* (**nu**₂) *negation* (e.g. **nu**₁ să **nu**₂ te duci), which can be interpreted as outside VG and inside VGK *quantifiers*. Similarly, there exist as VG *inside modifiers* the *special adverbs* (mai, cam, prea, și, tot), and the proper, VG *outside adverbs* (nu₁ să nu₂ te **tot**₂ duci **imediat**₁). The structure of VGK as the “inside” of VG, with a *syntactic head* (the *tense auxiliary*, bearing the number and person, when present lexically) and a *semantic one* (the *predicational head verb*, often called also the matrix, or lexical, or embedded verb), with clitics ‘inside’ and semantic (direct) arguments ‘outside’) the VGK, this verbal structure is playing an essential role in the development of the *lexical predication*.

VG may be seen as the ‘last’ shell of VGK, while the contents of VGK may be interpreted as the *clause-shadow* (of the regular clause) that projects itself onto the clause, as well as representing the projection(s) of the lexical-semantic head bearing the *predicationality feature* (e.g. [8]), using *diathesis transformations* and *semantic diathesis functions* associated with semantic restrictions on predication arguments (see [16], [23], [24], [1], [21]). Furthermore, along with its finite or non-finite *predicational* head, the VG may contain (some) other verbs,

including tense and passive auxiliaries, semi-auxiliaries, and restructuring (*modal, aspectual, motion*) verbs. Our goal is to investigate the VG syntactic-semantic substructures (Verbal SubGroups – VSGs), the relationship VG-predication-predicate, and whether a VG contains a single or a multiple predicate (thus clause, thus potential discourse segment).

This paper will examine several problems and their solutions for the syntactic theories of VG, as discussed in some fundamental papers such as [25], [16], [1], [21], and we shall try to offer our approach on the involved syntactic phenomena, the emphasis falling on the lexical semantics and intensional / extensional logic representations, our interests being mainly oriented towards VG parsing, VG substructures (VSGs), VSG composition and their FX-bar projections, VG phonology and (local) prosody [25; Chap.9], [14], [15].

1.1 The Classical Predication vs. Lexical Predications

The *classical predication* pair (Subject, Predicate) can be viewed as just *one of the facets* of the VG (verbal complex) whose semantic head bears the *predicationality feature* PREDF [8], the other ones, equally righted as “classical predications”, being instantiated by the predication verb (lemmatized form), endowed with clitic(s) as affixed inflexion(s), which are obligatory present when their valence-commanded arguments are personalized or focused (*i.e. theta-disordered*), doubled or not by the corresponding semantic arguments. Thus, the classical predication pair corresponds to the subject theta-role of “actor” or “actant”, while the other “classical” predications may associate, valence-driven, the theta-roles of “patient” and/or “receiver” and/or “addressee” to semantic arguments (but not adjuncts!). All these are commanded (or not) by the presence (or absence) of the PREDF predicationality feature, assigned at the *lexical level*, to the semantic head in VG.

In a *first move*, the classical predication pair (Subject, Predicate) should be reduced to the pair (Subject, PREDF_verb) corresponding to the *theta*-role of “actor” or “actant” in the valence-driven SUBCAT (or ARG-ST vector [26]), with 1 to 3 semantic ar-

guments. It is important to specify that there exist normally at least two SUBCAT lists: $SUBCAT_{oblic_order}$, containing the syntactic arguments of the $PREDF_verb$, in the order of increasing obliqueness, and $SUBCAT_{theta_order}$, enclosing the arguments in the $theta$ -order (or *systemic* order) for the valence-based arguments of $PREDF_verb$. Usually, (only) *for the active voice* and a normal semantics of predicationality, these arguments should coincide.

In a *second move*, to this first classical predication are added, equally righted in the $theta$ -semantics, the following similar “classical” predications (Figure 1):

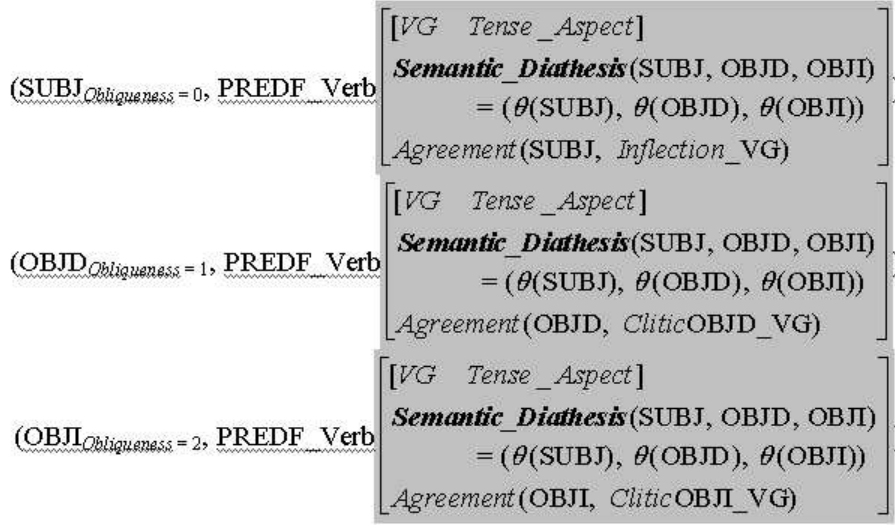


Figure 1. All the extended, valence-based ‘classical’ predications

In Figure 1, SUBJ, OBJD, OBJI represents the syntactic categories of subject and direct arguments (direct and oblique complements), respectively. The **Semantic_Diathesis** function, depending on the valence-value of the predicational verb at the lexicon level, *links* (in the sense of *linking theory* [28]) the grammatical (syntactic) arguments SUBJ, OBJD, OBJI (sometimes, OBJD2 at the shallow level) to their semantic, *theta roles* (e.g. Actant, Patient, Addressee etc.) θ (SUBJ),

$\theta(\text{OBJD})$, $\theta(\text{OBJI})$. The *Agreement* function establishes anaphoric local bindings between the verb inflection and its object (pronominal) clitics, on one hand, and the syntactic (SUBJ, OBJD, OBJI) arguments, respectively, on the other hand.

These are the new ‘traditional’ predications, with their real engine, *viz.* the *predicational feature* PREDF, installed on the (lexical) verb head of the verbal group VG. Similarly, non-finite forms of PREDF verbs may be associated to those Ns (Nouns called *nominalizations*) and/or As (Adjectives or Adverbs) that bear the feature PREDF.

In the ‘classical’ predications above, clitics may lack when the semantic arguments are of *non-person* or *non-animate* nature but are lexically present. This does not change the ‘equivalence’ of these newly devised valence-based predications. Such an interpretation of the VG structure has consequences in establishing the FX-bar (direct and inverse) VG projections (see the outlined solutions considered in the subsection 2.2 devoted to the problem of VG local structure and its FX-bar projections).

The problem of ‘classical’ *predication(s)* in HPSG theory [2], or the problem of the *special* role of the *subject* in the SUBCAT list of HPSG [26; Chap.9] are solved in the linguistic feature structures in Figure 1 above as follows: the feature *Semantic_Diathesis*(SUBJ, OBJD, OBJI) is not an elementary (atomic) feature value but a *function*, defined as follows: the input of the function is the VG shallow, *syntactic diathesis*, represented by the above mentioned $\text{SUBCAT}_{\text{obliq_order}}$, while the output (value) of the function is the VG *semantic diathesis*, *viz.* $\text{SUBCAT}_{\text{theta_order}}$ list. This solution forces the *subject-actor* and the *subject-least-oblique-element* (or grammatical subject) to take each one its own right place, in the right (possibly distinct) ordering.

Briefly, the *values* of the function *Semantic_Diathesis* are established as follows: the input value is represented by the tense and syntactic diathesis resulted from the VG shallow parsing. The output, or the value of the *Semantic_Diathesis* function, is obtained from the lexicon, where the head verb (*predication*) meaning is represented by specific standard lists of semantic arguments corresponding to the valence of that specific predicational category, and the syntactic diathesis

is transformed into a certain particular list of semantic arguments corresponding to the tense, diathesis, and predication meaning of that (verb) category. (See the mechanism of *dt* and *sd* functions in subsections 2.2 and 2.3, defined to make operational the FX-bar direct and inverse projections of VGK.)

In Figure 2 of the FX-bar scheme for *local structures*, the local (single-event) levels X0-X1-X2 express the *clause predication* depending on basic, lexical categories (V, N, A), while the levels CL0-CL1-CL2 express logical or (*second-order*) *predicational relations* on simple clauses.

1.2 Handing Down the Predication from Syntax into Lexis

The feature that we called *Predicationality* [8] borne at the *lexical* (even *lexicon*) level by the major lexical categories N, V, A, corresponds to what in the literature is called (more frequently, among other labels) as the *deverbal* property, or *deverbality*, of these categories. For an extended survey and analysis of this notion and its syntactic-semantic consequences, see [8]. We avoid the term *deverbality* because its meaning is *not necessarily specific* to Vs since this essential lexical-semantics feature is *equally shared* by Vs, Ns and As! Moreover, there are (classes of) verbs which do not bear this property, *e.g.* the *copulative* ones. The feature of *Predicationality* is assigned to those finite or non-finite Vs, Ns (often called *nominalizations*), and As, whose meaning involves a *process* event or *process* name. We abbreviated this feature as PRED(dication)F(eature), with two main values, PROC(ess) and STAT(e) (or EXIST).

We are mainly interested by those major categories providing a first-order predication feature, thus associated with first-order predicates whose arguments are noun groups (NGs) and not new finite VGs or VSGs. This is the *straight* or *canonical* type of predication, since there also exist second-order predications, *e.g.* as those assigned modal, inchoative, semi-auxiliary (or “restructuring” [25], *i.e.* *modal*, *aspectual*, *motion*) verbs. Thus it is important not only the valence (number of

arguments) of a predication category but also the sort-type of these arguments.

The classical notion of *predication* is known to be the pair (*Subject*, *Predicate*), an essentially *syntactic concept* meant to support the finite clause (proposition) structure. The predicate, either synthetic or analytic, encloses both *process* verbs and *state* verbs (the latter case for the nominal predicate) indiscernibly, despite the fact that only process (predicational) verbs entail a semantic argument-based syntactic distribution, corresponding to a proper *valence*. Furthermore, the feature of *predicationality* (or *deverbality*) is equally shared not only by process verbs but also by nominals *Ns* and modifiers *As* that are (in terms of lexical semantics) siblings of the corresponding predicational verbs, these non-verbal categories having a *similar syntactic distribution* of semantic arguments, with the same valence as their predicational, verbal counterparts.

Thus, the feature of *predicationality*, as a *lexical semantics* quality, is not necessarily related to the predicate (which is a syntactic construction): in the *nominal predicate*, the copulative verb is not a predicational one. The same goes for the auxiliaries incorporated within the VG, whose tense is based on compound syntactic constructions. This does not exclude, in the *nominal predicate*, that the *predicative nominal* (as semantic head of the construction) bears the feature of predicationality. *E.g.*, the predicative nominals *explanation*, *marking*, *receiving* etc. (which are *predicational nouns*) in the nominal predicates of the clauses *This is John's explanation (marking, receiving...) of the notion ...*

These reasons support the idea of *handing down* the notion of *predication* from its classical, *syntactic* level, to the *lexical*, word level of *representation* and *analysis*. The lexical semantics feature of *predicationality* (PREDF) has sometimes a contextual usefulness since the same word may, or may not, bear the feature PREDF, the process meaning depending on its contextual use. For instance, the noun *building* in languages such as English, French, Romanian, may have both the meaning of a process, with [PREDF +] (or simply, PREDF), and the meaning of an object (in this case, the process result), with [PREDF -]

(or STAT, or EXIST, or simply NPREDF values).

2 FX-bar Schemes and the Predicational Feature

We propose in Fig. 2 the following general FX-bar scheme [11] for (local and global) clause-level and RST discourse structures [22]. This FX-bar scheme is using the SCD marker classes and their graph-type hierarchy, an essential instrument to represent clause-level syntactic-semantic structures and to establish their (local and global level) dependencies, including VG as the representative structure for the *verbal predicate* and (finite) clause. In the same time, there exist *global structures* whose constructive bricks are not necessarily the finite-clause but the rhetorical *discourse-segment* of the RST discourse theory [22]. The dashed lines in Fig. 2 represent the special cases when a discourse segment is a proper subclause span and when a discourse segment splits a clause.

Compared to the version of FX-bar scheme exposed in [11], the *novelty* of *this* FX-bar scheme consists in the syntactic presence of traces I, J, K that corresponds to the valence of the VG semantic head, and embodies the (local) anaphoric agreement (and *linking* relations [28]) between VG and its *theta*-semantics *direct arguments* (*Agent*, *Patient-Object*, and *Receiver-Recipient*). The index I represents *the inflexion* of a semantic or syntactic verb head in VG that is in concord with the grammatical subject of the clause, while J and K, when lexically (overt) present, are *pronominal clitics*. Comprising the VG *direct-argument* (or *linking*) *indices* within the VG syntax represents an effective need in VG FX-bar projection and comes into play when applying the *linking algorithms* [28], fortifying the idea of viewing VG as the *clause-shadow* structure.

2.1 FX-bar Direct and Inverse Projections of VG

In the next subsection 2.2 we introduce *diathesis transformations* and *semantic diathesis functions* as useful tools in describing the *lexical*

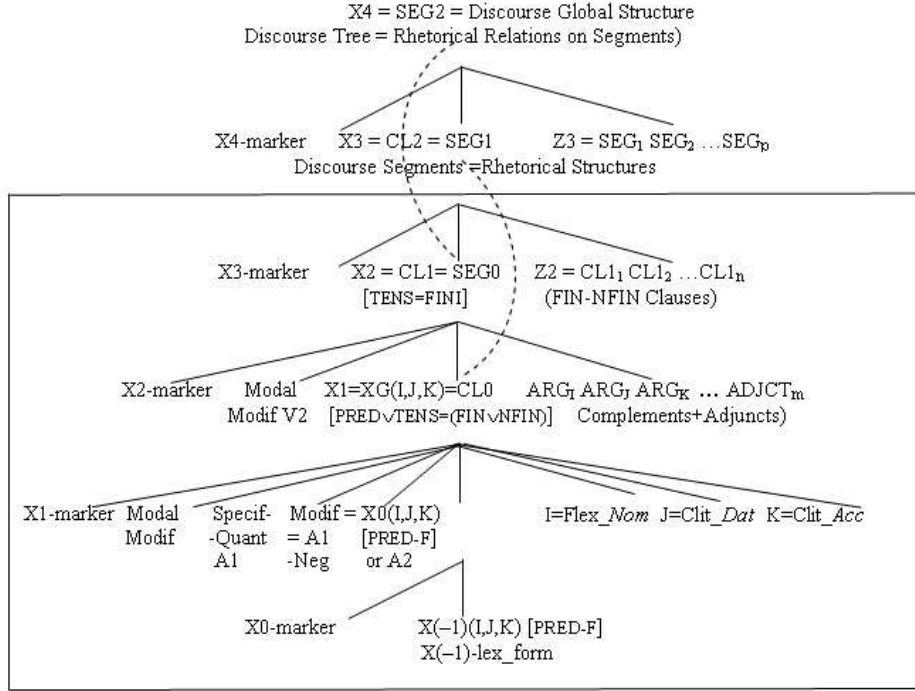


Figure 2. Global (clause and discourse) level FX-bar Scheme

predication metamorphosis from syntactic (shallow) diathesis to semantic diathesis as a top-down and bottom-up movement, from text to lexicon and backward. This mechanism may also be understood as procedures of direct and inverse FX-bar projection procedures of VG toward its (predicational) semantic head and to the clause, derived from the diathesis analysis (as in [20]), stated as solutions to the following VG FX-bar projection (Figure 2) problems:

FX-bar(VG): The problem of FX-bar direct projection of VG: To show how the *clause-shadow* information (see above) incorporated into VG is (directly) FX-bar projected into a (finite or non-finite) regular clause.

FX-bar⁻¹(VG): The problem of FX-bar inverse projec-

tion of VGK: To obtain an improved linguistic mechanism by which a *predicational category* (from the lexicon) is FX-bar projected on VG (VGK). This means to establish the FX-bar *inverse projection* $\text{Fx-bar}^{-1}(\text{VG})$ for VG, *i.e.* the morphologic-phonologic-syntactic-semantic restrictions on the (predicational) semantic head of VG that are necessary (and sufficient) to retrieve the VG (or VGK) *local* structure through FX-bar (direct) projection of its semantic head. The two projection functions are outlined in Figure 3.

The FX-bar inverse projection associates to VG a number of (possibly covert) *semantic heads*, corresponding to the meaning(s) of the lexical head entry, each semantic head observing the set of *sd* and *dt* functions and values, along with phonologic, lexical, morphologic, syntactic and semantic restrictions at lexical level on arguments, clitics, doubling etc. [29].

This is the starting point in the process of *text generation task*, when the first requirement is to generate one or several adequate VGs, satisfying the *planning* restrictions. For clause analysis / generation, the parsed VG (as *clause-shadow*) or the obtained VG(s) is FX-bar projected into one (or more) finite or non-finite clause(s), with its (their) arguments, constructed lexically from diathesis computations and linguistic restrictions.

2.2 Diathesis Transformations and Semantic Diathesis Functions

The definition given to the diatheses considers either the syntactic rapport between the subject and the verb complement(s), as arguments of the same predicational head category, either an ontological rapport between the action and its author, or even both realities. [3; p.87–91] distinguishes between active, passive, impersonal reflexive, and dynamical reflexive diatheses, according to the importance given by the speaker to the action presented. [19; p. 464] considers the realities between the syntactic positions (subject – verb – complement) and their semantic correspondences, (actant-process-patient). [17; p. 13–22] takes into account for the diathesis definitions, the reflection at the semantic level of

the verb of the extralinguistic rapport subject-action-object, meaning both the syntactic rapport verb-subject and the verb-complement one.

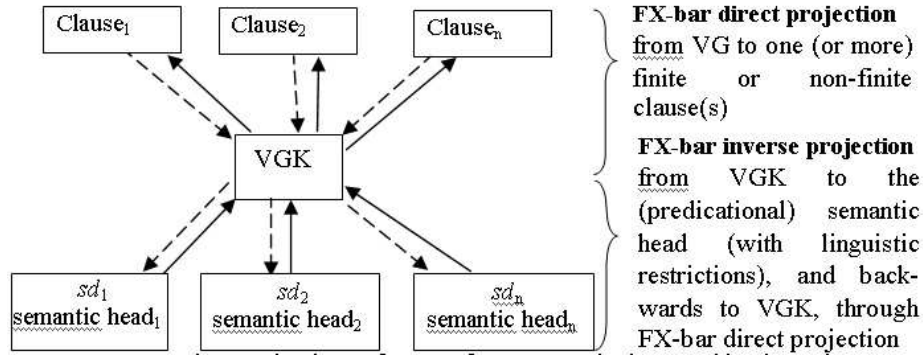


Figure 3. FX-bar projections of VGK, from text to lexicon and backwards

We try to solve the above mentioned problems of FX-bar direct and inverse projections for VG / VGK by defining *diathesis transformations* and *semantic diathesis functions*, following mainly the *semantic diatheses* (active, passive, reflexive, reciprocal, impersonal, and dynamic) developed in [20; p. 85–115]. The information sources for the projection processes are **(a)** VG / VGK parsing on one side, from which one can extract VG tense, syntactic (shallow) diathesis, (predicational or not) semantic head, clitics, quantifiers, internal and external (proper) modifiers, modalizers (Figure 3). With these elements, one moves down towards lexicon, where one should find **(b)** the second source of information: the valence (arity), type (sort, *e.g.* NG, VG, clause) of the VG semantic head, the *diathesis transformations*, and the values of the *semantic diathesis functions*. Necessary (and sufficient, if possible) constraints can be specified to ensure the uniqueness of the FX-bar projection(s), either *direct* or *inverse*, of a *semantic head*, through a VG, into a (finite) clause, and the reverse [8]. These constraints are similar to those met in *local linking algorithms* [28], including the diathesis (ARGLIST) transformations, *i.e.* VG semantic diatheses related to the TFA (Topic-Focus Articulation) ordering of clause arguments [18], [14].

The following Table 1 shows the mappings of the argument lists within the syntactic / semantic diathesis metamorphosis.

Table 1. Diathesis transformations from syntactic to semantic argument lists

SynD SemD	Active	Passive	Reflexive
Active	$dt([A1, A2, (A3)^{*}])$ $\downarrow\uparrow *$ $[A1, A2, (A3)]$	\emptyset	$dt([A1 \equiv \text{RefPron}, A2, \emptyset])$ (analysis) $\downarrow\uparrow$ (generation) $[A1, A2, \emptyset]$
Passive	\emptyset	$dt([A1, A2, (A3)])$ $\downarrow\uparrow$ $[A2, A1, (A3)]$	$dt([A1, A2 \equiv \text{RefPron}, (A3)])$ $\downarrow\uparrow$ $[X1^{***}, A1 \equiv A2, (A3)]$
Reflexive	\emptyset	\emptyset	$dt([A1, A2 \equiv A1, \emptyset])$ $\downarrow\uparrow$ $[A1, A2 \equiv A1, \emptyset]$
Reciprocal	\emptyset	\emptyset	$dt([A1 \equiv \text{RefPron}, (A2), \emptyset])$ $\downarrow\uparrow$ $A1 \equiv \{X1, X2\}$ $\{\{X1, A2, X2\},$ $\{X2, A2, X1\}\}$
Impersonal	\emptyset	\emptyset	$dt([(A1) \equiv \text{RefPron}, (A2), \emptyset])$ $\downarrow\uparrow$ $[X1, (A1), (A2)]$
Dynamic	\emptyset	\emptyset	$dt([(A1) \equiv \text{RefPron}, (A2), \emptyset])$ $\downarrow\uparrow$ $[(A1) \equiv X1, (A2), \emptyset]$

* $\downarrow\uparrow$ = analysis “ \downarrow ” and “ \uparrow ” generation tasks;

** (A_n) = argument optionally present;

***X = uninstantiated variable introduced to support semantically an argument;

The notation “ $(A1) \equiv \text{RefPron}$ ” means that the argument A1 is optionally present, the reflexive pronoun is overt (lexically present), and (clitic) doubling is possible.

The notation “ $dt([(A1) \equiv \text{RefPron}, (A2), \emptyset]) \downarrow\uparrow [X1, (A1), (A2)]$ ” means that the *semantic-diathesis* function $sd(\text{category, clitics, syntactic-diathesis, valence})$ is applied to the *Reflexive* diathesis list $[(A1) \equiv \text{RefPron}, (A2), \emptyset]$, the result being the semantic *Impersonal* diathesis list $[X1, (A1), (A2)]$ (see Table 1).

The *diathesis transformation* functions $dt(\text{List}_0) = dt_1 = \text{List}_1$ map

the list of syntactic (grammatical, shallow) arguments (corresponding to $\text{SUBCAT}_{\text{obliq_order}}$), into the list of semantic arguments corresponding to $\text{SUBCAT}_{\text{theta_order}}$. The result of transforming a syntactic diathesis into a semantic one is not a unique operation, and Table 1 gives the general dt functions, as a mapping of the *three* syntactic diatheses into the *six* semantic ones, and backwards. For a lexicon entry, the *semantic diathesis functions* take the form:

$sd(\text{category, clitics, syntactic_diathesis, valence}) = \{dt_1, dt_2, \dots, dt_n\}$, ($n = 1 \div 6$), where $dt_1 = dt(\text{List}_0)$, $dt_2 = dt(\text{List}_1)$, $dt_3 = dt(\text{List}_2)$, ..., accordingly to the lexical semantics meanings (*readings*) derived from the VG head category and additional information resulted from the VG parsing.

2.3 Diathesis Computing within FX-bar Projections of VG¹

As already mentioned in subsection 2.2, the semantic diathesis function is defined as $sd(\text{category, clitics, syntactic_diathesis, valence}) = \{dt_1, dt_2, \dots, dt_n\}$, $n = 1 \div 6$.

Using the verb *a-se-uita* (*to look at*) as example, the computation of dt and sd function values is realized in the following steps, derived from the operation sequence of FX-bar direct and inverse projections:

Step1. Extracting an *a-se-uita* derived VGK from a *concrete* but *arbitrary* clause that encloses it;

Step2. Handing down to the lexicon, with the semantic head of that VGK;

Step3. Computing the sd and the dt function values;

Step4. Retrieval of the same VGK as FX-bar projection of (one of the meanings of) *a-se-uita* semantic head, associated with the semantic diathesis computed values of dt and sd functions;

Step5. FX-bar projection of the VG into the n possible clause types, n corresponding to the *number* of (diathesis transformation) dt functions.

After choosing a VGK from an arbitrary clause in the text, VGK is completely parsed, being obtained the VG extracted semantic head,

¹The analysis of this subsection may also be found in [RevRoum06]

tense, syntactic diathesis, clitics etc. The next move is to step down with the VGK semantic head at the lexicon level, where the semantic diathesis transformations and restrictions are located. In our case, the following *sd* function value:

$$sd(se-uita, ReflPron_se, reflexive, 2) = \{active, reflexive, impersonal\} = \{dt_1, dt_2, dt_3\}$$

has to be found, meaning that the reflexive syntactic diathesis of *a-se-uita* (*to-look-at*) can be translated into the *active*, *reflexive*, and *impersonal* semantic diatheses. From the above *sd* values, using Table 1, one can compute the following values of *dt* functions:

$$\begin{aligned} dt_1(reflexive) &= active \Leftrightarrow dt_1([A1 \equiv ReflPron, A2, \emptyset]) \downarrow \uparrow [A1, A2, \emptyset] \\ dt_2(reflexive) &= reflexive \Leftrightarrow dt_2([A1, A2 \equiv A1, \emptyset]) \downarrow \uparrow [A1, A2 \equiv A1, \emptyset] \\ dt_3(reflexive) &= impersonal \Leftrightarrow dt_3([(A1) \equiv ReflPron, A2, \emptyset]) \downarrow \uparrow [X1, \\ &\quad (A1), (A2)] \end{aligned}$$

Since the valence of *a-se-uita* is 2, the resulted lists are reduced from 3 to 2 elements, the final value of *sd* being:

$$sd(se-uita, ReflPron_se, reflexive, 2) = \{[A1, A2], [A1, A2 \equiv A1], [X1, (A1)]\}$$

Due to different semantic diatheses, clause types with distinct readings are potentially parsed (in analysis task) or produced (in generation task). The following examples show the non-uniqueness for the *sd* function values at the lexical semantics level.

(1)(R) **Se uită** la fratele lui. (*sem. diathesis* = *active*)

(E) *He looks at his brother.*

(2)(R) **Se uită** în fața televizorului ore în șir.

(*sem. diath.* = *reflexive*)

(E) *He forgets himself in front of the TV.*

(3)(R) **Se uită** deseori semnificația zilei de 24 ianuarie.

(*sem. diath.* = *impersonal*)

(E) *The significance of 24 January is often forgotten.*

For a complete treatment of the verb *a-uita*, we describe hereafter the non-reflexive counterpart of its lexicon entry. The *sd* and *dt* functions may have, for instance, (some of) the following values for the (non-reflexive) *a-uita* (*to-forget*) entry:

$$sd(uita, Acc_Clitic, active, 2) = \{active\}$$

$sd(uita, \emptyset, \text{passive}, 2) = \{\text{passive}\}$.

The feature of argument optionality is transferred from larger to smaller number of arguments. These *sd* and *dt* computed values can be lexicalized in clause constructions like:

(4)(R) *lon a uitat-o pe Maria.* (*sem. diath.* = active)

(E) *John has forgotten Mary.*

(5)(R) *Geanta a fost uitată de lon.* (*sem. diath.* = passive)

(E) *The bag was forgotten by John.*

The mechanisms of computing syntactic and semantic diatheses on grammatical structures (from clause to VGK and its lexical semantic head – and backwards) viewed as FX-bar (direct or inverse) projections, and involving as essential incorporated element the *predicational* feature they bear or inherit, substantiate our attempt of taking apart the machinery and anatomy of linguistic *predication*.

The predication was handed down from the classical, syntactic level to the lexical, thus lexicon level, using mechanisms developed within FX-bar theory. The next Section 3 takes advantage of the lexical-level predication and FX-bar projections to propose a unified treatment of the (Romanian) predicate, either verbal or nominal, and to establish a consistent relationship between *predication* and *predicate* within the framework of *intensional / extensional logic*. Local and global sentence / discourse parsing, machine translation, and FramNet thematic roles assigning [27] are natural applications of the present approach.

3 Syntactic Substructures of the Romanian Predicate

3.1 VSGs of the Verbal Predicate

A. The Tense Auxiliary SubGroup

The most frequent and natural VSG is the Tense Auxiliary SubGroup (TASG): *voi fi, aş fi, am fi, sunt*. TASG is a non-saturated VSG that needs as semantic head a noun (N), adjective (Adj), or non-finite verb (V) form. For V finite forms, TASG is considered to be enclosed,

by default, within the synthetic inflection of the verb. TASG may receive the usual VG ingredients: *special adverbs*, *negation*. In its *basic (bare) form*, TASG cannot receive pronominal clitics (one can have “*i l-aş fi*” only in the presence of the 3-valued valence verbal form “*dat*”). TASG is non-predicational since its inner semantic head (in this priority order of occurrence) is either a copulative verb, passive auxiliary, or tense auxiliary. We shall use here TASG with the meaning of an (individual) assignment $x := y$, corresponding to the copulative relation between the terms x and y (as *intensive* or *extensive* variables or constants), in order to describe the intensional / extensional representations of the verbal and nominal predicate.

TASG is a natural *verbal subgroup* (VSG) of the Passive Tense Auxiliary SG (PTASG). (P)TASGs may have the same meaning of copular (assignment) relation, do receive special adverbs and negations, but they do not bear pronominal clitics (in their *basic form*). TASG and PTASG are both non-saturated, *i.e.* they need to receive a semantic head category, which is a *non-finite verb*, an *adjective* or a *noun*. Utterances such as “*Am fost.*” or “*N-am putut.*” (for modals) are verbal anaphora or anaphoric predicates.

PTASG may also contain variants of passive auxiliaries, conditional, passive conditional, etc. VSG types of TASGs.

B. The Modal Verb Subgroup

Another VG substructure refers to the Modal VSG (ModVSG). ModVSG derives normally from TASG, (not PTASG since one can not have “*s-ar fi fost trebuit [putut]*”), whose semantic head is *a-putea*, *a-trebui*. One may have “*[eu] pot; ar fi putut; s-ar fi putut; n-am fi trebuit*”, with possible insertions of *special adverbs* and *negations*, but not pronominal clitics (in their bare forms). The modals (*a-putea*, *a-trebui*) are predicational polymorphic verbs (*i.e.* whose predication orders or valences may be distinct for different senses). Thus, they may have as semantic arguments either predications of first or second order, saturated (clauses) or non-saturated (VGs) categories, or extensional categories.

Examples.3.1.1. (a) *Am fi putut alerga.* (b) *Se poate că lumina*

l-a speriat. (c) Îi trebuie apă [ca] să crească. (d) N-ar trebui să putem cheltui toți banii. (e) Ar fi trebuit să nu mai poată trece granița.

One may have various types of arguments for the modal verb head of the ModVSG, including (modal) recursion on its syntactic development. Once again, these VSGs receive the usual lexical insertions.

C. Special cases (whose analysis deserves a more detailed discussion, postponed here): **(a)** *Îi este ușor.* **(b)** *Îi voi fi coleg lui Ion.* **(c)** *Maria este colegă de clasă cu Ion.* **(d)** *Maria i-ar fi putut fi mamă loanei.* Hierarchical relations (*e.g. mother_of, colleague_of*) may be triggered by relational (but still non-predicational) nouns. Notice once again that, in their basic form, (P)TASGs and ModSVGs are non-saturated, polymorphic, and do not receive object (pronominal) clitics.

D. However, **the outcomes** of applying the basic VG substructures (P)TASG and ModVSG to their semantic heads may receive pronominal clitics according to the corresponding head valences. Two remarks:

(D1) When the head is a non-finite VSG, the potential clitics corresponding to the head valence are attached to the corresponding ModVSG.

(D2) When VSG has a finite head, the latter embodies its own clitics. From these observations, an important question is derived: utterances such as

- (a)** *Aș fi putut eu să i-o împrumut.*
- (b)** *Nu trebuia [ca] Maria să-l fi citit.*

are *biclausal* (as supported in [16]) or *monoclausal* constructions (as defended in [25])? A detailed analysis to decide on this question should be necessary. It is not essential whether the “modal clause” is applied to a finite VG or to another clause, but to decide if utterances such as **(c)** “*?Aș fi putut.*” or **(d)** “*?Nu trebuia.*” are truly finite clauses, while **(e)** “*Nu-i trebuia.*” is definitely finite. Utterances (c) and (d) in the previous sentence are *second-order non-saturated* predications, *i.e.* receive as their argument a finite clause. The problem is: can this type of predications be assimilated to a finite clause?

Formally, the functional representation is clear: one has $\text{ModVSG}(\text{Finite_Clause})$ or $\text{ModVSG}(\text{Non_Finite_VSG})$. The linguistic interpretation as monoclausal or biclausal is controversial. Furthermore, we can have *multimodal utterances* such as: (f) *Trebuie să poată să învețe*. (g) *Trebuie să poată învăța*. (h) *Se poate că Ion nu trebuia să vină aici*. We bet here on the monoclausal approach.

E. We definitely agreed on the necessity of completing the novel shape of the FX-bar scheme in Fig. 2 by representing the weak pronouns (clitics) or their *traces* (when covert) corresponding to the valence-based (direct) arguments of the VG predication (semantic) head. Each VG semantic head should receive, either in overt or covert forms, adequate morphologic-syntactic devices to realize (local) anaphoric agreement with its valence-based arguments. For Romanian (and other syntactically similar languages), the classical inflection of the head verb and its (diathesis-free) agreement with the grammatical subject, as well as the (lexical or virtual) presence of the clitics semantically attached to the same VG semantic head, express exactly this linguistic reality. These clitics may naturally be viewed as multiple “inflections”, valence-commanded, of the same VG semantic head. When covert, these valence-driven extended inflections of the VG semantic head (*viz.* the proper inflection of the head verb and the VG clitics) behave as veritable “linking” devices [28], including the local anaphoric binding, case marking etc. When covert (the head verb inflection is always covert), the clitics should receive the same bundle of linguistic features as they bear when overt.

Another problem to be solved (for Romanian, at least) is the distribution of the VG *linking indices* (or VG inside clitics, see also Section 2), either they are overt or covert). More precisely, the problem is to establish under which Verbal SubGroups (VSGs) of the VG the clitics are distributed (among which we include, once again, the inflection corresponding to the VG semantic head, accorded with the grammatical subject of the same clause) and, equally important, under what syntactic-semantic constraints. Interesting examples concerning this problem on the syntax-phonology interface are given in [25 :Chap.5, 244, Ex. (426)a.b.c.].

3.2 An Intensional / Extensional Modelling of the Verbal and Nominal Predicate

Let us consider the following *series* of predicates (we do not specify whether verbal or nominal ones).

Example 3.2.A. *a fost predată*

This VG is a *verbal predicate* in *passive diathesis* and *past tense*. “*predată*”, the head to which is applied the TASG, is a (non-finite) *intensional predicate* of valence 3. *E.g.*, the intensional representation of “*Lucrarea a fost predată.*” could be $lucrarea(Y) :=_{past} predată_{passive}(x, Y, z)$, where Y is an extensive variable, while x and z are extensional predicates. Here we take the extensional (context-depending) meaning of *lucrarea*, but it may also have an intensional (predicational) sense: *Lucrarea cu mîgală a pereților exteriori de către meșterii populari...*

Example 3.2.B. *a fost plecată*

This VG is a *nominal predicate* in the classical grammar. However, it may also be seen as a *verbal predicate* whose semantic head is a predication represented by a *non-transitive* (valence = 1) non-finite verb. Such a category has the representation $plecată(x(X))$, where x is the extensional predicate, and X is the extensional variable.

Example 3.2.C. *a fost frumoasă*

This is a clear *nominal predicate*, whose semantic head *frumoasă* is no more a predicational (intensional) category. Since any adjective, predicational or not, requires (at least one) *nominal* argument, written as the *extensional* predicate $x(X)$, the correct representation is $frumoasă(x(X))$, with x an extensional predicate and X an extensional variable.

Example 3.2.D. *a fost elevă*

This is a (classical) nominal predicate; the semantic head is representing the extensional predicate $elevă(X)$.

Example 3.2.E. *va fi trădarea*

This is also a *nominal predicate*, consisting of a TASG whose semantic head *predarea* is a *predicational noun* [8] (thus non-finite) category. The intensional representation is $P :=_{future-pres} trădarea(x, y)$, where

P is an *intensional variable* corresponding to the intensional predicate *trădarea*, and x and y are *extensional predicates* corresponding to the 2-valence predication (albeit) nominal category *trădarea*. For instance, P could represent the demonstrative pronoun (and *intensional anaphora*, as it follows) *Aceasta*, in the variant example: *Aceasta a fost predarea*.

Recapitulating this series of examples, one may see the structure of the *predicate* as a TASG applied to a verbal or nominal phrase whose *non-finite* heads vary as follows:

- *predată* = predication (intensional) V, valence-3, non-saturated; (Example 3.2.A)
- *plecată* = predication (intensional) V, valence-1 (non-transitive), non-saturated; (Example 3.2.B)
- *frumoasă* = non-predication (extensional) A, non-saturated, requiring an (extensional predicate) nominal head; (Example 3.2.C)
- *elevă* = non-predication (extensional) N, saturated; (Example 3.2.D)
- *trădarea* = predication (intensional) N, 2-valence, non-saturated; (Example 3.2.E)

3.3 A Smooth Transition from Nominal to Verbal Predicate

Examples 3.3. a. *Lucrarea a fost predată.* is represented as $lucrarea(Y) :=_{past} predată_{passive}(x, Y, z)$, where Y is an extensive variable, while x and z are intensive variables (Example 3.2.A).

b. *Ioana a fost plecată.* is represented as $A :=_{past; act\ diat} plecată(x(A))$, where A is an extensive constant (*Ioana*) and x is an extensional predicate.

c. *Ioana a fost frumoasă.* is represented as $A :=_{past; act\ diat} frumoasă(x(A))$, where x is an extensional predicate as nominal head, and A is an extensive constant (*Ioana*).

d. *Eleva a fost frumoasă*. is represented as $eleva(X) :=_{past; act\ diat} frumoasă(x(X))$, where *eleva* and *frumoasă* are extensional predicates and X is an extensional variable.

e. *Eleva frumoasă este studentă*. is represented as $frumoasă(eleva(X)) :=_{pres} studentă(X)$.

The **transition point** from *the verbal predicate* to *the nominal predicate* is located at Examples 3.2.B and 3.2.C (or 3.3.b and 3.3.c). As one can easily see, these predicates, although called verbal and nominal predicates, provide *the same intensional / extensional representations*, from different reasons: *plecată* is a predication-intensional category but with a *single extensional* argument (being non-transitive), while *frumoasă*, is a non-predicational thus extensional category, being non-saturated, and requires an extensional predicate as its nominal head. These predicates do not provide, from natural causes, passive diathesis, which is an exclusive (possible) attribute of predication categories with valence greater or equal than 2.

As in the case of FX-bar inverse projection of VG, from VGK towards (one of) its semantic heads [9], this function can not be specified without computing the VG semantic diathesis [12], relying basically on valence-arity and type-sort information of the semantic head (thus more than the simple presence of the predication feature [8]). The same observation, as proved here, is true for distinguishing between *verbal* and *nominal* predicates, having the same lexical semantics based on intensional / extensional representations.

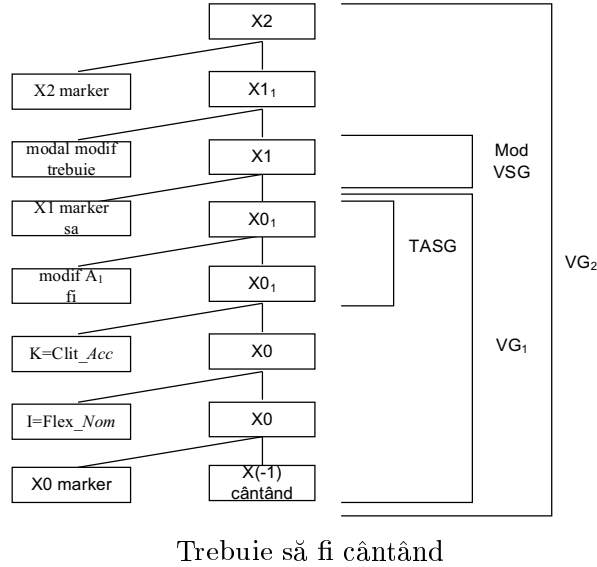
Furthermore, our approach provides an unitary taxonomy of the verbal and nominal predicate, based on intensional logic. Both constructions rely on (basic/applied) verbal subgroups (VSGs) that a VG is decomposed in, with tense auxiliary, copulative, modal, semi-auxiliary, restructuring head verbs. VSG substructures are (recursively) composing one another, using the predication feature (including valence / sort of arguments, if necessary) and polymorphism of their semantic heads, to obtain complex FX-bar projections representing the VG.

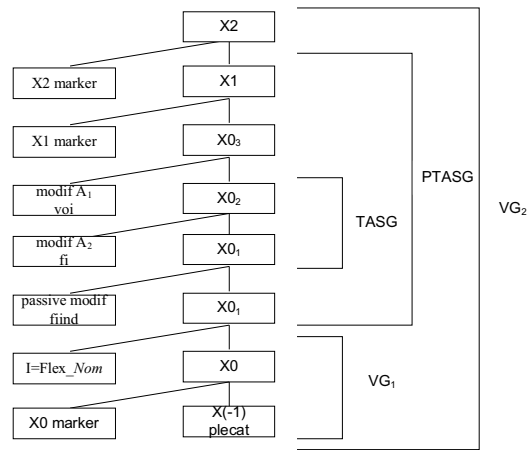
Although there is still a lot of work to be done for complete FX-bar scheme characterizations of various classes of (Romanian) verbs,

we designed here and in [9] detailed solutions that constitute a detailed solutions to VG analysis. The keypoint for the local, clause-level syntactic structures relies on the predicational feature and the newly defined *lexical predications* attached to the VG semantic head (see subsections 1.1, 1.2, and section 2), within the framework of FX-bar theory and lexical semantics – intensional logic formalism.

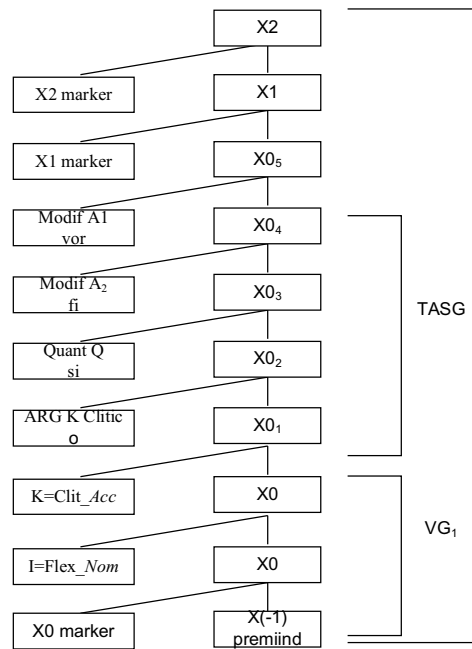
3.4 Examples of FX-bar Schemes Applied to VGs

In this subsection we expose several (linearized) FX-bar schemes (Figure 4), derived from the general FX-bar scheme (see Figure 2 above and [10], [11]), that mimics the decomposition model of the involved VSGs of the VG, described in previous subsections.

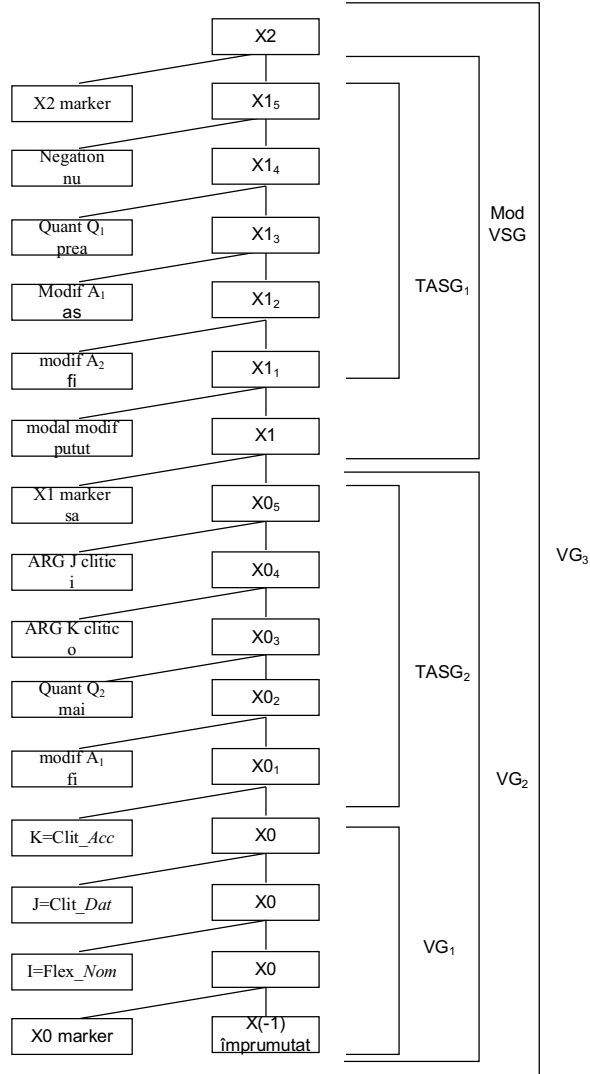




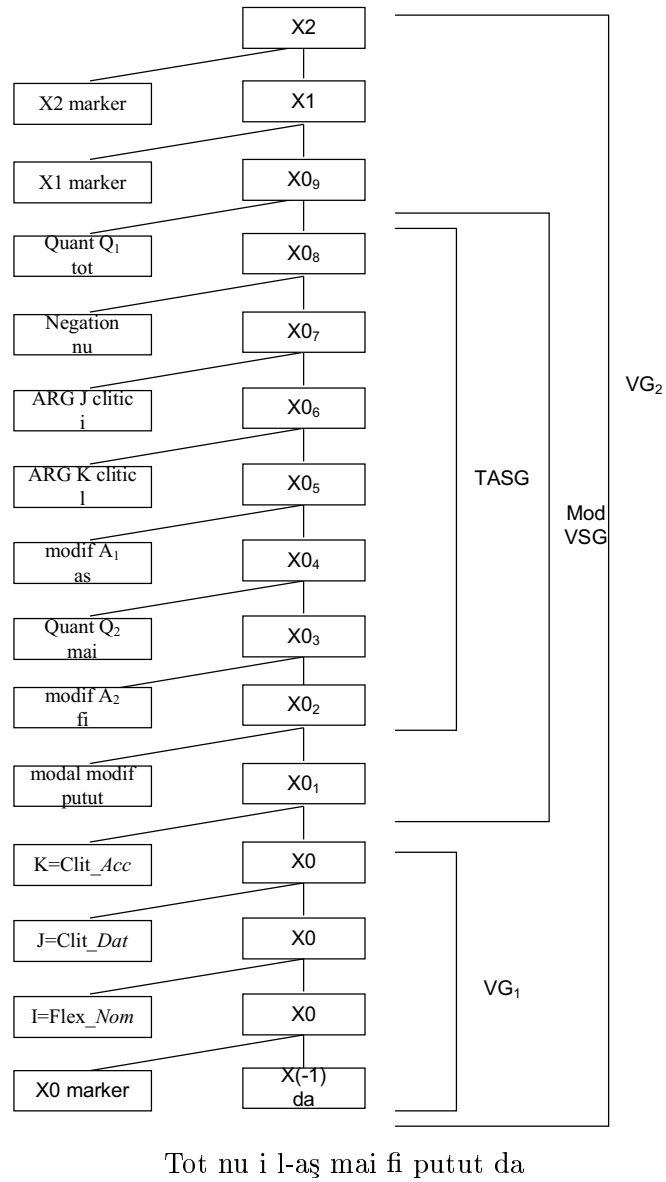
Voi fi fiind plecat

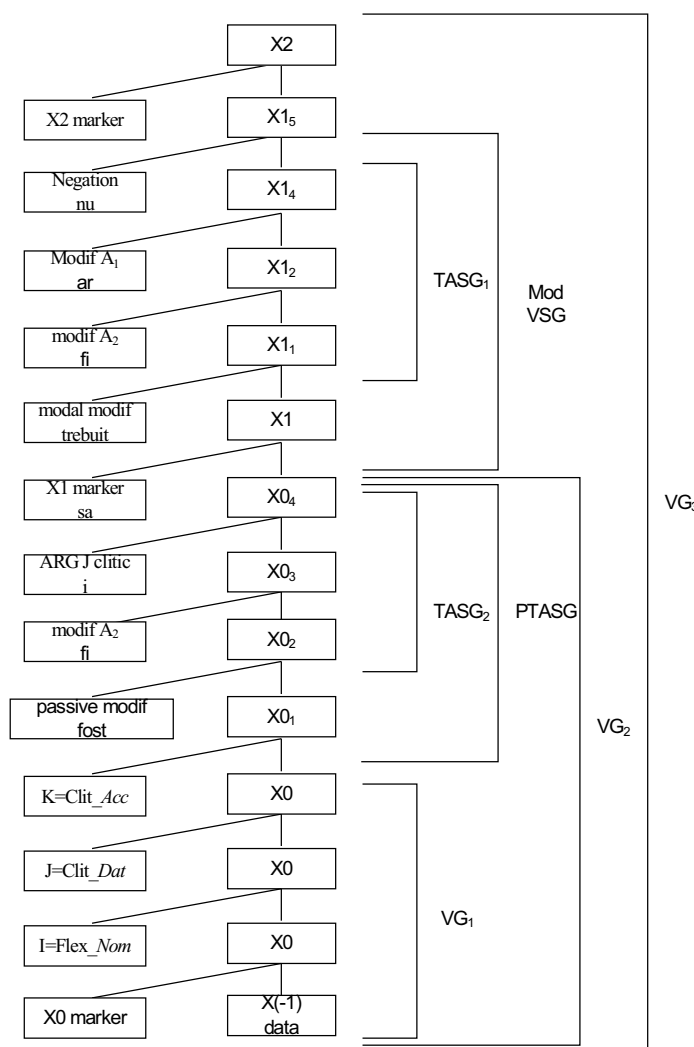


Vor fi și premiind-o



Nu prea aş fi putut să i-o mai fi împrumutat





Nu ar fi trebuit să-i fi fost dată

Figure 4. Examples of linearized FX-bar schemes for VGs

4 Conclusions

There is still a large quantity of linguistic data, some of them with subtle variations, to be analyzed as interesting for VG substructures, *i.e.* VSGs. In terms of FX-bar projections (Section 2), our aim is to reveal and classify as (unitary) VGs the categories and sorts of *verbal and nominal predicates*, looking for consistent VSGs (when they exist) and all the FX-bar projections as intermediate syntactic-semantic layers situated between the finite clause, its VG predicate (which includes the lexical or virtual clitics as VG *linking* indices), and the VG semantic (predicational) head. The significance of such an analysis should be remarkable: clearing up the *regime of predication*, the status of *verbal and nominal predicate* (as VGs), the structure and role of VSGs as verbal operators successively composing to re-construct the VG, thus the configuration of *local* (clause-level) and *global* (discourse segment) text structures. As current and future research topics, we consider that the present results on classical syntax of VG can provide the development basis for Topic-Focus Articulation (TFA) [15] and linking algorithms [28] at the clause / sentence level, to reveal the *information structure* (IS) syntax at the global (inter-clausal and discursive) level and the configuration of the syntax-prosody interface for Romanian [14], [15].

References

- [1] Barbu, Ana-Maria (1999): *The Verbal Complex*. Studii si Cercetari Lingvistice, L, no.1, Bucuresti, pp. 39–84 (In Romanian).
- [2] Barbu, Ana-Maria; Emil Ionescu (1996): *Contemporary grammatical theories: grammars of the phrasal head*. in Limba română, no. 45, (1-6) pp. 31–55 (in Romanian).
- [3] Combiescu, Mariana (1968): *On the definition of the Diatheses*. Limba Română, XVII, no. 1, pp. 87–91, (in Romanian).
- [4] Curteanu, Neculai (1988): *Augmented X-bar Schemes*. COLING'88 Proceedings, Budapest, pp. 130–132.

- [5] Curteanu, Neculai (2000): *Towards a Functional X-bar Theory*, Technical Report, Institute for Computer Science, Romanian Academy, Iaşi Branch, 32p.
- [6] Curteanu, Neculai (2002): *Elements of a Functional X-bar Theory Within the SCD Linguistic Strategy*, ECIT2002 Conference, Iaşi, România.
- [7] Curteanu, Neculai (2003): *Towards a Functional X-bar Theory*. In the volume “The Romanian Language in the Informational Society”, Dan Tufiş, F. Filip (Eds.), Edited by the Romanian Academy, Research Institute for Artificial Intelligence, Bucharest, pp. 51–86 (in Romanian).
- [8] Curteanu, Neculai (2003-2004): *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).
- [9] Curteanu, Neculai (2005): *Functional FX-bar Theory Extended to Discourse (Rhetorical) Structures*. In ‘Intelligent Systems’ Conference Volume, H.-N. Teodorescu *et al.* (Editors), Performantica Press, Iaşi (Romania), pp. 169–182.
- [10] Curteanu, Neculai; E. Zlavog; C. Bolea (2005): *Sentence-Based and Discourse Segmentation / Parsing with SCD Linguistic Strategy*. In ‘Intelligent Systems’ Conference Volume, H.-N. Teodorescu *et al.* (Editors), Performantica Press, Iaşi (Romania), pp. 153–168.
- [11] Curteanu, Neculai (2006): *Local and Global Parsing with Functional (F)X-bar Theory and SCD Linguistic Strategy*. (I.+II.), Computer Science Journal of Moldova, Academy of Science of Moldova, Vol. 14 no. 1 (40):74-102, Vol. 14, no. 2 (41):155–182.
- [12] Curteanu, Neculai; Diana Trandab (2006): *Functional (F)X-bar Projections for Local and Global Text Structures. The Anatomy of Predication*. Revue Roumaine de Linguistique, Romanian Academy Editorial House, Bucharest, No. 4 (to be published).
- [13] Curteanu, N; D. Trandab; M. Moruz (2006): *Substructures of the (Romanian) Predicate and Predication Using FX-bar Projection Functions on the Syntactic Interface*, (H.N. Teodorescu, Ed.)

- ECIT-2006, 4th European Conference on Intelligent Systems and Technologies, Iași, Romania.
- [14] Curteanu, N.; D. Trandabăț; M.A. Moruz (2007a): *Syntax-Prosody Interface for Romanian within Information Structure Theories*. (C. Burileanu and H.N. Teodorescu Eds.) Proceedings of SpeD-2007 Conference (Speech Technology and Human - Computer Dialogue), Romanian Academy Editorial House, Bucharest, pp. 217–228.
 - [15] Curteanu, N.; D. Trandabăț; and M.A. Moruz (2007b): *Topic-Focus Articulation Algorithm on the Syntax-Prosody Interface of Romanian*. Proceedings of TSD-2007 (Text, Speech and Dialogue) Conference, Plzen, Czech Republic, Springer Verlag, Berlin, LNCS / LNAI Series (to be published).
 - [16] Dobrovie-Sorin, Carmen (1994): *The syntax of Romanian. Comparative Studies*. Berlin: Mouton de Gruyter.
 - [17] Graur, Alexandru (1969): *Diatheses*. Studii si cercetari lingvistice, Vol. XX, no. 1, p. 13–22, (in Romanian).
 - [18] Hajiová, E., H. Skoumalova; P. Sgall (1995): *An Automatic Procedure for Topic-Focus Identification*, Computational Linguistics, 21 (1), 81–94
 - [19] Iordan, Iorgu; Vl. Robu (1978): *Contemporary Romanian Language*. Bucharest, Didactic and Pedagogic publishing house, p. 464, (in Romanian).
 - [20] Irimia, Dumitru (1997): *The Morphosyntax of the Romanian Verb*. The Editorial House of the “Al. I. Cuza” Iași University (in Romanian).
 - [21] Legendre, Géraldine (1999): *Optimal Romanian clitics: a cross-linguistic perspective*. In: V. Motapanyane (Ed.) Comparative Studies in Romanian Syntax. HAG, The Hague.
 - [22] Mann, William, Sandra Thompson (1988): *Rhetorical Structure Theory: A Theory of Text Organization*. Research Report RS-87-190, Information Sciences Institute, University of Southern California, Marina del Rey, California, 80 pp.

- [23] Monachesi, Paola (1998): *The Morphosyntax of Romanian Cliticization*. In: P.-A. Coppen *et al.* (Eds.), Proceedings of Computational Linguistics in The Netherlands 1997, pp. 99–118, Amsterdam-Atlanta: Rodopi.
- [24] Monachesi, Paola (2000): *Clitic placement in the Romanian verbal complex*. In: B. Gerlach and J. Grijzenhout (eds.) Clitics in phonology, morphology, and syntax. Linguistik Aktuell. John Benjamins. Amsterdam.
- [25] Monachesi, Paola (2005): *The Verbal Complex in Romance. A Case Study in Grammatical Interfaces*. Oxford University Press, Oxford Studies in Theoretical Linguistics.
- [26] Pollard, Carl; Ivan Sag (1994): *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, Chicago & London.
- [27] Trandabăţ, Diana; Maria Husarciuc; Monica Lupu (2005): *Towards an automatic import of English FrameNet frames into the Romanian language*, 1st ROMANCE FrameNet Workshop, held at EUROLAN 2005, Cluj-Napoca, Romania, pp. 28–36.
- [28] Van Valin, Robert D. Jr. (2005): *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
- [29] Zwicky, A. (1985): *Clitics and Particles*. Language, Vol. 62, No. 2, pp. 283–305.

Neculai Curteanu, Diana Trandabăţ, Mihai Alex Moruz, Received June 5, 2007

Institute for Computer Sciences,
Romanian Academy, Iaşi branch
B-dul Carol I, nr. 22A, 6600 IAŞI, ROMANIA
E-mail: {curteanu, dtrandabat, mmoruz}@iit.tuiasi.ro