FUNCTIONAL FX-BAR PROJECTIONS FOR LOCAL AND GLOBAL TEXT STRUCTURES. THE ANATOMY OF PREDICATION

NECULAI CURTEANU, DIANA TRANDABĂȚ

Abstract. This paper proposes and discusses issues on local and global text structures, all of them being connected to a **lexical** concept of **predication**. The main contributions of the present work comprise: **(a)** A novel functional X-bar (FX-bar) scheme is advised, aiming to reveal, model and relate the local, clause-level markers and text structures. **(b)** At global level, two FX-bar schemes are proposed, one pursuing the inter-clause level relations, and the other being of discursive, rhetorical nature. **(c)** There are described local and global classes of markers, together with their graph-based hierarchy, to be incorporated on the projection levels of FX-bar schemes and within SCD (Segmentation-Cohesion-Dependency) linguistic strategy algorithms. **(d)** The concept of functional generativity is discussed, with implications on parsing algorithm classification and FX-bar projection mechanism. **(e)** Local FX-bar projection functions have at their core the notion of **lexical predication**. Direct (towards clause) and inverse (towards lexicon) FX-bar projections of the verbal group (verbal complex) are shown to be efficient tools for a better understanding of the structure and functioning of the Romanian predication and predicate, and for supporting our challenging proposal of handing down the predication from the classical, syntactic level to the lexical one. **(f)** Finally, direct and inverse global FX-bar projections mediate between larger text spans and inter-clause vs. discourse trees, the intricate relationship between the finite clause and (sub-clause and multi-clause) discourse segment being highlighted.

1. INTRODUCTION

1.1. From Augmented to Functional X-bar (FX-bar) Schemes

In Curteanu (1988), in the context of the *Segmentation-Cohesion-Dependency* (SCD) *linguistic* (*parsing*) *strategy* (beginning with Curteanu 1983, 1990), there is defined a class of *augmented* X-*bar* (AX-bar) *schemes* intended to represent general syntactic *invariants* for the grammatical structures of NL, particularly for Romanian, as a solution to the problems raised by the automatic analysis and generation of NL text.

RRL, **LII**, *1–2*, p. 161–194, Bucureşti, 2007

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

Curteanu (2000, 2002, 2003) introduced *functional* X-*bar* (FX-*bar*) *schemes*, which complete and substantiate theoretically the *augmented* X-*bar* (AX-bar) *schemes* of Curteanu (1988), extended in Curteanu (2005) to *global* (discourse) *structures*, in the framework of SCD parsing strategy. The following points are aimed: **(i)** The proposed FX-bar general schemes stand for a theoretical submodule within the SCD linguistic strategy. Various versions of X-bar theory are structurally promoted within well-known linguistic theories, *e.g.* Chomsky (1995), Pollard and Sag (1994), Bouchard (1995) etc. **(ii)** The general FX-bar schemes should be seen as an uniform and unitary tool for ***representing*** the *local and global text structures*, incorporating (as much as possible) the *functional* and *relational properties* of language, and providing also a ***computational device*** when marker functions / relations, word lexical semantics and their dependency relationships are (partially) specified as in Curteanu (1994), Curteanu, Holban (1996), Curteanu *et al.* (2003, 2004, 2005).

1.2. Basic Assumptions

SCD considers four *major lexical categories* (and their functional projections within the FX-bar theory): *the Noun* (N) and *the Verb* (V) are the only lexical categories that have their own lexical (non-referential) meaning, and they are also saturated (representing their own semantic heads). Two other lexical categories play a central role in the syntactic organization of the *functional* X-bar (FX-bar) general schemes Curteanu (2002, 2005).

The Adjective (Adj) has its own (auto-semantic) meaning but it is not a saturated lexical category, since it represents a modifier function to be applied to its intrinsic referentially nominal category, *i.e.* Adj is a modifier function that requires an N-type argument head. The pronominal adjective has a similar interpretation.

The Adverb (Adv) plays the role of V modifier, role similar to the one of Adj *vs*. N. Often we denoted the modifier categories of Adj and Adv simply by A. It is important that this category is not confounded with the notation of A (Argument) positions (or A-bar, for non A-positions, a common representation in classical linguistic theories). A special question is whether there exist properly *predicational adverbs*, as adjectives do ("*predicational*" feature in the sense of Curteanu (2003-2004), often called *deverbal* property). It seems (at least for Romanian) that such adverbs do not exist properly. The first and most feasible explanation would be that the two predicational features of the verb and adverb would interfere, being too 'close' to each other. This is not a completely satisfying justification since there may coexist naturally both predicational noun and its predicational adjective pair!

These *four major* lexical categories are important because they may be endowed with *two* essential lexical-semantics features of the *local* (*i.e. clause* level,

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

which is also the *predicational* level) syntactic-semantic structures of language organization: *Tense* and *Predication* features.

The feature *TENSe* (*Time*) may receive at the lexical (syntactic) or phrase (analytical) level the values *FINIte* or *NonFINite* as well as various analytical combined values of tense and aspect for the temporal forms of the *verbal complex* (see Monachesi 1988, 2005, Barbu 1999, Legendre 1999, Curteanu 2006). The *FINIte* value of the feature *TENSe*, for each of the *four* major (lexical and) syntactic categories, is borne at (or inherited from) the lexical level by the verbal complex (to what traditionally is called *predicate*). For the structure of Verbal Complex, as in Monachesi (2005) and Barbu (1999), we shall continue to use the term "Verb Group" (abbreviated VG), in order to remain consistent with the notions, theoretical and computational approach of the functional FX-bar theory and SCD linguistic strategy. Both correspond, in a great measure, to the concept of *verbal predicate* in classical grammar.

V is the only *chosen* category for which the feature *TENSe* may receive its value *FINIte*. The other major categories N and A (Adj, Adv) receive a-priori the value *NonFINite*. These values of the feature *TENSe* involve the construction of the local syntactic structures: the Noun Group (NG), which is the classical NP with a single nominal head, VG (the Verbal Complex, already referred), and the finite and non-finite clauses.

Some few words about the *preposition* (P) category: P may play multiple roles, depending on the major category to which it is attached. It is also language-dependent. For instance, in Romanian (French, English etc.), P is mainly a case-marker or a theta-marker when attached to a nominal head. P may also play a relational role (called "predicational preposition" in Sag and Wasow 1999), *e.g.* "*on*", "*about*" (English), "*asupra*", "*despre*" (Romanian). When attached to V*s* or VG*s*, pre-positions or post-positions may change the meaning of the verb head (e.g. "*to look for*" in English, "*darstellen*" in German etc.). In all these situations, when represented in a functional model, P is not an argument but a function or an operator *P*, of one or several variables: the classical PP is either *P*(NP), when P is a case- or theta-marker, or $P(NP_1, NP_2,\ldots)$ when P is a particle introducing a structure relation. In the former situation, since P is recursively embedded into NP, one may write $NP_2 = P(NP_1)$, or even NP = *P*(NP). Using the notations XG, X = N, V, A above, we have similarly $NG_2 = P(NG_1)$, or even NG = *P*(NG) (see Section 2 for the FX-bar theory). Many other premises and principles are taken into consideration in order to support the settings and functioning of FX-bar schemes (see Curteanu 2003, 2003-2004, 2005, Curteanu *et al.* 2003, 2004).

1.3. Handing Down the Predication from Syntax to Lexis

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

The feature that we called *Predicationality*, 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 the notion and its syntactic-semantic consequences, see Curteanu (2003-2004). We avoid the term *deverbality* because its meaning is *not necessarily specific* to V*s* since this essential lexical feature is *equally shared* by V*s*, N*s* and A*s*. 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 V*s*, N*s* (often called *nominalizations*), and A*s*, whose meaning involves a *process* event or *process* name. We abbreviated this feature as PRED(dication)F(eature), with two main values, PROC(cess) and STAT(e) (or EXIST).

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 an 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 N*s* and modifiers A*s* that are (in term of lexical semantics) siblings of the corresponding predicational verbs, these non-verbal categories having a *similar syntactic distribution* of 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 (verbal group, or verbal complex) 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* …".

This is a reason to 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, thus the process meaning depends on its contextual use. For instance, the noun "*building*" in languages like 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

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

[PREDF –] (or STAT, or EXIST, or simply NPREDF values, see also Curteanu (2003, 2005)).

2. FX-BAR SCHEMES FOR LOCAL AND GLOBAL TEXT STRUCTURES

2.1. Clause-Level Local Structures

We pointed out within the SCD (Segmentation-Cohesion-Dependency) linguistic strategy that the natural language (NL) text is constructed from *local* and *global structures*. We consider *local structures* as those structures that build a single finite-clause or a single (finite or non-finite) lexical predication (including both), in sum, *finite* or *non-finite sub-clause* and *clause-level structures*.

Thus a *local structure* is one of the following FX-bar structures: **(a)** single- (or multiple-) head *noun phrase*, together with its (their) FX-bar linguistic projection(s) (the single-headed noun phrase is called *noun group* NG in SCD); **(b)** single- (or multiple-) head *adjective phrase*, with its (their) FX-bar linguistic projection(s); **(c)** *finite verbal group*, as defined in Curteanu (1988, 1994), known also under the label of *verbal complex*, Monachesi (1998, 2000) and Barbu (1999), with its FX-bar projection elements (corresponding to what is also known to be, in Irimia (1997), the *verbal predicate*, either the *synthetic* or *analytic* one); **(d)** *non-finite* VG, whose head is a non-finite V, bearing or not the predicational (deverbal) feature, and whose FX-bar projection is similar to that of the finite VG (verbal complex); **(e)** *finite clause*, viewed as the FX-bar projection of a finite VG; **(f)** *non-finite clause*, whose head is a *lexical* but *non-finite predicational* category (which can be a *predicational* but *non-finite* V, or a *predicational* N, or a *predicational* Adj), together with its FX-bar projection.

Specif (or Spec) is also postulated in SCD to be a *functional* category bearing *quantificational features* at the lexical level (in particular, the negation at the X1 level), including (lexical or non-lexical) (in)definiteness, thus overlapping sometimes on the X1-marker functional features such as agreement.

The agreement (functional) relations are essential for what is called (*local*, *syntactic*) *cohesion* within SCD strategy: X0-Modif and X0-Specif agreement at the X1-level, 'Predicate-Subject' and Compl-$\text{Pron}_{\text{Clitic}}$ agreement at the X2-level etc. These kinds of (agreement, reference, and co-reference) *local cohesion* relations are responsible for a large category of *local dependencies*, including 'long-distance' dependencies. We claim that *all* these 'agreements' are *local cohesions* derived from the *predicational* head and its argument(s). *Global cohesion* of Marcu (1997), representing a chain of co-references for the same individual, is the discourse-level counterpart of a similar set of syntactic devices, but at the global level of text. There exist a close relationship between SCD local

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

concepts and global approaches to text analysis, *viz.* rhetorical-based discourse issues (RST, Mann & Thompson 1988), as suggested within the following table:

Local (Clause-level) Parsing	Global (Discourse) Parsing
◘ **Segmentation** with marker classes (including discourse markers) at the phrase, clause, and sentence level; ◘ Syntactic, **local-Cohesion** at the N, V, A phrase and clause level; ◘ **Dependency** establishing among local clause-level text.	◙ **Segmentation** with discourse markers (clue phrases) at the clause-like units (discourse segments); ◙ Lexical-semantics, **global-Cohesion** used to reveal the structure of text ◙ **Dependency** establishing between discourse segments (discourse tree).

The FX-bar scheme for *local text structures* (including VG) is enclosed into the *line-bordered* (common) *part* of the figures **2.1.** and **2.2.** that present the *clause-level*, respectively *discourse-level global*, FX-bar schemes.

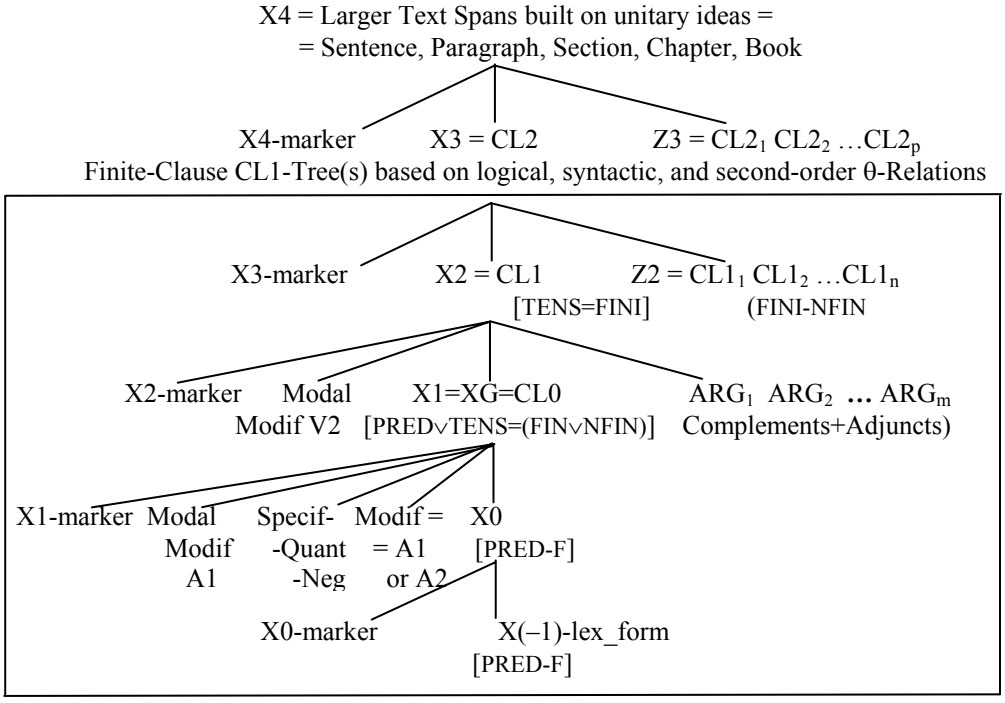


Fig. 2.1– FX-bar scheme for clause-level local (and global) text structures.

2.2. Two Types of Global Text Structures

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

2.2.1. The *FX-bar Scheme for Global Clause-Level (Fig. 2.1)*

Global structures could be classified into (at least) two main categories: **(1)** There exist global structures built from *finite-clauses* or *lexical predications* (including both) using *logical* operators, *syntactic* operators (*e.g.* for the *relative clause*), and *second-order theta-relations* (*i.e. second-order predicational* relations, *e.g.* for the so-called subjective, predicative, direct-completive clauses etc.). **(2)** The usual *clause-based global text structures* are the sentence, paragraph, section, chapter etc. The *clause-level global structures* correspond to the general FX-bar scheme whose elementary constructive element is the finite-clause (Fig. 2.1 above).

2.2.2. The *FX-bar Theory Extended to Global Discourse-Level (DFX-bar)*

There exist *global structures* whose constructive bricks are not necessarily the finite-clause but the rhetorical *discourse-segment* of the RST discourse theory of Mann & Thompson (1988), Marcu (1997, 2000). The FX-bar general scheme extension to RST *discourse-segment global structures* is presented in Fig. 2.2.

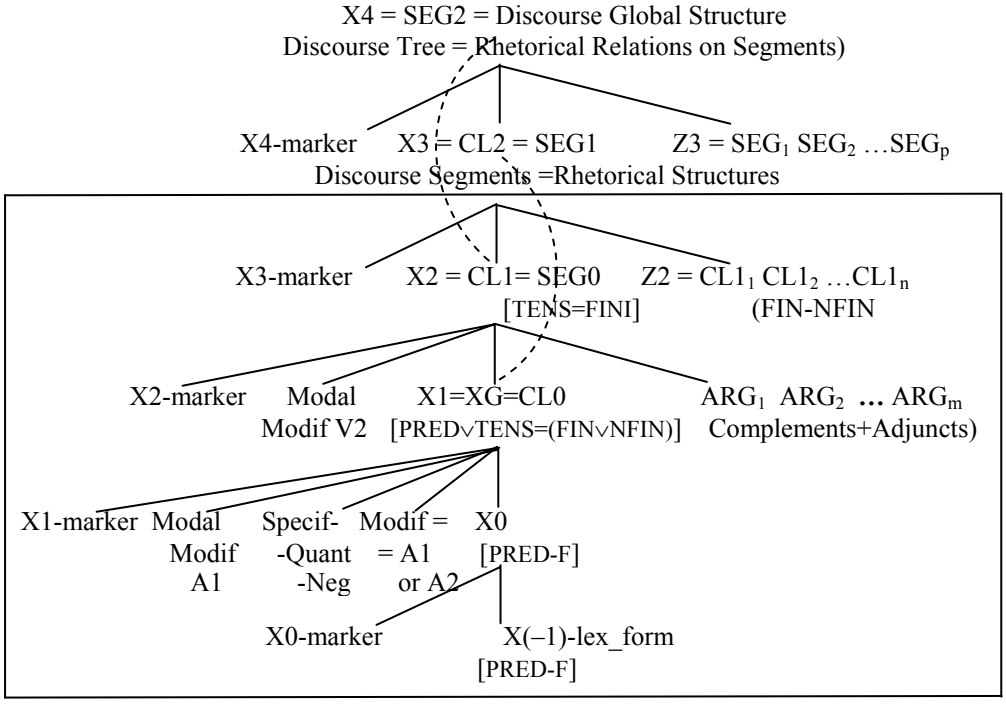


Fig. 2.2 – Discourse-level FX-bar (DFX-bar) Scheme.

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

Remark. Dashed lines represent the special cases when a discourse segment is a proper subclause span and when a discourse segment splits a clause.

In general, a RST discourse segment is made up of one or several finite-clauses. Actually, there exist an intricate relationship between the RST discourse segment and the finite clause, explored in Curteanu (2005). Briefly, we have underlined that there exist sub-clausal discourse segments (*e.g.* a discourse segment constituted from a single NG), or that a discourse marker may split up a (finite) clause into text spans that belong to distinct discourse segments (see §5.2).

Significant elements involved by the *new linguistic projections* incorporate the discursive functionality within the currently proposed DFX-bar scheme (Fig. 2.2), while the categories and structures specified by the projection principles in the *'*old*'*, *local*, *clause-level* FX-bar scheme and theory remain the same (the bordered part in the figures 2.1 and 2.2).

The marker classes are represented into FX-bar schemes at the X*n*-marker levels, $n = 0, 1, 2, 3, 4$. The *explicit description* of the marker classes that are used within the FX-bar schemes in Fig. 2.1 and 2.2, and the graph-based hierarchy of marker classes are exposed in Section 3. The idea of incorporating markers as explicit and stable elements of X-bar schemes, belonging however to an independent hierarchy, began in Curteanu (1988, 1990), and was constantly improved in various papers since then, in Curteanu (1994, 2003, 2006).

The SCD markers are relational (actually, functional) *lexical* (overt) *categories*, *e.g.* (sub)clause-level and discourse collocations (cue phrases, connectors, etc. called *clause* and *discourse markers*), but also lexically empty (covert) functional categories, such as T(ense) (or INFL) in Chomsky (1981, 1986, 1995), or the intrinsic presence of *predicational feature* that can be ascribed, possibly and equally, to each one of the major lexical category N, V, or A, (and inherited by the XG phrase where that category is embedded in), see Curteanu (2003-2004). The dependency graph of the hierarchy of SCD marker classes is also represented in Curteanu *et al.* (2004, 2005) (X*n*-markers in Fig. 2.1-2.2, **M***pq* classes of markers in Section 3 and Fig. 3.2 below).

3. THE SCD MARKER CLASSES AND THEIR DEPENDENCY HIERARCHY

3.1. The SCD Marker Classes

The SCD parsing strategy extends from three to *four* the representation level of marker classes, having functions of setting the boundaries of the main syntactic structures: XG (X = N, V, A), clause, inter-clause, and discourse elementary unit (segment). These classes are presented in detail subsequently.

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

The *first marker class*, denoted **M0** (or M00), applied to the *word dictionary form*, is represented by the *functional role* of *morpho-grammatical inflection*, and corresponds to the lexical (text) level of each word.

a) M1 Class = **{***markers delimiting (introducing) XG structures* **}**.

The M1 class of markers *consists of* X1-*level markers*, (X = N, V, A), *i.e.* markers to be applied to the X1-*level syntactic constructions* (also denoted XG, and called X *groups*). These syntactic constructions consist basically of a *semantic head* (N, V, A category) surrounded by *modifiers* (adjectives or adverbs), and/or by (generalized) *quantifiers* (this includes determiners, negation, etc), *modal modifiers* of level 1-bar (*e.g.* the A1 adverb "*poate*" (*maybe*)) or 2-bar (*e.g.* the V2 modal verb "a putea" (*can-may*)), and/or functionally marked by *pre-positions* (in English, French, Romanian) or post-positions (in German or English) that express the *case* (for N), *aspect* or *meaning* (for V), etc. The main elements of an XG structure provide also the marker subclasses of M1. It is important to mention a certain linguistic (but not linear) *order* of these components of the XG, coming from the *distance* of these elements to the left or to the right of their semantic head, *e.g.* for the noun: the closest to the head are the modifiers, followed by quantifiers, the farthest to the head being the pre- or post-position functional particles. For VG (or verbal complex), the *predicational marking* and FX-bar *projections* are by far more elaborated operators and operations.

M1 can be split into subclasses of markers that are useful in delimiting the XG (X1) substructures, X = N, A, accordingly to criteria such as the above-mentioned *distance* to the X0 semantic head of the surrounding elements, a head which ultimately is always an (overt or covert) objectual common noun, proper noun, or personalized reference to a noun.

M11 = {M11N, M11P }
- M11N = {the occurrence of an objectual, non-predicational common noun, or of a proper noun }
- M11P = {the occurrence of an accentuated or non-accentuated pronominal form}

M12 = {M12N, M12V}
- M12N = {the occurrence of a noun modifier (adjective, pronominal adjective)}
- M12V = {the occurrence of a verb modifier (adverb)}

M13 = {the occurrence of a (generalized) quantifier}

M14 = {pre-positions or post-positions expressing the case (for N), aspect or meaning (for V), etc. }

b) M2 Class = **{***markers that introduce a (finite or non-finite) clause, or a syntactic category group phrase with the semantic head N, V, A* **}**. XG syntactic compound, (X = N, V, A), may be assimilated with a (degenerated) non-finite

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

clause for X = N, A. M2 is split into the following subclasses (in decreasing order of priority when introducing dependency relations):

M25 = {markers that introduce the *relative clause*}.

The explanation for M25 *tag* (and its place in the dependency graph of Fig. 3.2) is that the relative clause represents the most complex syntactic compound playing the role of a modifier, to be applied to its NG head argument. The relative clause is an A2-level modifier in the FX-bar scheme, *i.e.* a modifier of 2-bar (clause) level of FX-bar projection.

M24 = {the occurrence of a *finite verbal group* (FVG) or, simply, the occurrence of the FINIte feature value assigned to a verb, introducing a *finite* VG thus *clause*}.

The whole VG may inherit the FINIte feature value if its (predicational V) *semantic head*, or its (auxiliary V) *syntactic head* for VG with complex tenses, bears this feature value.

M23 = {the occurrence of PREDF = PROCess non-lexical feature value assigned to any of the major categories N, V, A (since the lexicon encoding), thus introducing a clause}.

M22 = {the occurrence of the TENSe=NonFINite feature value assigned to the category V}. See Barbu (1999), Monachasi (2005), Dobrovie-Sorin (1994), Legendre (1999) for various analyses of the *verbal complex*, *i.e.* VG in FX-bar terms.

M25, M24, M23 and M22 marker classes introduce X2-level structures, *viz.* finite or non-finite clauses, made up of an X1 phrase (or XG group, X = N, V, A) that represents the semantic (either finite, non-finite or predicational) head of the X2-level structure, *followed* by the corresponding NG-type (including prepositional-headed) arguments and/or adjuncts within the same clause. Some of the arguments, such as the classical case of the *grammatical subject* (or *all* the arguments, as it is possible in German), may *precede* the X1-type semantic head of the clause to which they belong, see Curteanu (2003: 73). Note that there exists a *systemic* (*canonical*) *order* of the clause compounds (Sgall *et al.* (1986)), or '*actants*' (Arguments and Adjuncts) in a (finite or non-finite) clause: ACT(or), PAT(ient), ADDR(essee), ORIG(ine), LOC(ation), etc. The systemic order of the arguments within a clause is (a *theta-order*) specific to each NL, being obtained as a result of a very careful linguistic and statistic research.

M21 = {markers that introduce JOIN-type relations, *i.e. conjunctions* of the type "*and*", "*or*", "*as_well_ as*", "*together_with*"}.

M20 = {COMMA}.

Classes M21 and M20 comprise markers with an important degree of ambiguity since they may introduce any structure of type X1 (XG groups, X = N, V, A) or X2 (finite or non-finite clauses).

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

c) M3 Class = {*inter-clausal (discourse) markers*}.

The M3 class markers are functions, or relations (when correlated), having as arguments two or several finite (some of them may be non-finite) clauses. These markers are what Curteanu (1994), Marcu (1997), Curteanu *et al.* (2005) and other approaches are calling inter-clause, 'clause-like', or discourse markers, and applying to the X2 = CL1 FX-bar projections of clause-type.

M3 may be partitioned into the following subclasses (in decreasing order of priority when introducing dependency relations):

M34 = {punctuation (pragmatic) markers that separate clauses, *e.g.* ".","!","?"… }

M33 = {inter-clausal / discourse markers that introduce (unambiguous) *strict super-ordination* clausal dependency }. *Strict super-ordination* means the effective *raising* of (*at least*) *one level* of clausal dependency, and is represented by such markers as "*then*", "*else*", etc.

M32 = {inter-clausal / discourse markers that introduce *super-ordination* clausal dependency, including *punctuation marks* such as *colon*, *semi-colon*, *closed parenthesis, second-paired dash*, etc.}. *Super-ordination* means *raising* one (or several) level(s) of clausal dependency, or remaining on the same dependency level within a *coordination*-type dependency. Typical examples of markers from M32 class are: "*but*", "*therefore* (*thus*)", "*even*", "*equally_*(*to*)", "*in_comparison_with* (*compared_to*)", etc.

M31 = {inter-clausal (discourse) markers introducing one (or several) *sub-ordination* clausal dependency level(s), including *punctuation marks* such as *open parenthesis, first-paired dash*, etc.}. This is a large class of discourse markers bearing various types of relations between clauses: logical, syntactic (of several types), semantic, pragmatic, etc.

As mentioned above, each of the M33, M32, and M31 classes may, at their turn, be partitioned into sub-subclasses that contain relational-type markers (expressed by correlation) as relations on clauses, or as functions (with at least two arguments) on clauses.

d) M4 Class = {*discourse markers, which determine the rhetorical relations that can be established between discourse segments***}.**

The *elementary discourse units* (EDU*s*, or *segments*) are identical to clauses in most of the cases, but exceptions can be found, that is, some segments can be constituted of several clauses and, remarkably, sub-clausal segments (non-finite clauses or groups different from the verbal one, but which still contain a covert predication) can also exist (see §5.2). Some of the *discourse markers* are also M3 level markers, i.e. they also have an inter-clausal relation determination role.

The same rhetorical marker can introduce several types of rhetorical relations, the disambiguisation being resolved by additional methods (statistical results, anaphora resolution, and lexical chains).

The M4 level markers can be classified accordingly to several criteria:

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

i) *According to the type of rhetorical relation introduced*;

The M4 level markers determine certain types of rhetorical relations, similar to those described in Mann and Thompson (1988), where the number of these relations is approximately 25, the list being extended in Marcu (1997). The M4-level markers can be classified by the type of the established relations:

Antithesis: *dar, însă, cu toate acestea, ci, dacă nu, numai nu*;
Concession: *deşi, cu toate că, cel puțin*;
Detail: *în acelaşi mod, la fel cum, cât despre*;
Duration: *niciodată, încă o dată, după ce, în tot acest timp*;
Elaboration: *pe deasupra, şi încă, în acea perioadă, la care*;
Justify: *dar şi, însă, de asemenea*;
Purpose: *pentru că, ca să, fiindcă, cu scopul*;

ii) *According to the type of discourse units introduced*;

The discourse markers can be classified depending on the type of the discourse segments they introduce: markers that introduce *nucleus-type* discourse units (*dar, însă, atunci, altfel, în primul rând*) and discourse markers that introduce *satellite* units (*chiar dacă, cu toate că, din cauza, dacă*).

iii) *According to the complexity of the introduced relations*;

Applying these criteria, the M4-level markers can introduce: **(a)** binary relations – most of the relations between the discourse segments are binary; for example, the *Elaboration* relation, introduced by markers like "*în plus*", "*pe lângă acestea*", "*de asemenea*", "*în afară de acestea*"; **(b)** *n*-ary relations, (n>=3).

There are some rhetorical relations which can have as arguments more than two discourse segments. Among these are the *Joint* relation (introduced by markers like "*şi*", "*sau*"), the *Contrast* relation (introduced by markers like "*dimpotrivă*", "*deşi*", "*ca şi cum*"), the *List* or *Sequence* relations.

An important aspect that has to be considered in establishing the rhetorical relations between discourse units is *marker correlation*. This is also used to establish dependencies between clauses, but at the discourse level it is essential if we want to build the discourse trees correctly.

An obvious example of *correlation* at the M4 level is the 3-uple (*dacă* S1) – (*atunci* S2) – (*altfel* S3) (*if-then-else* relation). In this case, the tree corresponding to the sentence must be built taking into consideration not only the relations between the S1, S2 and S3 segments, established on the markers, but also the relations between the markers, relations that determine the structure of the tree built from the discourse units.

Fig. 3.2. presents the hyper-graph hierarchy of the SCD marker classes. This hierarchy is considered valid for Romanian. Certain modifications could be necessary from a NL to another. When we situate within restrained field of Indo-European languages (such as French, English, German, possibly Russian), one can appreciate that the proposed marker classes and structures (in Fig. 3.2.) remain similar, possibly submitted to slight modifications from a particular NL to another.

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

3.2. The Graph-Based Hierarchy of FX-bar Marker Classes

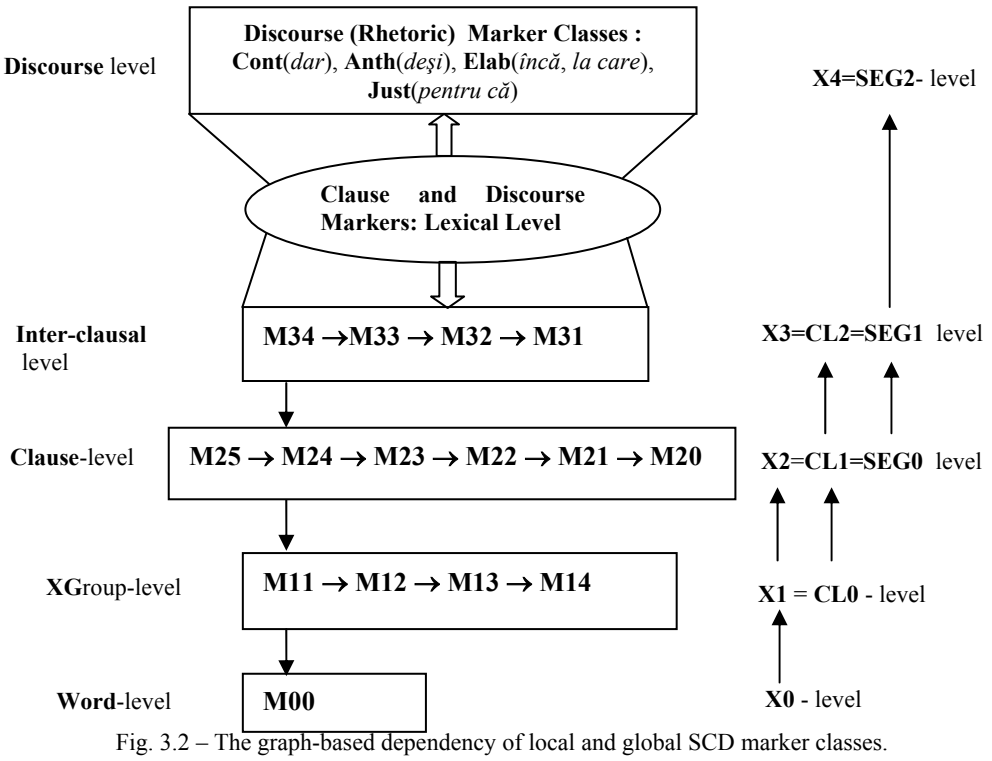


Fig. 3.2 – The graph-based dependency of local and global SCD marker classes.

3.3. The Functional Generativity of Local and Global Markers

In Curteanu (2005), we defined the *functional generative capacity* for *phrase markers* (such as those in SCD or in Marcu's (1997, 2000) parsing algorithms) as follows: when applied at *lexical level*, the phrase markers provide *strong functional generativity*, and when applied at (marker) *class level*, the phrase markers provide *strong functional generativity*, and when applied at (marker) *class level*, the phrase markers provide *weak functional generativity*. An observation is necessary: while a structure built from lexical preterminals N, V, and/or A is *strongly generated* and a sequence of lexical categories (words) is *weakly generated* in the classical sense of *categorial generativity* (Kornai, Pullum 1990, Miller 1999), we consider *lexical markers* as being of *strong functional generativy* since the sharper *functional meaning* of *a lexical marker* entails (is stronger than) the *functional meaning* of a whole *class of markers*.

For example, the categorial strongly generated structure Det A N "implies" (its less informative meaning subsumes the one of) the weakly generated sequence

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

the beautiful flower, and also the functional strongly generated *and*(XG_1, XG_2) implies (in the partial, reverse ordering of semantic meanings) the functional weakly generated *conjunction*(XG_1, XG_2), since the *more informative* meaning of the lexical conjunction "*and*" *subsumes* the inherently *less informative* meaning of the class-depending *conjunction* marker. To further support the proposed definition of functional generativity we observe that the 'strong' lexical marker "*dacă*" entails the 'weaker'-level phrase-marker *conjunction* (a marker class comprising several lexical conjunctions), since the information it holds is richer (e.g., in the sense of subordinate type determination) than the information held by the less informative *conjunction* class (which can only determine the subordinate, but not its type).

Thus, these definitions preserve the general entailment "*strong* implies *weak generativity*", but with the essential *difference* that while "strong categorial implies weak lexical generativity", we need a "strong *lexical marker*" to be *functionally* applied to an utterance to entail a "weak *marker class*" that is applied to the same utterance similarly.

Related to the manner in which markers or marker classes are applied to local or global text structures, the concept of *functional generativity* has immediate consequences on the FX-bar *projections* of local and global text structures, hence within the segmentation / parsing algorithms whose task is to handle the recognition / generation of these entities efficiently.

For instance, using clause markers at *lexical* level in a segmentation task entails a weaker *categorial* generativity and a higher complexity of the algorithms. The same task, worked with *classes* of clause markers, strengthens the expressional generativity and decreases the algorithm complexity.

At discourse level, and especially for the parsing (dependency-establishing) task, it is more profitably to use the lexical markers in order to obtain a stronger *functional* generativity. The usage of the marker classes at this level, either for segmentation or parsing task, determines a certain degree of generality-ambiguity in determining the discourse units and rhetorical relations, resolved by the use of markers at lexical level (see Marcu (1997) and Curteanu *et al.* (2005) for more general tables to be used in global parsing).

In Curteanu *et al.* (2005) we analyzed and classified several classes of local and global segmentation / parsing algorithms, based on such criteria as: **(a) categorial generative capacity** (or categorial generativity); the *strong* and *weak generativity* of major (N, V, A) preterminal and lexical categories, Kornai and Pullum (1990), Miller (1999); **(b) functional generative capacity** (or functional generativity); the *new concept*, introduced in Curteanu (2005), of *strong* and *weak functional generativity* of lexical and, respectively, classes of clause / discourse markers, as the *functional* counterpart to the corresponding notions of *categorial generative capacity* defined for major lexical categories;

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

(c) processing task: *segmentation* or (dependency-establishing) *parsing*; **(d) the output structure targeted**: *clause* level or *discourse segment* (clause-like) level, *i.e.* local and/or global structures as outcome.

We end this discussion on *strong* and *weak generative capacity* mentioning that our concept of *functional generativity* differs essentially from what in Miller (1999: 140-141) is called *derivational generative capacity*, a notion that is related to the derivation trees of functor-argument clause-type in TAGs (*e.g.* Joshi and Schabes (1997), Joshi's previous papers on TAG*s*). We see *derivational generativity* as a generalized form of strong categorial generativity applied to (derivation) trees instead of the simpler categories (or lexical preterminals) N, V, A. The essential distinction between *categorial* (with the more general *derivational*) *generativity* and the concept of *functional generativity* is that they represent different components of the mathematical function object *f*(*X*): the first concept corresponds to the argument *X*, while the second concept corresponds to the *f* function name and role.

4. LOCAL FX-BAR PROJECTIONS

4.1. Verbal Group, VG Kernel, and the Predicational Feature

The verbal group (VG), as XG structure in the BAR = 1 projection level of the FX-bar scheme, contains a *semantic head* verb, *around* which one can find pronouns (only in unaccentuated forms, *i.e.* clitics), special adverbs, auxiliaries, modal verbs (or adverbs), negation. VG is also better known under the label of *verbal complex* (see Monachesi 2005, Barbu 1999, Legendre 1999), and constitutes what is traditionally called *verbal predicate* for the classical clause (proposition). The VG Kernel (VGK) was initially introduced in Curteanu *et al.* (2005) (under the name of *default verbal kernel*), and represents a basic substructure in the VG parsing. The *typical difference* between VG and VGK is that VGK is missing the *proper adverb* of VG (that may syntactically commute with VGK to accomplish the VG).

Examples of VGK*s* (VGK is represented in parentheses, included in VG; unaccentuated pronouns (clitics) in VGK are in *italics*) are: « nu că (nu *mi*-*l* va mai şi plăti) greu; (nu-*i* cunoşteam); (*li se* cereau) ; (*îşi* mai recăpătase) ; (Ai consultat) ; (ar fi simțit) ; (*i se* aşternea) ; (să *se* întâmple) ; (nu *se* putea abține); (n-*o* putea lua); (Nu *i*-ar fi trecut); (să poată afla); (să *te* intimideze); (să *vă* văd lucrând) », see Curteanu *et al.* (2005), Curteanu (2006).

VG may be seen as the 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.* Curteanu (2003-2004), using *diathesis*

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

transformations and *semantic diathesis functions* associated with semantic restrictions on predication arguments.

As Barbu (1999) 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 *modifiers* the (VGK) *special, inside adverbs* ("cam", "mai", "prea", "şi", "tot"), and the proper, VG *outside adverbs* ("*nu*$_1$ *să nu*$_2$ *te* ***tot***$_2$ *duci* ***imediat***$_1$"). The structure of VGK as the "inside" of VG, with a *syntactic head* (the auxiliary bearing the number and person, when present lexically) and a *semantic one* (the predicational verb), with clitics 'inside' (and arguments 'outside') VGK playing an essential role in the development of the *lexical predication* (subsection 4.4) should be further explored, both in linguistic theory and parsing.

In a *verbal group* (VG), the "positive" feature values such as PROC and FINI are *inherited* from the tensed V head by the whole VG phrase, or may be *cumulatively acquired* through morpho-syntactic FX-bar projection.

Somehow similarly (preserving proportions) to the A. Joshi's well-known *tree adjoining grammar* (TAG) and *lexicalized* TAG (LTAG), *e.g.* Joshi, Schabes (1997), SCD strategy may also be seen as a *theory* of (D)FX-bar *scheme* (thus *tree*) *checking and adjoining*. Since in LTAG one considers the *initial trees* to be of the form 'functor-arguments', thus one begins in phrase generation with a clause shell, our VGK, whose structure is a *clause-shadow*, may constitute a substantiation argument for *initial trees* in lexicalized TAGs.

In Curteanu (2003-2004), we discussed a suitable taxonomy for classical predications, involving the classical predicates (VG or verbal complex), based on the lexical property of predicationality PREDF, in agreement also with the extensional / intensional logical representations of these structures.

A typical example of the SCD predicational taxonomy is given by the two main categories of common nouns: **(i)** non-*predicational nouns*, corresponding to *existential-type*, *object-denoting*, *non-event individuals*, whose predicational feature PREDF value is EXIST (*e.g.* [Eng: *student, table*; Rom: *elev-student, masă*]), and whose functional representation in *extensional logic* is done by predicates depending on a single, *extensional variable*: *student*(X), *table*(X) etc.

Our interest is however in **(ii)** *predicational nouns* (often called *nominalizations*), whose predicational PREDF feature value is PROC, *e.g.* [Eng: *meeting, envy, marking,* etc.; Rom: *întâlnire*, *invidie*, *marcare*, etc.], whose functional representations depend on several *intensional variables*, *e.g. întâlnire*(*x*, *y*, …), *invidie*(*x*, *y*, …), *marcare*(*x*, *y*), *donaţie*(*x*, *y*, *z*) etc. Proper nouns and/or personifications are encoded either as constants or variables of extensional nature on which the above extensional / intensional predicates are applied. Other examples: the common nouns *car* and *man* are non-predicational individuals, represented extensionally as *car*(*X*) and *man*(*X*), the adjective *red* is a basic, also *extensional* predicate *red*(*X*), while *leaving* is a *predicational* (process-event)

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

nominal (called *nominalization*) which is represented as an intensional (unsaturated) predicate *leaving*(*x*, *y*), with *x* and *y* as *intensional* arguments.
[Eng: *boy, pencil*; Rom: *băiat, pix*] PREDF = EXIST, and TENS = NFIN;
[Eng: *attempt, showing, proved*; Rom: *încercare, arătând, demonstrat*]
PREDF = PROC, and TENS := NFIN;
[Eng: *are*; Rom: *sunt*] PREDF = EXIST, and TENS := FINI;
[Eng: *gives*; Rom: *dă*] PREDF = PROC, and TENS := FINI.

The *predicational nouns* are typical non-verbal categories whose distributional behaviour is perfectly similar to their verbal counterparts included in VG*s*.

4.2. Classical and Lexical Predications

Without coming into details, the *classical predication* pair (Subject, Predicate) can be viewed as just *one of the facets* of the VG (verbal complex) whose semantic head bears PREDF, the other ones, equally righted as "classical predications", being instantiated by the predicational verb (lemmatized form), endowed with clitic(s) as affixed inflexion(s), which are obligatory present when their valence-based arguments are of personalized semantic nature and optionally present otherwise, doubled or not by the corresponding valence-commanded arguments. Thus, the classical predication pair corresponds to the subject theta-role of "actor" or "actant", while the other "classical" predications 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), at the *lexical level*, of the PREDF feature assigned to the semantic head in VG (verbal complex).

Thus, 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 vector (with 1 to 3) semantic arguments. It is important to specify that there exist normally at least two SUBCAT lists: $\mathrm{SUBCAT}_{oblic_order}$, containing the syntactic arguments of the PREDF_verb, in the order of increasing obliqueness, and $\mathrm{SUBCAT}_{theta_order}$, 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 predication are added, equally righted in the *theta*-semantics, the following similar "classical" predications:

$$(\mathrm{SUBJ}_{Obliqueness=0}, \mathrm{PREDF_Verb}\begin{bmatrix} [VG \quad Tense_Aspect] \\ \textbf{\textit{Semantic_Diathesis}}(\mathrm{SUBJ}, \mathrm{OBJD}, \mathrm{OBJI}) \\ \qquad = (\theta(\mathrm{SUBJ}), \theta(\mathrm{OBJD}), \theta(\mathrm{OBJI})) \\ Agreement(\mathrm{SUBJ}, Inflection_\mathrm{VG}) \end{bmatrix})$$

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

$$(\mathrm{OBJD}_{Obliqueness=1}, \mathrm{PREDF_Verb}\begin{bmatrix}[VG \quad Tense_Aspect] \\ \textbf{\textit{Semantic_Diathesis}}(\mathrm{SUBJ}, \mathrm{OBJD}, \mathrm{OBJI}) \\ \quad = (\theta(\mathrm{SUBJ}), \theta(\mathrm{OBJD}), \theta(\mathrm{OBJI})) \\ Agreement(\mathrm{OBJD}, Clitic\mathrm{OBJD_VG})\end{bmatrix})$$

$$(\mathrm{OBJI}_{Obliqueness=2}, \mathrm{PREDF_Verb}\begin{bmatrix}[VG \quad Tense_Aspect] \\ \textbf{\textit{Semantic_Diathesis}}(\mathrm{SUBJ}, \mathrm{OBJD}, \mathrm{OBJI}) \\ \quad = (\theta(\mathrm{SUBJ}), \theta(\mathrm{OBJD}), \theta(\mathrm{OBJI})) \\ Agreement(\mathrm{OBJI}, Clitic\mathrm{OBJI_VG})\end{bmatrix})$$

Fig. 4.2 – All the extended, valence-based 'classical' predications.

In Fig. 4.2, SUBJ, OBJD, OBJI represents the syntactic categories of subject, direct and indirect 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*) the grammatical, *direct arguments* SUBJ, OBJD, OBJI (sometimes, OBJD2 at the shallow level) to their semantic, *theta-roles* (*e.g. Actant*, *Patient*, *Addressee* etc.) θ(SUBJ), θ(OBJD), θ(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) direct arguments, respectively, on the other hand.

These are the new 'traditional' predications, with their real engine, *viz.* the predicational feature PREDF, installed on the verb head of the verbal group VG. Similarly, non-finite forms of PREDF verbs may be associated to those N*s* (called nominalizations) and/or A*s* 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 4.4 devoted to the problem of VG local structure and its FX-bar projections).

The problem of 'classical' *predication*(s) in HPSG theory in Barbu & Ionescu (1996), or the problem of the *special* role of the *subject* in the SUBCAT list of HPSG in the classical (Polard & Sag 1994: Chap.9) are solved in the linguistic feature structures in Fig.4.2 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 $\mathrm{SUBCAT}_{obliq_order}$, while the output (value) of the function is the VG *semantic diathesis*, *viz*. $\mathrm{SUBCAT}_{theta_order}$ list. This solution forces the subject-*actor* and the subject-

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

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 predicational meaning of that (verb) category. (See the mechanism of *dt* and *sd* functions in subsections 4.4 and 4.5, defined to make operational the direct and inverse FX-bar projections of VGK.)

In Fig. 2.1–2.2 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, while the levels CL0-CL1-CL2 express logical or (second-order) predicational relations on simple clauses. The two global FX-bar schemes work in a (top-down and bottom-up) recursive manner, both in the analysis and generation tasks of the parser, in close relationship with SCD linguistic strategy, its marker classes and hierarchy, and its meta-algorithms for analysis-generation (see *e.g.* Curteanu 1994, 2005, Curteanu *et al.* 1996, 2003, 2004).

4.3. The Problem of Direct and Inverse FX-bar Projections of VG

In the next subsection 4.4 we introduce *diathesis transformations* and *semantic diathesis functions* as useful tools in describing the lexical predication metamorphosis from syntactic (shallow) diathesis to semantic diathesis as a top-down and bottom-up movement, from text to lexicon and backwards. This process may also be understood as direct and inverse FX-bar projection procedures of VG (VGK) towards its (predicational) semantic head and to the clause, derived from the diathesis analysis inspired by Irimia (1997), stated as solutions to the following VG (VGK) FX-bar projection problems:

FX-bar(VG): The problem of direct FX-bar 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}(VGK): The problem of inverse FX-bar projection 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* FXbar^{-1}(VGK) for VGK (or VG), *i.e.* the morphologic-phonologic-syntactic-semantic restrictions on the (predicational) semantic head of VGK that are necessary (and sufficient) to retrieve the VG (VGK) *local* structure through (direct) FX-bar projection of its semantic head.

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

The inverse FX-bar projection associates to VGK a number of (virtual) 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.

This is the starting point in the process of *generation task*, when the first requirement is to generate one or several adequate VGs, satisfying the text *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.

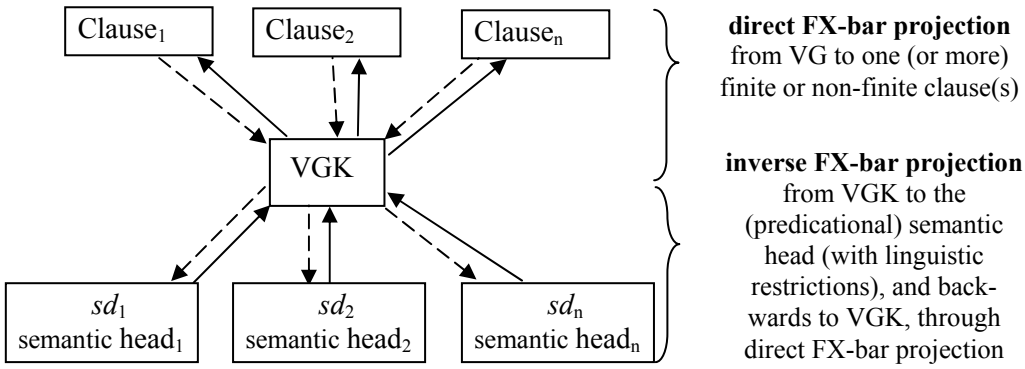


Fig. 4.3 – FX-bar projections of VGK, from text to lexicon and backwards.

4.4. 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. Combiescu (1968: 87-91) distinguishes between active, passive, impersonal reflexive, and dynamical reflexive diatheses, according to the importance given by the speaker to the action presented. Iordan and Robu (1978: 464) considers the realities between the syntactic positions (subject – verb – complement) and their semantic correspondences, (actant-process-patient). Graur (1969: 13-22) considers for the diatheses 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.

We try to solve the above mentioned problems of direct and inverse FX-bar 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 Irimia (1997:

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

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 (Fig. 2.1–2.2). With these elements, one moves down towards lexicon, where one should find **(b)** the second source of information: the valence (number of proper predicational arguments, if any) of the VG semantic head, the *diathesis transformations*, the values of the *semantic diathesis functions*, and the necessary (and sufficient, if possible) restrictions that ensure, as much as possible, the uniqueness of the FX-bar projection values, either direct or inverse, with a target or another.

Here there are tables with the (necessary) information at the lexicon level for diathesis metamorphosis.

Table 4.4.1

Diathesis transformations from syntactic to semantic argument lists

SynD / SemD	Active	Passive	Reflexive
Active	*Dt*([A1, A2, (A3)**]) ↓↑* [A1, A2, (A3)]	Ø	*dt*([A1≡ReflPron, A2, Ø]) (analysis) ↓↑ (generation) [A1, A2, Ø]
Passive	Ø	*dt*([A1, A2, (A3)]) ↓↑ [A2, A1, (A3)]	*dt*([A1, A2≡ReflPron, (A3)]) ↓↑ [X1***, A1≡A2, (A3)]
Reflexive	Ø	Ø	*dt*([A1, A2≡A1, Ø]) ↓↑ [A1, A2≡A1, Ø]
Reciprocal	Ø	Ø	*dt*([A1≡ReflPron, (A2), Ø]) ↓↑ A1≡{X1,X2} {[X1, A2, X2], [X2, A2, X1]}
Impersonal	Ø	Ø	*dt*([(A1)≡ReflPron, (A2), Ø]) ↓↑ [X1, (A1), (A2)]
Dynamic	Ø	Ø	*dt*([(A1)≡ReflPron, (A2), Ø]) ↓↑ [(A1)≡X1, (A2), Ø]

*↓↑ = analysis "↓" and "↑" generation tasks;
**(A_n) = argument optionally present;
***X = uninstantiated variable introduced to support semantically an argument;

The notation "(A1)≡ReflPron" means that the argument A1 is optionally present, the reflexive pronoun is lexically present, and the doubling is possible.

The notation "*dt*([(A1)≡ReflPron, (A2), Ø]) ↓↑[X1, (A1), (A2)]" means that the *semantic_diathesis* function *sd*(category, clitics, syntactic_diathesis, valence) is

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

applied to the *Reflexive* diathesis list [(A1)≡ReflPron, (A2), Ø], the result being the semantic *Impersonal* diathesis list [X1, (A1), (A2)] (see Table 4.4.1).

The *diathesis transformation* functions $dt(\text{List}_0) = dt_1 = \text{List}_1$ map the list of syntactic (grammatical, shallow) arguments (corresponding to $\text{SUBCAT}_{oblic_order}$), into the list of semantic arguments (corresponding to $\text{SUBCAT}_{theta_order}$). The result of transforming a syntactic diathesis into a semantic one is not a unique operation, and Table **4.4.1.** 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(category, clitics, syntactic_diathesis, valence) = $\{dt_1, dt_2, \ldots dt_n\}$, (n = 1 ÷ 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 the other information resulted from the VG parsing.

The next table 4.4.2. presents samples of the phonological, lexical, morphological, syntactical and semantic restrictions on VGK semantic head, its arguments, clitics, and clitics-doubling in diathesis transformations.

Table 4.4.2

Restrictions on head, arguments, clitics, doubling in diathesis transformations

SynD ↓ SemD →	**Restrictions on head and arguments**†	**Examples**
Active → Active	– The VGK can contain maximum two personal pronouns (clitics), one in dative and one in accusative, in the above mentioned order. – Doubling: Obligatory doubling when the object is animated, obligatory non-doubling of the accusative object when the corresponding clitic is postposed inside the VGK. For the other cases, the doubling is facultative.	Nu i-aș mai cere-o *(cartea) lui Ion. *I wouldn't ask John for the book.* Mi-l va da mâine. *He would give it to me tomorrow.*
Passive → Passive	– The VGK can contain one clitic pronoun, unaccentuated form, in dative. The doubling is facultative. – The VGK contains the verb "a fi" (*to be*) and the semantic verb in inflected past tense.	Nu i-ar mai fi fost dată cartea lui Ion dacă nu vorbea cu directoarea. *The book would not have been given to John if he didn't have spoken with the manager.*
Reflexive → Active	– The presence of the Reflexive Pronoun, third person, as clitic in the VGK. – Lexical semantics features: [+activity] – The grammatical subject is also the logical subject (the actant)	Nu m-am mai uitat niciodată la emisiunea aceasta. *I have never looked at this TV Show.*

(To be continued)
(Continuation table 4.4.2)

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

Reflexive → Passive	– The presence of the Reflexive Pronoun, third person as clitic in the VGK. – Lexical semantics features: [+passive], [+transitive] – The grammatical subject fulfils the Object thematic role. – Optionally presence of an agent complement outside the VG, introduced by the preposition "de" or "de către" (*by*).	Protestul s-a semnat de toți. *The protest was signed by everybody.* Școala se va închide de către autorități. *The school will be closed by the authorities.*
Reflexive → Reflexive	– The presence of the Reflexive Pronoun as clitic in the VGK. – Subject – Reflexive Pronoun – Verb agreement in person and number. – The clitic can be doubled by the accentuated form of the reflexive pronoun preceded by the preposition "pe" *(on)*, occasionally followed by the demonstrative pronoun. – The subject and the object refer to the same entity ([+animate]); Coreferentiality between the subject and the object.	Mă cunosc destul de bine [pe mine însumi] *I know myself pretty well.* Mă întreb când vor ajunge? *I wonder (myself) when they are coming?*
Reflexive → Reciprocal	– The presence of the Reflexive Pronoun as clitic in the VGK. – Lexical semantics features: [+ reciprocity] – Outside the VG, a reciprocity complement may appear as a noun (or its pronominal substitute) preceded by prepositions such as "cu" (*with*), "între" (*between*) or "pe" (*on*), the expression formed by the indefinite pronoun *unul (each),* a preposition and the indefinite pronoun *altul (other),* or the expression indefinite pronoun *unul* in accusative, followed by the indefinite pronoun *altul* in dative. – The action is alternatively made by two subjects which are concomitantly two patient complements. – Complementarity between the subject and the object.	Își dau binețe (unul altuia) în fiecare zi. *They greet (each other) every day.* M-am întâlnit ieri cu Maria. *I have met Maria yesterday.*
Reflexive → Impersonal	– The presence of the Reflexive Pronoun, third person. – Lexical semantics features: [+perception verb], [+declaration verb], [+attitude verb]. – At semantic level, this diathesis express the absolutism of the verbal predication in the linguistic interpretation of the reality; only the action development is kept, therefore suspending the relation with a grammatical subject (the agent being therefore usually not lexicalized). – Verb at third person only.	Se dă câte o pâine tuturor. *A bread is given to each person.* Se spun multe. *Lot of things are said.*

(To be continued)

(Continuation table 4.4.2)

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

Reflexive Dynamic →	– The presence of the Reflexive Pronoun. – Lexical semantics features: [+subjectivity].	S-a îngălbenit de spaimă. *He grew pale of fear.* El îşi aminteşte anii copilăriei. *He recalls the childhood years.*

†Phonologic, lexical, morphologic, syntactic and semantic restrictions

4.5. The Diathesis Computing as a Solution to FX-bar Projections of VGK

As already mentioned in subsection 4.4, the semantic diathesis function is defined as follows:

sd(category, clitics, syntactic_diathesis, valence) = $\{dt_1, dt_2, \ldots dt_n\}$, 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 direct and inverse FX-bar projection operation sequence:

Step 1. Extracting an "*a se uita*"-derived VGK from an *arbitrary* clause that encloses it;

Step 2. Handing down to the lexicon, with the semantic head of that VGK;

Step 3. Computing the *sd* and the *dt* function values;

Step 4. Retrieval of the same VGK as FX-bar projection of (one of the meanings of) "*a se uita*" semantic head, associated with the diathesis computed values of *dt* and *sd* functions;

Step 5. 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, 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 4.4.1, one can compute the following values of *dt* functions:

dt_1(*reflexive*) = *active* ⇔ dt_1([A1≡ReflPron, A2, Ø]) ↓↑ [A1, A2, Ø]
dt_2(*reflexive*) = *reflexive* ⇔ dt_2([A1, A2≡A1, Ø]) ↓↑ [A1, A2≡A1, Ø]
dt_3(*reflexive*) = *impersonal* ⇔ dt_3([(A1)≡ReflPron, A2, Ø]) ↓↑ [X1, (A1), (A2)]

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

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

sd(*se_uita*, ReflPron_*se*, reflexive, 2) = {[A1, A2], [A1, A2≡A1], [X1, (A1)]}

Due to the different semantic diatheses, different clause types with distinct readings are potentially parsed (in analysis task) or produced (in generation task).

The following example shows the non-uniqueness for the *sd* function values at the lexical semantics level.

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

(E) *He looks at his brother.*

(2)(R) **Se uită** în faţa televizorului ore în şir. (*sem. diath.* = *reflexive*)

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

(3)(R)**Se uită** deseori semnificaţia zilei de 24 ianuarie. (*sem. diath.* = *impersonal*)

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

For a complete treatment of "*uita*" verb, 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) "*uita*" (*to forget*) entry:

sd(*uita*, Acc_Clitic, active, 2) = {active}

sd(*uita*, Ø, passive, 2) = {passive}.

The corresponding values of *dt* function are:

dt_1(*active*) = *active* ⇔ dt_1([A1, A2, (A3)]) ↓↑ [A1, A2, (A3)];

dt_2(*passive*) = *passive* ⇔ dt_2([A1, A2, (A3)]) ↓↑ [A2, A1, (A3)].

Now, the effective values of *sd* function follow:

sd(*uita*, Acc_Clitic, active, 2) = [A1, (A2)]

sd(*uita*, Ø, passive, 2) = [A2, (A1)].

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) Ion **a uitat-o** pe Maria. (*sem. diath* = active)

(E) *John has forgotten Mary.*

(5)(R) Geanta **a fost uitată** de Ion. (*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 (direct or inverse) FX-bar 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*. Local and global sentence / discourse parsing, machine translation, and FramNet thematic roles assigning are natural applications, *e.g.* Trandabăţ *et al.* (2005).

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

5. GLOBAL FX-BAR PROJECTIONS

5.1. Direct and Inverse Global FX-bar Projections

It is a common fact in classical grammar to expand the thematic arguments (*theta*-roles) from inside the clause to the inter-clausal relations of the same type, using labels such *subjective*, *predicative*, *direct-completive* etc. *clauses*. Such a clause tree, whose *inter-clausal* relations are based on *logical-type* operators (conjunction-disjunction, implication, conditional, concession, consecution-purpose, correlated operators such as *if-then-else* etc.), (purely) syntactic-type relations (such as the *relative clause*), and *second-order theta-semantics* relations (as the above mentioned *theta*-role clauses) could be considered the *global linguistic projection* of the *saturated matrix (root) clause* of the *clause-level tree*. Other (*linking* or grammatical-oriented) weaker or stronger semantics, together with node operations on the clause-tree may be taken into account.

A *similar problem* can be stated for the discourse segments, in particular for the *discourse tree* evolved from the RST inter-segment *rhetorical relations* (Mann, Thompson 1988). Questions of theoretical and practical (computational) importance: which is the discourse projection nucleus for a resulted RST discourse tree, and what is the relationship between the corresponding clause and discourse segment trees?

In terms of discourse tree, we may state the following conjecture: a *text could be seen as* the "global" *projection of its discourse tree* (or of certain significant subtrees of the discourse tree). It is a kind of "summarization" of the text through its main rhetorical components (discourse segments), hierarchically organized as its discourse tree (see Cristea *et al.* (2005) for *summarization* through discourse structure). A similar conjecture may be stated in terms of the corresponding *finite clause-level tree* (or certain subtrees), as a hierarchically organized tree (or graph) of the text enclosed events.

As one can see in the Figs. 5.1.1–4 and examples Ex. 5.1.1–2, the clause-level trees are not necessarily embedded into the corresponding discourse segment trees: in Ex. **5.1.1.**, a subclause phrase makes a unitary segment with clauses in the following sentence(s), and in Ex. 5.1.2 a subclausal phrase (a non-finite clause) is detached and adjoined to the next clause, making a discourse segment.

Making comprehensible the way the (global) projection function is important also from another point of view: *referring* an individual (object or person), a process (event or existence), or a whole bunch of actions corresponding to a larger text span is equivalent not only with the *anaphora resolution* that relates the referring expression to the corresponding text, but also with equating the referred

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

concept as the '*inverse*' *of the linguistic projection function*. In other words, a phrase that refers to a specific text span would be naturally associated with the *head* (or *nucleus*, kernel, projection tree, or another linguistic object) that is (locally or globally) projected into that text, thus with the value of the *inverse projection function* applied to that text. This perspective, also met in Cristea *et al.* (2005), shows one more time the importance of the linguistic projection mechanism and its specification.

Two simple examples may give a better idea of the approach we propose: "*Plecarea vânătorilor în munții Călimani pe o vreme atât de rea a fost pe nepregătite.* Aceasta *le-a fost fatal*." The demonstrative pronoun "*Aceasta*" refers the whole previous sentence, and one could associate it with the sentence (and finite clause) predicate head "*a fost pe nepregătite*", or even corroborate it with the predicational head "*plecarea*" of the enclosed non-finite clause. These phrases represent "*inverses*" of the projection function, applied to the whole sentence at the local, clause-level.

Another possible example is to associate the phrase that refers a whole story within a (larger) text span to the discourse or clause-dependency tree of that text, *i.e.* to the value of '*inverse*' *projection function*, applied to that text, at the global level. This correspondence relates the story reference expression to specially computed nodes and/or subtrees in the mentioned trees.

The problems of *linguistic projection* at the *global level*, floors 3 and 4 in DFX-bar scheme (Fig. 2.2), are especially complex. Two (counter)examples show that an RST discourse segment is not necessarily the projection of its enclosed *saturated matrix clause,* as one would expect. Corroborated with the fact of subclausal discourse segments, pointed out in Curteanu *et al.* (2003, 2004), this gives the flavour for the difficulty of the problems for specifying the *discourse* (*global*) *projection function*, as well as *its* '*inverse*' one, *i.e. the nucleus* (or *head*) *structure* whose projected value is a certain (larger or smaller) text span.

Ex. 5.1.1 (Mann & Thompson 1988: 68) **(1)** *Un nou număr din broşură este în curs de apariție,* **(2)** *şi acest lucru înseamnă o şansă pentru noi propuneri de proiecte*. **(3A)** *Oricine* **(4)** *doreşte să actualizeze intrările în broşură* **(3B)** *ar trebui să aibă copia până la 1 Decembrie.* **(5)** *În caz contrar va fi utilizată intrarea existentă*.

Mann & Thompson (1988) notices that, for the rhetorical relations **condition** and **otherwise**, their classical constructions for RST diagrams do not cover the text (similar to classical programming) conditional "*If* A, *then* B. *Otherwise* C". These syntactic constructions receive a special attention in the latest version of SCD,

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

falling under the category of *correlated constructions* (and *clauses in correlation*), see Curteanu *et al.* (2005), Popârda, Curteanu (2002).

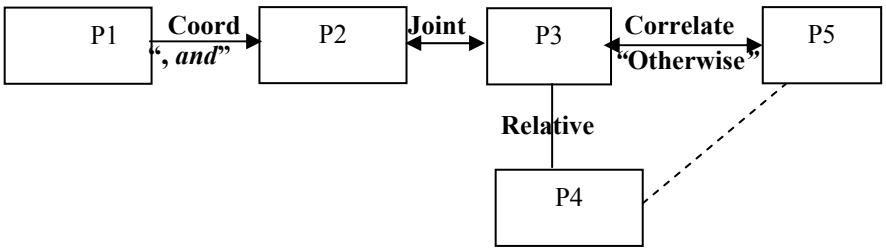


Fig. 5.1.1 – Inter-clause tree inherent to the segment tree of Ex. 5.1.1.

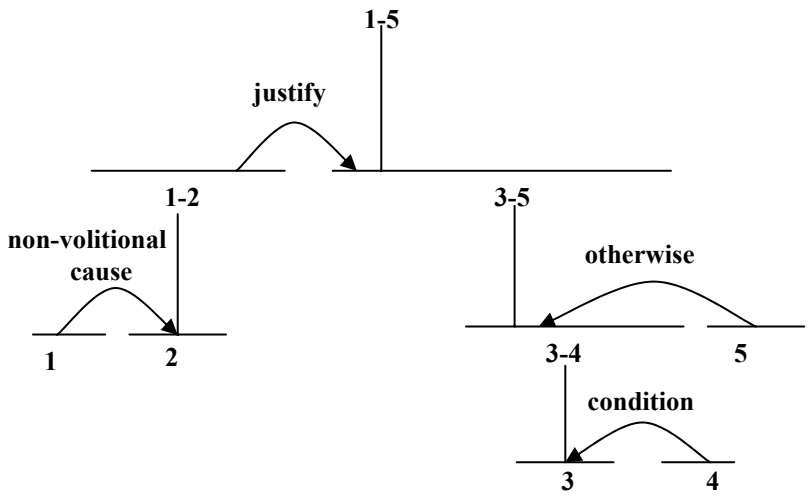
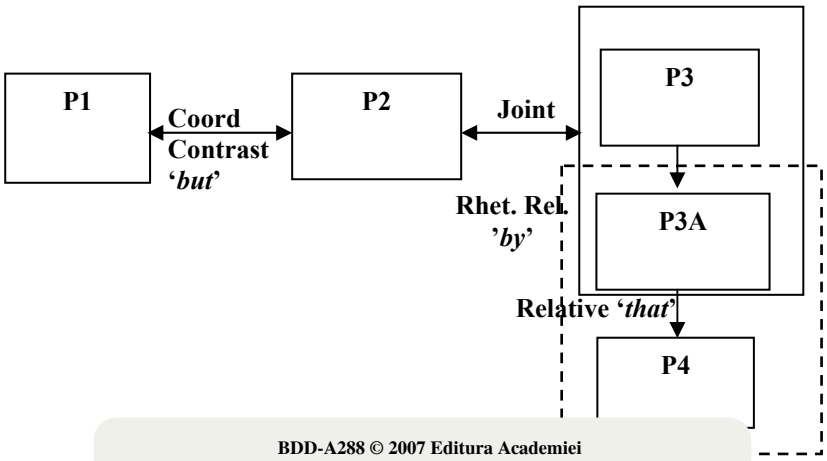


Fig. 5.1.2 – The RST segment tree for Ex. 5.1.1. [Mann & Thompson (1988)]

Ex. 5.1.2 (Mann & Thompson 1988: 76) **(1)** *Animalele se vindecă*, **(2)** *dar arborii se compartimentează.* **(3)** *Ei rezistă întreaga viață la răni şi infecții* **(4)** *prin instalarea unor granițe care rezistă la extinderea microorganismelor invadatoare.*

The figures showing the inter-clausal relations and discourse trees for the texts in Ex. 5.1.1-2 support our statements concerning the global projections at these clause and discourse levels.



BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

Fig. 5.1.3. Clause-level tree inherent to the segment tree of Ex. 5.1.2.

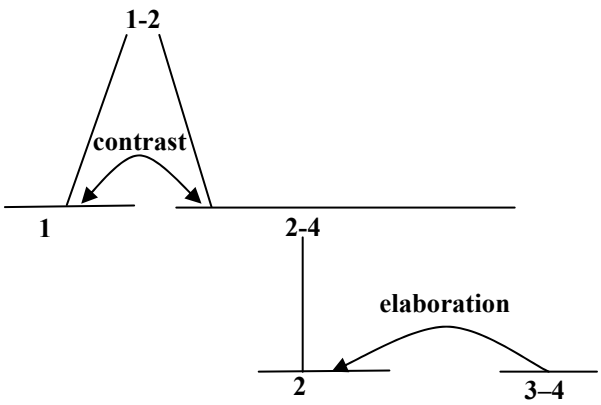


Fig. 5.1.4 –The RST segment tree for Ex. 5.1.2. [Mann & Thompson (1988)]

5.2. A Special Case: Sub-Clausal Discourse Segments

The essential difference between Marcu's discourse segmentation, in Marcu (1997, 2000), and the SCD syntax-driven segmentation, in Curteanu *et al.* (2004, 2005) is the type of target structures that the two algorithms are looking for: Marcu's algorithm's objective is to obtain structures derived from RST rhetorical relations, while SCD's main purpose is to reveal the sentential, syntactic-semantic structures, at the sentence level, from syntactic category-headed phrases and non-finite clauses, to finite clauses and inter-clausal (syntactic and logical-semantic) relations (see Popârda & Curteanu (2002) for a SCD-based approach for a diacritical analysis of global text structures). Between rhetorical relations and inter-clausal relations of syntactic-semantic nature there is a subtle, distinctive, however

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

very close relationship. The following examples from Marcu (1997: Appendix A) illustrate some aspects of this situation:

Ex. 5.2.1. (Marcu 1997: 269, Text A.4) [*Every rule has exceptions*,] [*but the tragic and too common tableaux of hundreds or even thousands of people snake-lining up for any task with a paycheck illustrates a lack of jobs*,] [*not laziness.*]

Comments. The discourse segment [*not laziness.*] is actually a clause, defective of its (finite) predicate "*illustrates*".

Ex. 5.2.2. (Marcu 1997: 269, Text A.6) [*Cleaning agents on the burnished surface of the Ectype coating actually remove build-up from the head*,] [*while lubricating it at the same time.*]

Observation. The situation when an elementary discourse unit (EDU, or discourse segment) is properly embedded into a (finite) clause is very close to that when a discourse marker splits a finite clause into two spans, each span belonging to distinct EDU*s*. This does not necessarily mean that the two EDU*s* are both enclosed into the same finite clause; the most frequent situation is when a discourse segment tears a phrase from a clause and continues its span on the next clauses(s).

These examples bring further arguments that the relationship between the *discourse segment tree* and its underlying *clause-level tree* is an intricate interplay, falling both in the field of lexical semantics for *local structures* but especially in the area of clause-level and discursive semantics for *global structures* of the text organization. Segment tree projection function is closely related to the composition between the discourse marker semantics and the clause-level predications involved by the subsumed clauses. Until one would know more about the projection functions at global level, about the relationship between clause-dependency and discourse trees of a text span, and how the global projection functions and their inverses work, these issues remain still open and challenging problems.

6. CONCLUSIONS

The *phrase markers* play a fundamental role in delimiting the syntactic (and also, semantic) structures and establishing their dependencies. This role was emphasized since the beginnings of SCD (see Curteanu (1988) and earlier). One can see now a whole movement toward rediscovering the essential role of markers, especially on the *discourse* and higher levels of the text. The SCD linguistic strategy, in particular the local and global FX-bar theory, is trying to use and to put to work not only the '*connectives*' of several types, '*cue phrases*', '*discourse markers*', etc. but the *whole palette* of text markers, for *all* the (local and global) levels of analysis-generation of NL, from lexical to discourse ones, especially *in the syntax*. SCD makes a special effort to maximize the use of the lexical-semantics

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

and syntactic means in discovering the logical-semantics and discursive structures of NL.

The main novel aspects that make the difference between FX-bar schemes and previous X-bar type theories may be summarized as follows: **(a)** The two *global* FX-bar schemes represent extensions to the *clause* and *discourse levels*, both enclosing an improved shape of the same *local* FX-bar *scheme*. **(b)** The *graph-based hierarchy* of the SCD *marker classes* are used within the FX-bar general schemes, for the local and global levels. **(c)** Theoretical and empirical arguments support a *lexical-level predication*, based on the *predicational feature* assignment to the major lexical categories N, V, A, since the lexicon description. **(d)** Local and global FX-bar projections proved to be powerful tools for solving local (including VG and the predicate) and global (clause-level and discourse) NL text structures. **(e)** The maximal use of the functional (predicational) and relational features of the local and global markers represent an adequate framework for defining the concept of *functional generative capacity*, with consequences on the design and taxonomy of local / global text segmentation / parsing algorithms.

FX-bar scheme may be associated also to a *language-dependent automaton* (working similarly, however, for a large class of NL*s*) that starts with a sentence, receives *on-line* each word of it, and stops at the final punctuation sign. For adequate values of the parameters like *word* (*argument*) *ordering* and *projection direction* of the major categories and markers, the FX-bar scheme can properly represent the correct dependency of linguistic structures.

As a basic component of the SCD linguistic strategy (see Curteanu (1994) and afterwards), the *local and global* FX-bar *theory* may also be seen as a *procedural mechanism* providing a consistent set of principles and rules that ensure a sound functioning of the FX-bar *schemes*, from the lexicon to the discourse level organization of the NL analysis / generation processes. Continuing this perspective and paraphrasing A. Joshi's well-known *tree adjoining grammar* (TAG), SCD strategy may further be understood as a *theory* of FX-bar *scheme(s)*, thus *tree*, *checking and adjoining* (see subsections 3.3. and 4.1.). The same role of procedural mechanism for FX-bar scheme(s) is envisaged for the related but more general model of *Marcus contextual grammars* (see Păun 1997), as a down-to-language strategy putting to work (highly)-*contextual mechanisms* (such as SCD marker classes and dependency principles) for the NL phrase structure recognition and generation. These are just possible interpretations of the challenging role that an evolved FX-bar theory can play within the NL theory and technology.

The global FX-bar scheme exposed in Fig. 2.2. represents an essential extension to the global approach in the context of SCD linguistic strategy. Each of major lexical categories X = N, V, A, along with the grammatical category CL and the discourse category SEG, are projected (recursively) on *three* bar levels (BAR = 0, 1, 2), within *five* local-global levels of FX-bar linguistic projection process. All these structures, except the lexicon normalized X0-lex form, are functionally

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

and/or relationally "marked", through multiple applications, by the *four*-level local / global markers (on the first level of the hierarchy, followed by other sub-hierarchies) whose classes are better specified within the SCD linguistic strategy.

Functional properties of the (predicational and relational) categories can be assigned even from the lexicon level, but the semantic and/or pragmatic context may entail temporarily loosing or gaining such a quality. This is also true for phrases and collocations resulting from the lexical analysis. Discovering and pointing out of the *functional* (*functorial*) and *relational properties* of the words and phrases is an essential task of the NL *parsing* (analysis and generation) processes; this is specific not only to SCD linguistic strategy, but also to *principle-based parsing* strategies, *e.g.* the *rhetorical parsing* of Marcu (1997, 2000). The proposed FX-bar schemes, consolidating the basic ideas of AX-bar schemes in Curteanu (1988, 1994), and FX-bar theory in Curteanu (2000, 2003, 2005), provide both a theoretical support and practical tool for local and global parsing / generation text processing tasks.

A central issue for obtaining a solution to the *direct* and *inverse* FX-bar *projection problems* of VG (VGK) consists in defining and computing *diathesis transformations* and *semantic diathesis* functions, showing that these function values may characterize the way VG structure is (reversely) FX-bar projected into its (predicational) semantic head, as well as (directly) FX-bar projected, on the lines of its semantic head meaning(s), into the corresponding clause(s). These mechanisms support a better understanding of the anatomy and functioning of the (*lexical*) *predication* and (either verbal or nominal) *predicate*. As shown in Tables 4.4.1–4.4.2, specification of the argument lists for predicational categories, from the *syntactic* to *semantic diatheses* and *backwards*, is equally important for the *analysis* as well as for the *generation* phase of natural language text processing.

REFERENCES

Barbu, A.-M., 1999, "The Verbal Complex", *Studii şi cercetări lingvistice*, L, 1, 39-84.

Barbu, A.-M., E. Ionescu, 1996, "Contemporary grammatical theories: grammars of the phrasal head", in *Limba Română*, 45, 1–6, 31–55.

Bouchard, D., 1995, *The Semantics of Syntax. A Minimalist Approach to Grammar*, Chicago & London, The University of Chicago Press.

Chomsky, N., 1981, *Lectures on Government and Binding*, Dordrecht, Foris.

Chomsky, N., 1986, *Barriers*, Cambridge, Massachusetts, The MIT Press.

Chomsky, N., 1995, *The Minimalist Program*, Cambridge, Massachusetts, The MIT Press.

Combiescu, M., 1968, "On the definition of the Diatheses**"**, *Limba Română*, XVII, 1, 87–91.

Cristea, D., O. Postolache, I. Pistol, 2005, "Summarisation through Discourse Structure", in *Proceedings of CiCling 2005*, Springer LNSC, vol. 3406.

Curteanu, N., 1983, "Algorithms for the syntactic analysis of Romanian sentence and clause", in *Proceedings INFO-IAŞI'83*, 533–549.

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

Curteanu, N., 1988, "Augmented X-bar Schemes", in *COLING'88 Proceedings*, Budapest, 130–132.
Curteanu, N., 1990, *A Marker-Hierarchy-based Approach Supporting the SCD Parsing Strategy*, Research Report no. 18, Institute of Technical Cybernetics, Bratislava, Slovak Republic.
Curteanu, N., 1994, "From Morphology to Discourse Through Marker Structures in the SCD Parsing Strategy. A Marker-Hierarchy Based Approach", in *Language and Cybernetics*, Akademia Libroservo, Prague, 61-73.
Curteanu, N., G. Holban, 1996, "SCD Linguistic Strategy Applied to the Analysis and Generation of Romanian", in: D. Tufiş (ed.), *Language and Technology*, Romanian Academy, Bucharest, 169–176.
Curteanu, N., 2000, "Towards a Functional X-bar Theory", Technical Report, *Institute for Computer Science*, *Romanian Academy*, Iaşi Branch.
Curteanu, Neculai, 2002, "Elements of a Functional X-bar Theory Within the SCD Linguistic Strategy", *ECIT2002 Conference*, Iaşi, România.
Curteanu, N., 2003, "Towards a Functional X-bar Theory", in: D. Tufiş, F. Filip (eds), *The Romanian Language in the Informational Society*, Edited by the Romanian Academy, Research Institute for Artificial Intelligence, Bucharest, 51–86.
Curteanu, N., D. Gâlea, C. Linteş, (2003), *Segmentation Algorithms for Clause-Type Textual Units*, in: D. Tufiş, F. Filip (eds), *The Romanian Language in the Informational Society*, Edited by the Romanian Academy, Research Institute for Artificial Intelligence, Bucharest, 165–190.
Curteanu, N., D. Gâlea, C. Butnariu, C. Bolea, 2004, "Marcu's Clause-like Discourse Segmentation Algorithm and SCD Clause Segmentation-based Parsing", in: H.-N.Teodorescu, (ed.), *Intelligent Systems,* Selected Papers from ECIT2004 Conference, Iaşi, România, 59-86.
Curteanu, N., 2003-2004, "Contrastive Meanings of the Terms "*Predicative*" and "*Predicational*" in Various Linguistic Theories (I, II)", *Computer Science Journal of Moldova*, 11, 4, 2003 (I); 12, 1, 2004 (II).
Curteanu, N., 2005, "Functional FX-bar Theory Extended to Discourse (Rhetorical) Structures", in: H.-N. Teodorescu *et al.* (eds), *Intelligent Systems* Conference, Performantica Press, Iaşi, 169–182.
Curteanu, N., E. Zlavog, C. Bolea, 2005, "Sentence-Based and Discourse Segmentation / Parsing with SCD Linguistic Strategy", in: H.-N. Teodorescu *et al.* (eds), *Intelligent Systems* Conference, Iaşi, Performantica Press, 153–168.
Curteanu, N., 2006, "Local and Global Parsing with Functional X-bar Theory and SCD Linguistic Strategy (II.)", *Computer Science Journal of Moldova* (to appear).
Dobrovie-Sorin, C., 1994, *The syntax of Romanian. Comparative Studies*, Berlin, Mouton de Gruyter.
Graur, A., 1969, "Diatheses", *Studii şi cercetări lingvistice*, XX, 1, 13–22.
Iordan, I., V. Robu, 1978, *Contemporary Romanian Language*, Bucharest, Didactic and Pedagogic publishing house, p. 464.
Irimia, D., 1997, *The Morphosyntax of the Romanian Verb*, Iaşi, The Editorial House of the "Al. I. Cuza" Iaşi University.
Joshi, A. K., I. Schabes, 1997, "Tree Adjoining Grammars", in: A. Salomaa *et al.* (eds), *Handbook of Formal Languages and Automata*, 3, Heidelberg, Springer-Verlag.
Kornai, A., G. Pullum, 1990, "The X-bar Theory of Phrase Structure", *Language*, 66, 1, 24–50.
Legendre, G., 1999, "Optimal Romamian clitics: a cross-linguistic perspective", in: V. Motapanyane, (ed.) *Comparative Studies in Romanian Syntax*, HAG, The Hague.
Mann, W., S. 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.
Marcu, D., 1997, *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*, Ph.D. Thesis, Univ. of Toronto, Canada..
Marcu, D., 2000, *The Theory and Practice of Discourse Parsing and Summarization*, Cambridge, Massachusetts, The MIT Press.

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)

Miller, P., 1999, *Strong Generative Capacity. The Semantics of Linguistic Formalism*, CSLI Publications, Stanford, California.

Monachesi, P., 1998, "The Morphosyntax of Romanian Cliticization", in: P.-A. Coppen *et al.* (eds) Proceedings of *Computational Linguistics in The Netherlands 1997*, Amsterdam-Atlanta, Rodopi, 99–118.

Monachesi, P., 2000, "Clitic placement in the Romanian verbal complex", in: B. Gerlach, J. Grijzenhout (eds), *Clitics in phonology, morphology, and syntax*, Linguistik Aktuell, Amsterdam, John Benjamins.

Monachesi, P., 2005, *The Verbal Complex in Romance. A Case Study in Grammatical Interfaces*, Oxford University Press, Oxford Studies in Theoretical Linguistics.

Păun, G., 1997, *Marcus Contextual Grammars*, Dordrecht, Kluwer Academic Publishers.

Pollard, C., I. Sag, 1994, *Head-Driven Phrase Structure Grammar*, Chicago & London, The University of Chicago Press.

Popârda, O., N. Curteanu, 2002, "L'évolution du discours juridique français analysé par la stratégie linguistique SCD", in : D. Lagorgette, P. Larrivée, (eds), *Repreésentations du Sens Linguistique*, LINCOM Europa, Studies in Theoretical Linguistics, München, Germany, 487–502.

Sag, I., T. Wasow, 1999, *Syntactic Theory. A Formal Introduction*, Stanford, CSLI Publications.

Sgall, P., E. Hajičova, J. Panevova, 1986, *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*, Dordrecht, Kluwer Academic Publishers.

Trandabăț, D., M. Husarciuc, M. Lupu, 2005, *Towards an automatic import of English FrameNet frames into the Romanian language*, 1st *ROMANCE FrameNet* Workshop, *EUROLAN 2005*, Cluj-Napoca, 28–36.

Received January 2006

BDD-A288 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 12:53:25 UTC)