Chapter 6

Constructing Compound Verbs in Bangla An Implementation in LKB

6.0 Introduction

The present chapter proposes a semantic constraint to be imposed on verbal lexical signs, which are arranged in a system of multiple inheritance hierarchy. This constraint controls the unification of a V1 with a V2 to ensure that the grammar licenses only well-formed CV sequences. Since the combinatorial well-formedness of a CV sequence depends on the semantic compatibility between the constituent verbs the constraint is defined on the semantic component of a V1 and a V2.

In section 1, I present an overview of this chapter. Section 2 presents the syntactic and semantic description of the type *pole verb word* (V1) that participates in *periphrastic compounding*. This section also demonstrates that the meaning component of a CV is not constructed compositionally from that of its constituent members. I propose a *semantic constraint on periphrastic compounding* in section 3. This constraint regulates the unification of V1 and its associate V2 in order to ensure the construction of well-formed CV sequences. Finally I present an implementation program using LKB platform in section 4.

6.1 Overview

I have argued in chapter 2 that CV sequences represent a single predicate, that is, one functional-semantic unit. In Bangla and many other Indo-Aryan languages certain predicates are customarily expressed by multi-verb expressions as in the case under consideration. This implies that the same or similar semantic information might be conveyed by a one-word

expression in another language. For instance, the meaning expressed by the Bangla CV sequence *ḍeke paṭhano* ‘send for’ has a one-word equivalent ‘summon’ in English. Bangla as well as many other Indo-Aryan languages have two different patterns for expressing predicates on the surface: one that creates a new *word* type from a verbal lexeme, for which I use the term simple verb expression; and another that derives a multi-verb expression by unifying two verbs. CV sequences belong to the second pattern. The two sentences in (1a) and (1b), for instance, contain the past forms of the simple and the compound verb variants of the basic verb *kena* ‘buy’:

1a. *ritu boi̯-ṭa* ***kin-lo***
Ritu book-cl buy-3 pt
‘Ritu bought the book’

b. *ritu boi̯-ṭa* ***kin-e ni-lo***
Ritu book-cl buy-cp take-3 pt
‘Ritu bought the book (implying Ritu is self-beneficiary)

The verbal expressions *kinlo* “buy-3 pt” in (1) and *kine nilo* “buy-cp take-3 pt” in (2) are evidently related. Both verbs represent the third person past form of the basic verb lexeme *kena* ‘buy’. Both verbs license two arguments. However they represent two distinct predicates because they have distinct meaning. In *kine ne̯a* “buy-cp take” (see 1b) the focus is on the buyer’s being self-beneficiary. When a speaker decides to include categorically this semantic specification, he / she chooses the expression *kine ne̯a* “buy-cp take” instead of the simple verb form *kena* ‘buy’ as illustrated in the sentence in (1b). Thus the CV sequence *kine ne̯a* “buy-cp take” is a lexical variant of the verb lexeme type *kena* ‘buy’. Similarly if the speaker intends to express that the buyer has accomplished the event of *buying* in anticipation of future need, he / she will preferably use the CV sequence *kine rakha* “buy-cp keep”. This is illustrated in (1c):

1c. *ritu du-ṭo šaṛi* ***kin-e rakh-lo***

Ritu two-cl saree buy-cp keep-3 pt

'Ritu bought two sarees (anticipating future requirement)'

Speaker's choice and the context of discourse determine which variant of a verb will be suitable in a given context. Just like simple verb forms, CV sequences project the phrase structure of the sentences they head. Figure 1 and 2 below show the position of the head daughter in a tree structure representation for the sentences in (1a) and (1b) respectively:

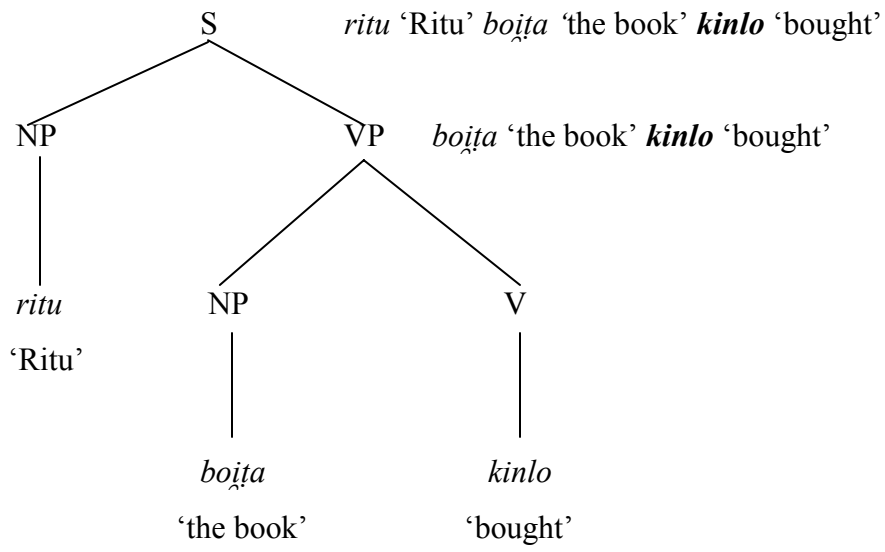


Figure: 1

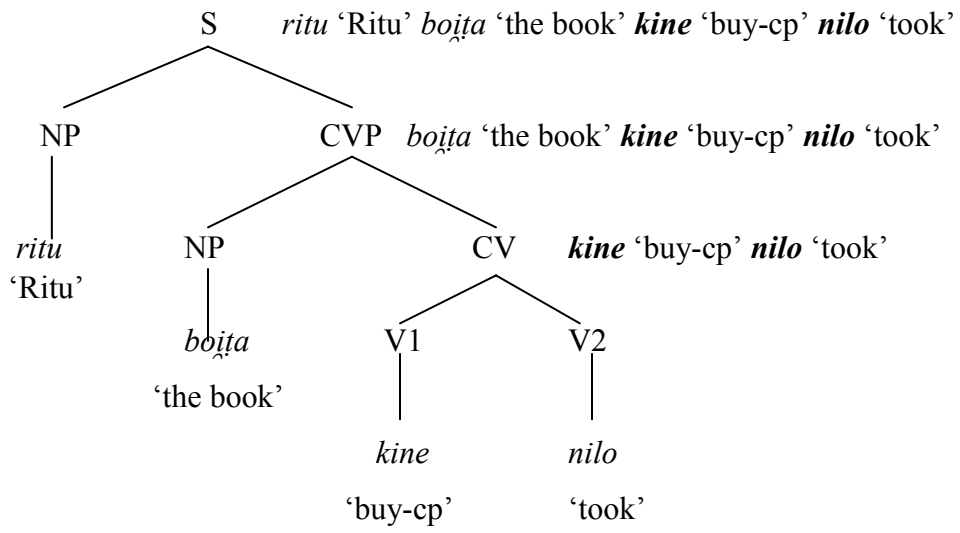


Figure: 2

I presume that CVs are lexical variants of their V1 associate. The resultant construction expresses the core meaning of the V1, modified or extended[1]. Syntactically a CV copies the arguments of its V1 constituent except for the few cases recorded in the previous chapter. The close syntactic and semantic association of a CV with its V1 constituent motivates the present analysis, which considers the V1 as the salient constituent of a CV structure. Nevertheless the following factors underscore the role of V2s in determining the categorial states of CV sequences:

1. A V1 unifies with a V2 only when the meaning component specified at the *participant level* of the two verbs are compatible. In other words the semantics of V2 plays a decisive role in determining whether the unification under consideration is legitimate or not.
2. The semantic information related to telicity, modality, duration and aspect for a CV is always determined by its V2 associate.

[1] Chapter 3 proposes that the semantics of V2 *profiles* part of the base event denoted by the semantic representation of V1 by (i) highlighting (and also sometimes by revising) the manner of involvement of the participant(s) engaged in the base-event and (ii) by adding temporal and aspectual nuances on the semantics of the resultant predicate.

3. The morphosyntactic features assigned to the V2 bear the categorial information for the whole CV construction (see sentences in 1b and 1c).

Since V2 determines the categorial characteristics of a CV sequence, grammatical theories such as GB, which assumes that verbal inflection is the head of a sentence, will favor the position that V2 is the syntactic head of a CV sequence. We face the daunting task of determining which constituents, one that is salient for its contribution of semantic and subcategorization information, i.e. V1 or the one which bears categorial information, i.e. V2, should be given the status of the head daughter within the phrase structure of a CV construction. Irrespective of which of the constituents is designated as the head, it is implicit that a lot of information from the other constituent will have to be copied into it, often rather arbitrarily. One might consider an alternative proposal of having two heads for the CV phrasal structure, both heads independently contributing their share to the lexical description of the CV sequence. In this approach the V1 within a CV will be the syntactic-semantic head (since the argument structure of CV is a copy of that of its V1 associate and also its meaning is an extension or modification of the meaning of the V1) and V2 will be the functional head. Under these circumstances the following stipulations would have to be made within the grammar of HPSG:

1. The *head feature principle* (see chapter 1 and also the later part of this chapter) will refer to the V2 and not the V1 as the head daughter because the HEAD value of the mother needs to be token-identical to the head which bears categorial information;
2. The *valence principle* (Sag and Wasow 1999, p. 86), on the other hand, will refer to the V1-head and not to the V2-head because the subcategorization information is contained in the ARG-ST feature of the V1.

As maintaining the proposal of co-headed structure would imply violation of the fundamental notion that every syntactic phrase has one head daughter, I would not consider it any further. Working within the framework of HPSG, Hinrichs and Nakazawa (1990) have taken the

inflected auxiliary verb (V2 within CV constructions) to be the head of the phrasal structure of complex predicates[2]. Their position stands in conformity with the conventional viewpoint, which considers the functional category to be the syntactic head of the phrase structures that it projects. The proposal of Hinrichs et al., known as *argument composition mechanism*, specifies the following stipulation on the composition of complex predicates: a complex predicate arises when the argument structure list of a functional head (here V2) includes a predicative word (here V1) and its arguments. This theory has been widely adopted for constructing structural descriptions of complex predicates in various languages (discussed in chapter 2). Butt (1995) follows a similar line of thought (within the LFG framework) as she proposes an *event fusion* mechanism[3] for the composition of Hindi-Urdu CV sequences[4] at the syntactic level. According to her, V2s are *incomplete* predicates and their structural description contains a *transparent event* argument. A *transparent event* in contrast to the simple event cannot stand on its own. It needs to combine with the a-structure[5] of another predicate. Thus the function of this *transparent event* in the grammar is to trigger *event fusion*. Ackerman et al. have mainly discarded Hinrichs et al.'s proposal on the following ground:

> It "motivates the need for the creation of new argument structures in the syntax on the basis of constructions involving a combination of two verbs which jointly define the semantic, functional and categorial properties of a clause …"
>
> (Ackerman et al. 1998, p. 17)

The *argument composition mechanism* allows a lexical entry to subcategorize for another lexical entry as a complement. As a consequence, the selecting lexical entry (V2s in the cases under consideration) might inherit some or all of the selectional properties of its V1

[2] I have discussed the characteristic features of complex predicates in chapter2. Like Indo-Aryan CV sequences, these constructions are also multi-word expressions. The difference is that the second constituent of these constructions (as notified for non-Indo-Aryan languages) is an auxiliary, while the V2 of CV sequences is not completely delexicalized.
[3] A variant of *predicate composition mechanism* proposed by Alsina (1993)
[4] Butt has termed this construction as aspectual complex predicate.
[5] Butt's idea of a-structure is described in chapter 3, section 3.1.2.

complement. Since the identity of the complement verb will be known once the two verbs appear together in phrase structure, the argument structure of the V2 will be finally specified only in the syntactic component. According to Ackerman et al., this is a clear violation of the principle of *lexical addicity* which stipulates the following constraint:

> The addicity of a lexical item is lexically fully determined and cannot be altered by items of the syntactic context in which it appears.
>
> (Ackerman et al. 1998, p. 17)

Thus Ackerman et al. object to the proposal of Hinrichs et al. on the ground that the argument structure of a verb, which is supposed to be its lexical property, cannot be determined at the level of syntax. Furthermore, the fact that V2s bear inflection is not a sufficient criterion for assigning to this constituent the status of a syntactic head. There are languages where the categorial information is distributed between the two constituents of an analytic expression representing one predicate. For example, Ackerman et al. show that in Hungarian as well as the Algonquian language Fox the markers which appear on a predicate when it is expressed synthetically are distributed among the pieces of predicates (see chapter 2 for illustration). I will present here an instance from Hindi-Urdu in support of my contention. When a verbal predicate is expressed synthetically, both tense and gender markings appear on the verb. This is exemplified in (2):

2a. *ritu-ne kha:na:* ***kha:e̯-a:***
Ritu-erg meal eat-pt.masc
'Ritu took (her) meal'

The verb agrees in gender with the object *kha:na:* 'meal'. However sometimes, the tense and aspectual information are distributed between syntactically independent auxiliary(s) and verbal forms. This is illustrated in (3) and (4):

3a. *li:na:-ne kha:na:* ***kha:e̯-a: th-a:***
Leena-erg meal eat-pft.masc be("pt")-masc
'Leena had taken (her) meal'

b. *jɔn-ne kha:na:* ***kha:e̯-a: th-a:***
John-erg meal eat-pft.masc be("pt")-masc
'John had taken (his) meal'

4a. *li:na kha:na:* ***kha: rɔ̃h-i: th-i:***
Leena meal eat be("impft")-fem be("pt")-fem
'Leena was having (her) meal'

b. *jɔn kha:na:* ***kha: rɔ̃h-a: th-a:***
John meal eat be("impft")-mas be("pt")-masc
'John was having (his) meal'

The categorial information is spelt out over several members in these multiverbal constructions. The verbal base is *kha:*- 'eat' in (3) and (4). The two verbal elements *rɔ̃h*- and *th*- are auxiliaries. The feminine gender marker in Hindi-Urdu is *–i:*, while the masculine gender is marked by the suffix *–a:*. The gender information is attested on both the main verb and the auxiliary in (3). It spreads over both auxiliaries in (4). Since the subject marker is ergative in (5), the verb and the auxiliary agrees with the object *kha:na:* 'meal'. Since the word *kha:na:* 'meal' is masculine in Hindi-Urdu the verbs bear the masculine gender marker. Bangla does not show gender distinction. The tense and aspect information represented by the analytic expressions in (3) and (4) are conveyed by one-word expressions in Bangla, as shown in (5):

5a. *ritu khabar* ***kha-cchilo***
Ritu meal eat-3 pt cont
'Ritu was having her meal'

b. *ritu khabar* ***khe-lo***
Ritu meal eat-3 pt
'Ritu took (her) meal'

The foregoing discussion related to Hindi-Urdu confirms that morphosyntactic evidence need not be taken as the seminal factor for determining which of the constituents within a multi-word expression (that represents one predicate) is to be assigned the status of the head. Instead, I have chosen to bring into focus the fact that a CV sequence is a lexical variant of its V1 participant. This motivates my proposal that the lexical description of V1 projects the phrase structure of CV sequences. The presence of V2s along with their functional, syntactic and semantic contribution is guaranteed through the lexical entry of the V1. The requirement that a CV sequence will have one *vector* in its surface spell-out is specified within the syntactic description of its V1 participant. For that purpose I introduce a feature VEC within the value of SYN in V1. The assignment of the feature VEC to the structural description of the *verb lexeme* type provides a unified lexical description of synthetic and periphrastic verb forms. The analysis of the simple verb form *kinlo* 'bought' (see 1a) differs from the periphrastic compound *kine nilo* 'bought (implying the sense of actor's being self-beneficiary)' (see 1b) in that the former has an empty VEC, whereas the VEC list of the periphrastic verb contains the V2 *neo̯a* 'take'. I have already noted that a V1 does not combine with every V2. This observation prompts my postulation of a semantic constraint on the meaning component of V1, which controls the unification of V1s and V2s. I refer to it as *semantic constraint on periphrastic compounding*.

Following the proposal of Ackerman et al. regarding the construction of the head feature of verbal forms, I postulate two categorial features CR and HD within the HEAD feature of V1s. When a V1 will have a non-empty VEC list the HD value of the V1 will be token-identical to that of the V2. Otherwise the CR and HD value for V1s will be token-identical. Thus the categorial information borne out by V2s will be copied into the lexical description of the head daughter V1. At the level of syntactic phrase formation V2s play a purely categorial role.

The type *root* is declared to correspond to the phrases that can stand alone as complete utterances of the language. This phrase is represented by the highest node S in the above tree structures (see figure 1 and figure 2). The syntactic-semantic structure of the *root* type must satisfy the following description:

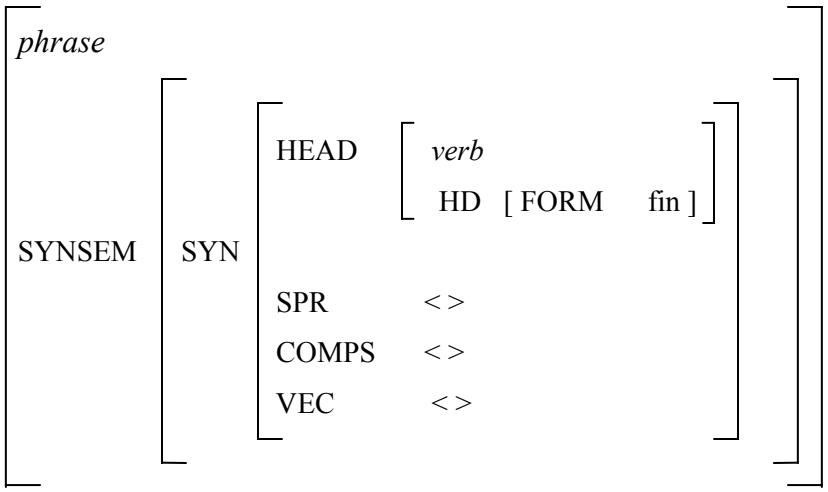


Figure: 3

As shown in figure 3, the root phrase must be headed by a verb in finite form. All syntactic requirements exhibited by the value of SPR, COMPS and VEC must be saturated at this level. The next section discusses the constraints that build the syntactic and semantic structure of a verb that participates in *periphrastic compounding*.

6.2 Structure of SYN and SEM

The following feature structure description represents the minimal syntactic and semantic information associated with a *verb lexeme* type:

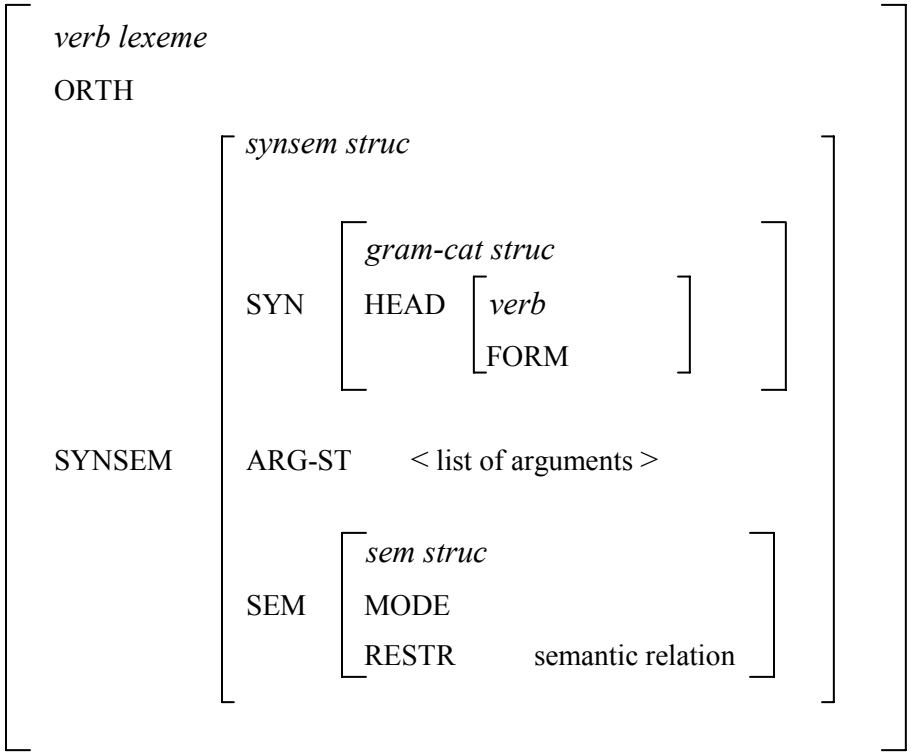


Figure: 4

This feature structure representation of *verb lexeme* type closely resembles that depicted in Sag and Wasow (1999)). When a verb participates in morphological process such as inflection or derivation, the value of FORM for the derived word is formally determined. For instance, a finite verb has the value "fin" for FORM, while the perfective form of the verb has the FORM value "pft". The value of MODE indicates which type of sentence the verb heads. When the verb heads a declarative sentence, the feature MODE is assigned the value "prop". Similarly for a question or an imperative sentence, the value of MODE is defined to be "ques" and "imp" respectively. As illustrated in (1) and (2) in the previous section, a *verb lexeme* type can either be realized as a simple verb or a compound verb on the surface. Accordingly I have distinguished two *word* types in the following hierarchy of *word*:

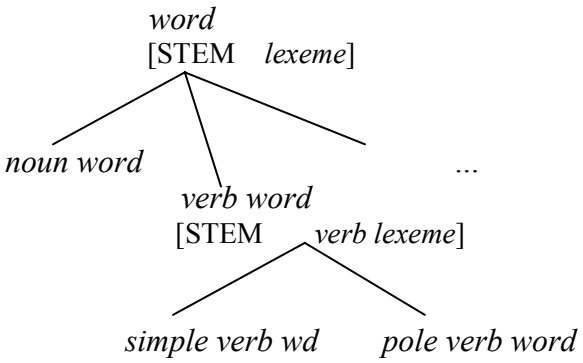


Figure: 5

Words are derived morphologically (either by inflection or by derivation) from its corresponding lexeme. The description of the lexeme is represented as the value of the feature STEM within the word type in figure 5. The word type *verb word*, for example, is morphologically derived from its STEM value *verb lexeme* type. The syntactic-semantic descriptions for the types *simple verb word* and *pole verb word* will be distinct. The former does not participate in *periphrastic compounding*, while the latter does. The following two subsections propose a set of constraints on the verbal word structures.

6.2.1 Structure of SYN in verb word type

I introduce a new feature VEC within the SYN of the *verb lexeme* type. Like the ARG-ST list the value of VEC is declared to be a list. The list contains the *synsem struc* of the V2 that a verb selects in order to constitute a well-formed CV sequence. I propose to keep the HEAD and VEC features for the *pole verb word* type distinct from those for the *simple verb word* type. First I will examine the structure of the HEAD.

The value of the feature HEAD represents the categorial information about a lexical type. I postulate a *pos* type, which is defined to subsume various category types such as *noun, verb, adverb* and so on. This provides a way to determine which features are appropriate for which part of speech. The type *noun*, for example, has a complex feature structure, which represents information related to agreement (number, gender and person) and case such as

nominative, objective and so on. This type is declared to be the value of the HEAD feature within the *noun lexeme* and the *pronoun lexeme*. The value of the HEAD for the *verb lexeme* type is defined to be of the type *verb*. I propose to draw a distinction between CR = "core head" and HD = "function head" features within the type *verb*. I set up a correlation between the status of VEC and the value of CR and HD at the word level. When the VEC list within a *verb word* type is empty, the value of CR and HD will be token-identical. The empty VEC list on the *word* type entails that the *verb word* description is morphologically represented by one word expression. The syntactic description of the finite form of a simple verb expression is given in figure 6:

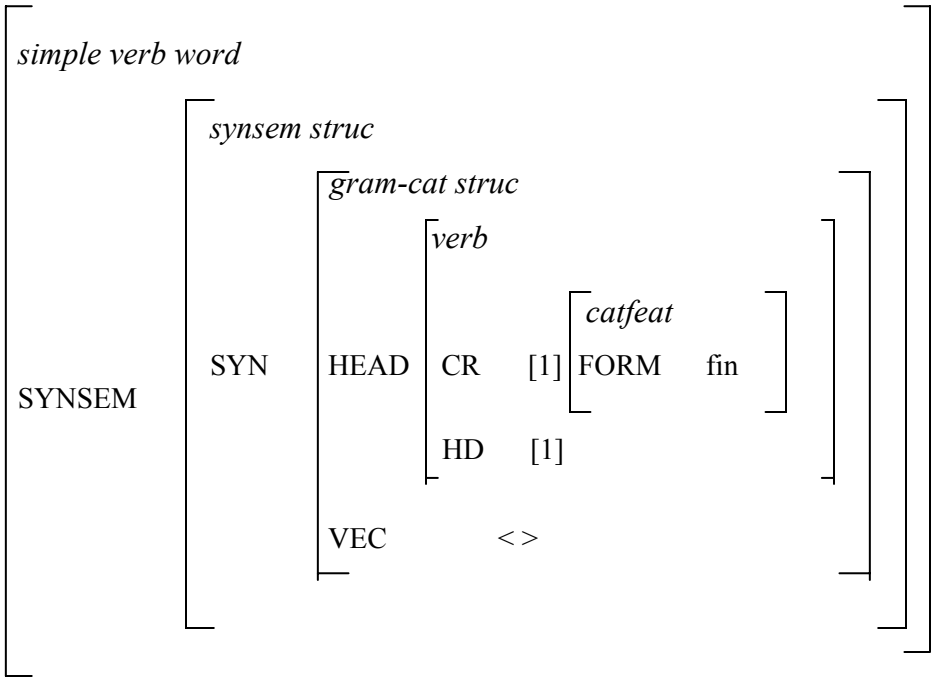


Figure: 6

When a verb has a non-empty VEC list it participates as pole verb (V1) in *periphrastic compounding*. The *pole verb word* type on the word hierarchy in figure 5 represents these verb forms. They copy the HD feature of their V2 associate. Thus the categorial information attested on the HEAD of the V2 participant, becomes part of the HEAD value of the V1. This is illustrated in figure 7:

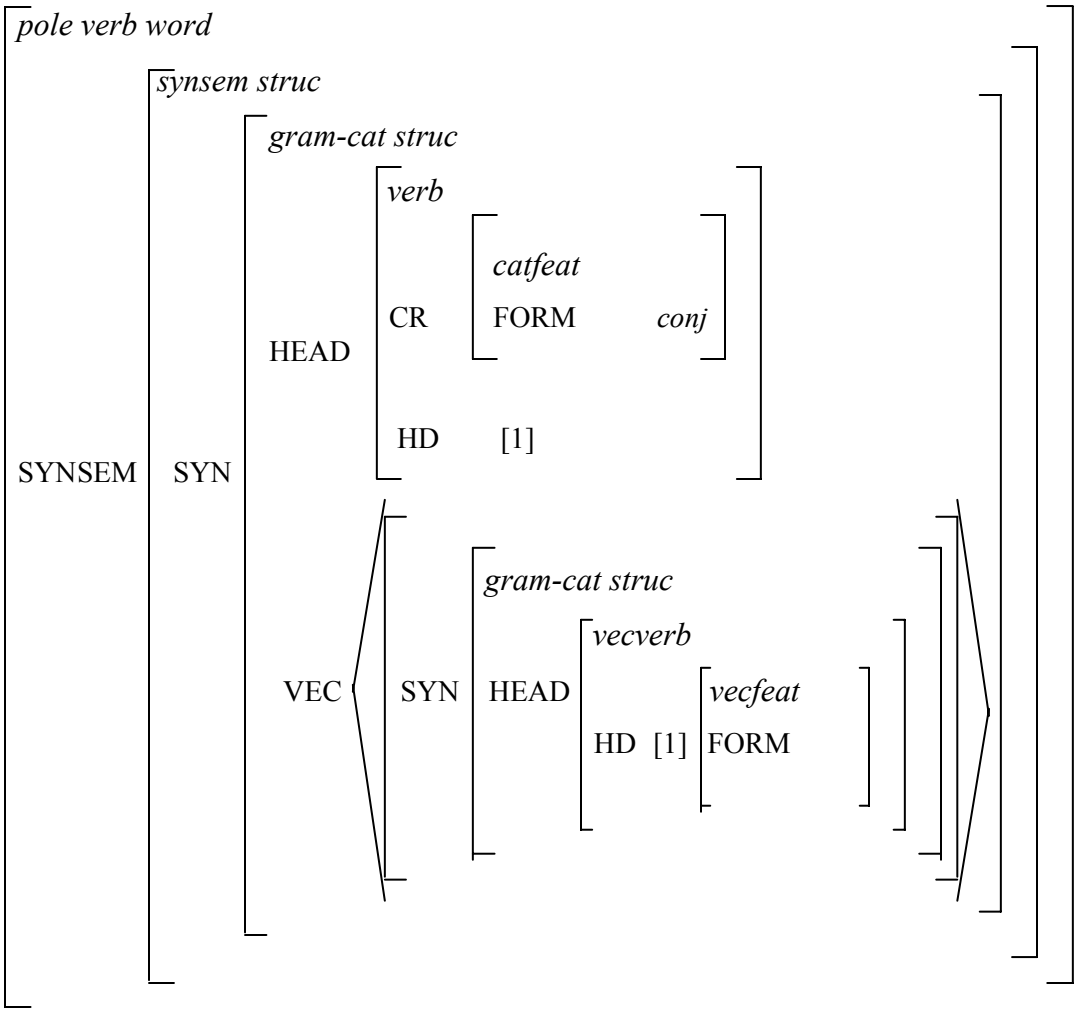


Figure: 7

The following constraint determines the categorial description for the *pole verb word* type:

> A pole verb's (V1's) SYN | VEC is non-empty. The value of the V1's HEAD | HD feature is token identical to that of its V2 associate, which is the value of its VEC.

In a similar fashion I will introduce a constraint on the MODE value of the *pole verb* in the next subsection.

6.2.2 Structure of SEM in verb word type

The value of SEM contains the meaning component. The SEM value of a verb lexeme is represented as a complex feature structure that contains a feature MODE as shown in figure 4. For the *simple verb word* type, the MODE value of the corresponding lexeme is copied to the word level. However, the *pole verb word* type inherits the value of MODE from its V2 associate. The constraint is illustrated in figure 8:

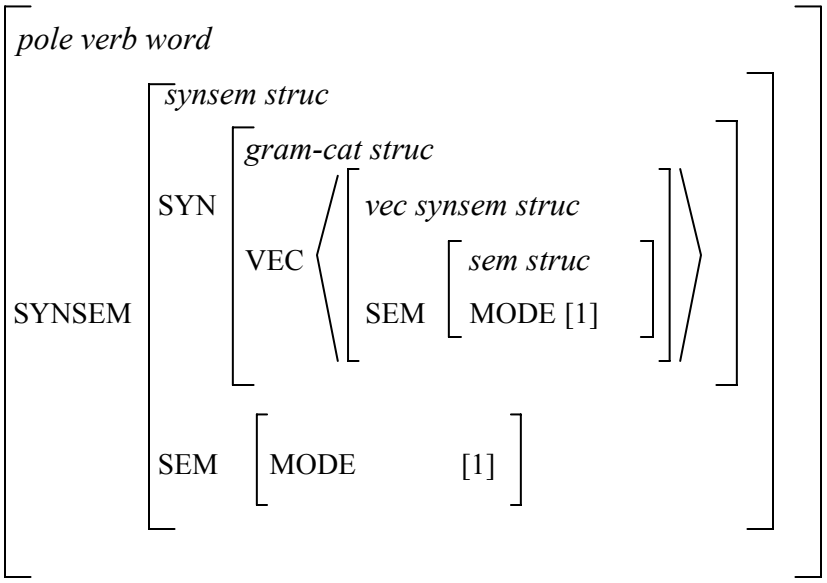


Figure: 8

The following constraint determines the MODE value for the *pole verb word* type:

> A pole verb's (V1's) SYN | VEC is non-empty. The value of the V1's SEM | MODE feature is token-identical to that of its V2 associate, which is the value of its VEC.

I discussed earlier in chapter 3 the need for representing the verb meaning at two separate levels. I refer to them as *participant level* and *supra-lexical level*. The *participant level* contains a linguistically relevant[6] subset of relations among participants in the event or state denoted by the semantic predicator. These semantic informations are represented by various *semantic relations* (see chapter 4). The *supra-lexical level* contains information related to duration, temporality and situational viewpoints or aspects.

I will introduce two features, THEM(ATIC) and GRAM(MATICAL) within SEM. The value of THEM represents semantic information at *participant level*. The value of GRAM presents information related to telicity, duration, modifier and aspect such as *perfectivity, imperfectivity, inceptive, inchoative* and so on.

The THEM and GRAM features are presented as attributes of PREDS in the following feature structure representation:

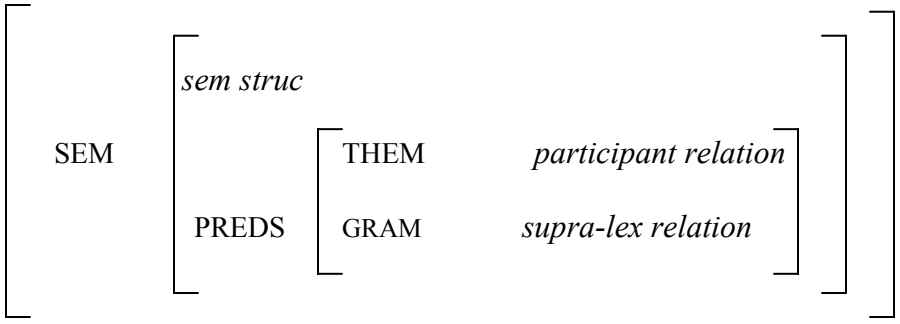


Figure: 9

Below we will see that the information on the supra-lexical level of a CV sequence is determined by the V2 participant.

[6] By the term 'linguistically relevant' I understand the piece of information required for the following two purposes:

1. Defining linking constraint between the semantic roles and syntactic arguments and also
2. Constraining the V2 selection by a *pole verb* (V2)

Verbs, which are popularly known as accomplishment verbs, represent an event type that includes a process and an outcome or change of state. The process advances in successive stages to the point of culmination. The endpoint denotes a change of state. Like activity verbs these verbs can also occur in the context of *adverb of duration*. For example, the sentences in (6) exhibit that the accomplishment verbs such as *ãka* 'draw', *lekha* 'write' and *banano* 'build' are compatible with the *durative adverbials*:

6a. *ritu* <u>*šɔkal theke šondhe porjonto*</u> *chobi* ***ãk-lo***
Ritu morning from evening till picture draw-3 pt
'Ritu drew the picture from morning till evening'

b. *chele-ṭa* <u>*šaradin dhore*</u> *ækṭa ciṭhi* ***likh-che***
boy-cl whole day for one letter write-3 pr cont
'The boy is writing a letter for the whole day'

c. *mɔhešbabu* <u>*tinbɔchor dhore*</u> *baṛi-ṭa* ***bana-len***
Mahesh Babu three years for house-cl build-3 hon pt
'MaheshBabu built the house for three years'

In chapter 3 I discussed the *durative adverbials* in some detail. I noted that post-positional phrases (... *theke ...porjonto* 'from … till …' in (a) and (*tinbɔchor) dhore* 'for (three years)' in (b)) express duration. When the accomplishment verbs (as illustrated in (6)) select the V2 *phæla* 'drop', the end segment of the V1 event is only *profiled* and the resultant CV event becomes non-durative in nature. These CVs do not, therefore, expresses a span of time in the context of *durative adverbials* as illustrated in (7):

7a. **ritu* <u>*ækghɔnṭa dhore*</u> *chobi-ṭa* ***ẽk-e*** ***phel-lo***
Ritu one hour for picture-cl draw-cp drop-3 pt
'Ritu completed drawing the picture for one hour'

b. **chele-ṭa* <u>*šaradin dhore*</u> *ækṭa ciṭhi* ***likh-e phel-che***

boy-cl whole day for one letter write-cp drop-3 pr cont

'The boy is completing writing a letter for the whole day'

c. *******mɔhešbabu tinbɔchor dhore baṛi-ṭa* ***bani-e phæl-len***

Maheshbabu three years for house-cl build-cp drop-3 hon pt

'Maheshbabu completed building the house for three years'

The cases illustrated above confirm that the SEM | PREDS | GRAM value of the V2 constituent determines the supra-lexical information of the resultant CV sequence.

The next section examines the semantic constraint imposed on the value of SEM | PREDS | THEM of a verb's lexeme type. The values of THEM for both V1 and V2s are chosen from a set of semantic relations, which have been identified in chapter 3 and 4. These semantic relations are typed feature structures arranged in a multiple inheritance hierarchy system. We will see in the next section that the semantic constraint imposed on the semantic relations being arranged in a multiple inheritance hierarchy controls the well-formedness of the resultant CV constructions.

6.3 Semantic Constraint on Periphrastic Compounding

The *semantic constraint* is declared on the value of THEM of the V1 and its associate V2 inside the typed feature structure that represents a verb lexeme. This is illustrated in the following feature structure description of the *verb lexeme type:*

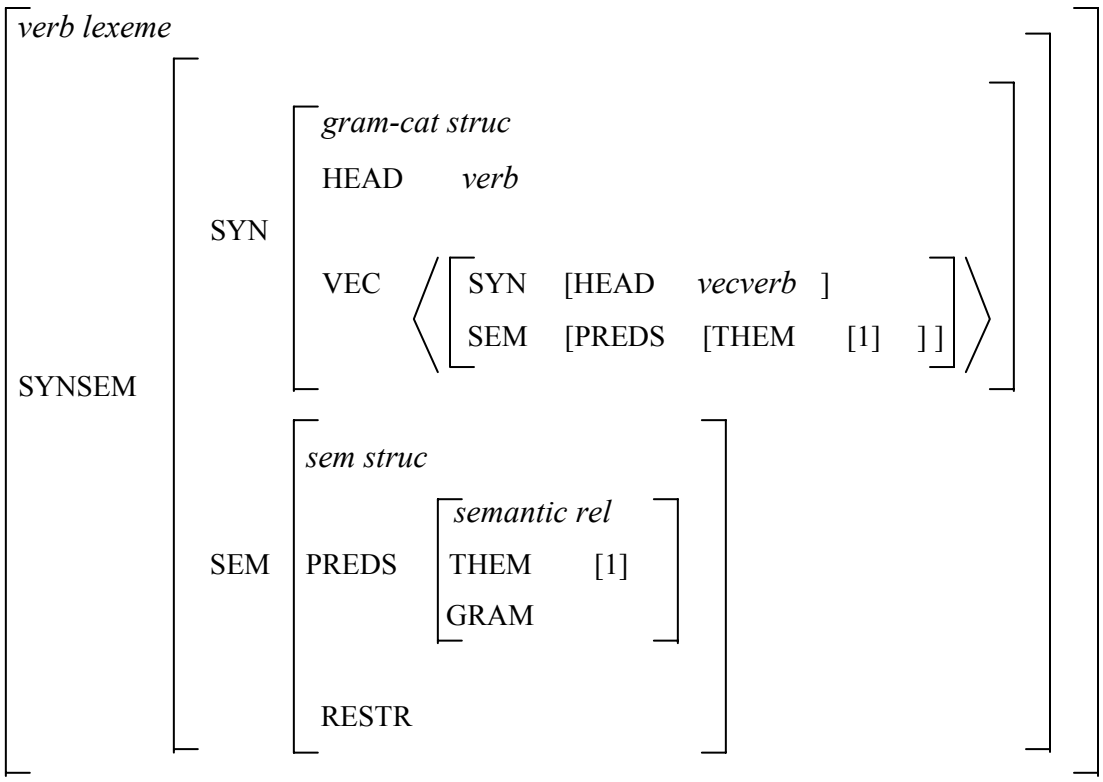


Figure: 10

The constraint states that the value of THEM of V1 and its V2 associate (declared as the value of VEC), must be token identical. The token identity as shown in figure 10 implies that the semantic types representing the value of THEM for V1 and V2 must be consistent. This is possible only when the semantic type representing the value of THEM of one V participant is a subtype of the other or the two semantic types have a common subtype in the hierarchy of the semantic types. In the following sections, I will present a miniature of hierarchy of semantic types in order to illustrate the semantic constraint on *periphrastic compounding*.

The verb *oṭha* 'rise' is a V2 in Bangla. The semantic type *upward manner relation* (see chapter 3, sec 3.5.2.1) constitutes the THEM value of this V2. The following sets of V1s (pole verbs) select the V2 *oṭha* 'rise':

1. ***phoṭa*** ‘blossom’, ***paka*** ‘ripe’

These are inchoative verbs. As discussed in chapter 4, these verbs belong to the semantic type *inchoative upward manner relation*. These verbs denote an event type in which the participant undergoing a change of state is entailed to attain a state of fullness or wholeness.

2. ***jaga*** ‘get up’, ***šara*** ‘cure, heal’, ***bãca*** ‘become alive’, ***khæpa*** ‘become angry’,
mata ‘become engaged in something’

Like verbs of class 1, these are also *inchoative* verbs. In contrast to another set of *inchoative verbs* such as *ghumono* ‘sleep’, *jhimono* ‘doze’ that denote an event type in which the participant is entailed both to undergo a change of state and to attain a state of inactivity, verbs of class 2 imply that the participant comes to a state of being active. As discussed in chapter 4 (see section 4.1.3.6), the semantic types of these verbs are also subtypes of the type *inchoative upward manner relation*.

3. ***kãpa*** ‘tremble’, ***dola*** ‘rock’, ***baṛa*** ‘increase’, ***naca*** ‘dance’, ***laphano*** ‘jump’

In chapter 4, I identified the semantic types of these verbs to be the subtype of *nondislocative motion upward relation* as opposed to other *nondislocative motion verbs* such as *ḍhola* ‘doze’ that entails that the participant is approaching towards an inactive state as he or she dozes. The semantic type of the verb *baṛa* ‘increase’ clearly indicates a rising motion, an entailment of growth, as illustrated in the following sentences:

8a. *gach-ṭa tɔrtɔr kore* ***baṛ-che***
tree-cl rapid prt grow-3 pr cont
‘The tree is growing rapidly’

b. *mee-ṭa-r jɔr* ***beṛ-eche***
girl-cl-gen fever grow-3 pr pft
‘The girl’s temperature has gone up’

4. ***haša*** ‘laugh’, ***cæ̃cano*** ‘shout’, ***kãda*** ‘cry’, ***bɔla*** ‘speak’, ***ḍaka*** *(pakhi* ‘bird’*)* ‘chirp’

These verbs are subtypes of *volitional sound emission relation* which is a subtype of *act relation*.

All the aforementioned semantic relations are arranged in a multiple inheritance hierarchy network as presented in figure 11. Given the constraint stated on the value of SEM | PREDS | THEM of V1s and V2s, the following hierarchy of *semantic relations* determines the rightful unification of the V2 *oṭha* ‘rise’ with the verbs enlisted in the above four classes.

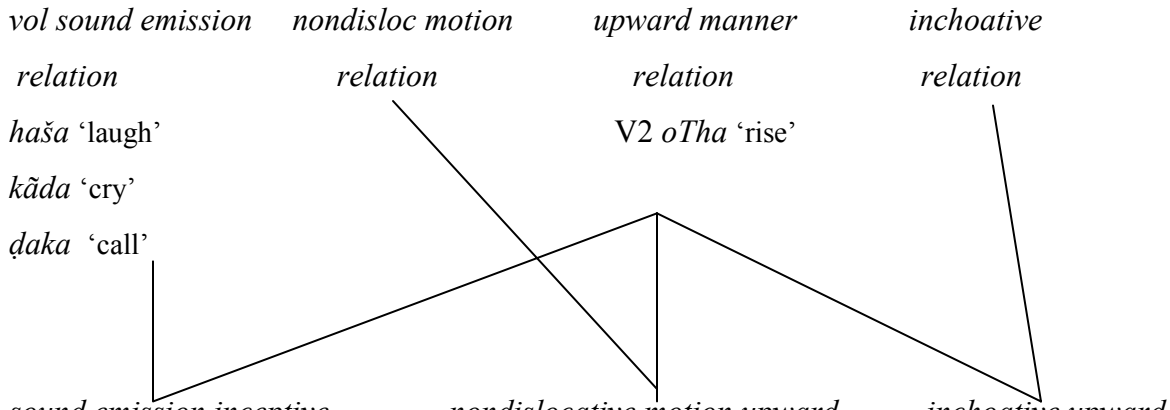


Figure: 11

As shown in figure 11, the *sound emission inceptive relation* type is declared to be the greatest lower bound of the two types, *volitional sound emission relation* and *upward manner relation.* Thus these two types become *type consistent*. Consequently verbs of the semantic type *volitional sound emission relation* (i.e., verbs of class 4) and the V2 *oṭha* 'rise' of the type *upward manner relation* can unify. The semantic types of the verbs of class 1 and 2 are declared to be subtype of *upward manner relation* in the hierarchy. Therefore these verbs also combine with the V2 *oṭha* 'rise'.

The following inheritance hierarchy network illustrates the position of the semantic types of the verbs such as *ghumono* 'sleep'*, jhimono* 'doze' and their CV variants such as *ghumie pɔḭra* 'fall asleep'*, jhimie pɔḭra* "doze-cp drop"*, ghumie neo͡a* 'take a nap (self-beneficiary)' and *jhimie neo͡a* 'take a doze (self-beneficiary)' on the hierarchy.

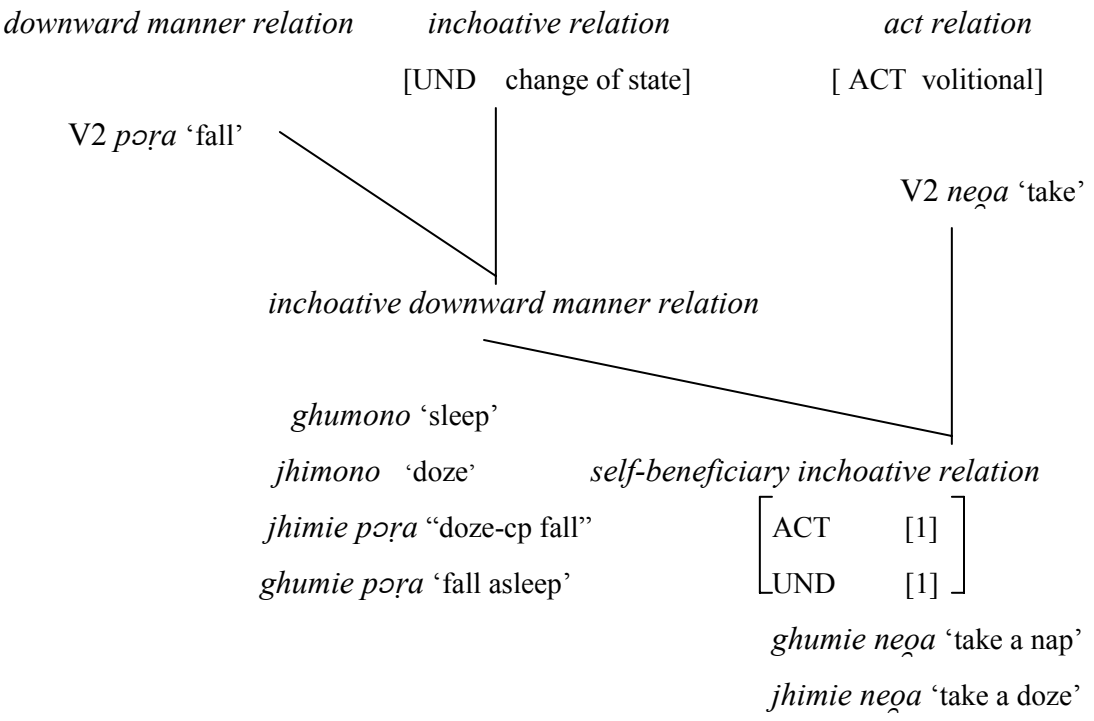
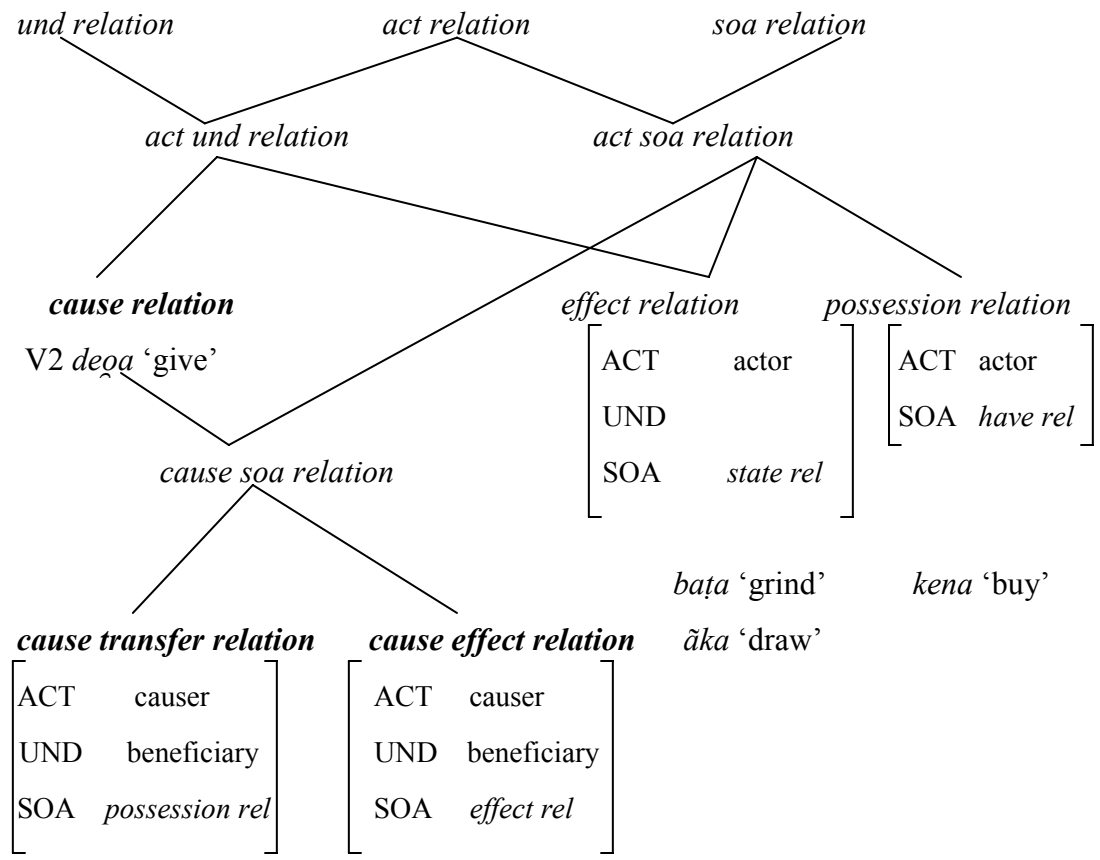


Figure: 12

The THEM value of the V2 *p ṛa* 'fall' is a subtype of *downward manner relation*. The greatest lower bound of the *inchoative relation* and the *downward manner relation* is the *inchoative downward manner relation*. Since this semantic type subsumes the semantic types of the verbs *ghumono* 'sleep' and *jhimono* 'doze', they are compatible with the V2 *p ṛa* 'fall' that belongs to the semantic type of *downward manner relation*. The V1 *ghumono* 'sleep' and *jhimono* 'doze' also unify with the V2 *neo a* 'take'. The resultant CVs imply that the participant involved in the CV event is self-beneficiary. Thus the participant is entailed to have some control over the situation. This semantic information is expressed by the semantic relation *self-beneficiary inchoative relation*, which is the common subtype of *act relation* and *inchoative downward manner relation* in the hierarchy. Since the THEM value of the V2 *neo a* 'take' is also of the type *act relation*, the verbs *ghumono* 'sleep' and *jhimono* 'doze' unify with the V2 *neo a* 'take' and the THEM value of the resultant CV sequences *ghumie neo a* 'sleep (self-beneficiary)' and *jhimie neo a* 'doze (self-beneficiary)' belong to the *self-beneficiary inchoative relation* type. The value of the two proto-role attributes ACT and UND within the *self-beneficiary inchoative relation* structure shares because the participant that undergoes a change of state as he/she sleeps is also entailed to be a volitional and sentient entity.

Finally I will discuss cases where the argument structure of CV is not a copy of that of its V1 constituent. In chapter 5, I presented a semantic account for the argument structure modification. My contention was the following: when the V2 *deo a* 'give' is selected by the verbs of semantic type *effect relation* and *possession relation*, the semantic structure of the resultant CV contains a new semantic role of beneficiary which is required to be linked to an argument on its argument structure list. This prompts a formal argument structure modification for the CVs under consideration. For instance, the THEM value of the verb *kena* 'buy' is *possession relation* and that of the verbs *baṭa* 'grind' and *ãka* 'draw' is *effect relation*. In chapter 4, the structural representation of these semantic relations has been presented. As discussed in chapter 3, the semantic type of the V2 *deo a* 'give' is *cause relation*. The following hierarchy contains the semantic type of the V2 *deo a* 'give' and various other semantic types, which are consistent with the semantic type of this V2:



bæca 'sell'
bece deọa 'sell'
kine deọa 'buy for somebody'
beṭe deọa 'grind for somebody'
ẽke deọa 'draw for somebody'

Figure: 13

As illustrated in figure 13, I postulate two subtypes of *cause relation*. They are *cause tranfer relation* and *cause effect relation*. The semantic type of *kine deọa* and also that of the verbs *bæca* 'sell' and *bece deọa* 'sell' belong to the type *cause transfer relation*. The semantic type of the CVs *ẽke deọa* 'draw for somebody' and *beṭe deọa* 'grind for somebody', on the other hand, instantiate the type *cause effect relation*. The types *possession relation* and *effect relation* are assigned to the value of SOA of the types *cause transfer relation* and *cause effect relation* respectively. This implies that the event of *kine deọa* 'buy for somebody' is

accompanied with a subevent that entails the event of buying. Similarly the event of the CV *beṭe deo̲a* 'grind for somebody' embeds a subevent of 'grinding'. In this way I define a correlation between the semantics of *kena* 'buy' and its CV alternant *kine deo̲a* 'buy for somebody' and that of *baṭa* 'grind' and *beṭe deo̲a* 'grind for somebody'. Thus I assume a bi-event analysis for these CVs. I noted earlier that these CVs license three arguments while their V1 associate takes only two. In this thesis, I formulate a lexical rule, which modifies the argument structure of the verb of *poss lxm* and *effect lxm*[7] and re-map the arguments to the semantic roles within the resultant semantic structure. The rule stipulates the following: if the semantic type of the input is *effect relation* or *possession relation* and it selects a V2 of semantic type *cause relation* (i.e., the V2 *deo̲a* 'give'), then the output verb will have the corresponding semantic type *cause effect relation* or *cause possession relation* respectively (which must be a subtype of *cause relation*). The rule also stipulates the linking constraint in the following manner: the three arguments on the argument structure list of the output will be linked in a left to right order to the value of the top-level ACT, to the value of the top-level UND and the value of the embedded UND.

I have so far attempted to develop a constraint-based analysis for the CV constructions in Bangla. I have also proposed the design of a lexicon in which the general syntactic types, semantic types and their greatest lower bounds representing the supertypes of lexical entries and also various phrasal types are arranged in a multiple inheritance hierarchy network. The principles and constraints determining the well-formed structure for lexical entries and phrasal structures are also declared on the members of this hierarchy. The following section will present the design of a LKB based implementation, which provides a testbed for the analyses being developed for Bangla CV sequences in this thesis.

[7] *poss lxm* and *effect lxm* are the two verb lexeme types that have a SEM value of the type *possession relation* and *effect relation* respectively. In the present discussion, the verb *kena* 'buy' instantiates the *poss lxm* type, while the verbs *baṭa* 'grind' *effect lxm* type

6.4 A Fragment of LKB Grammar

I make use of the grammar and lexicon development environment of the LKB system and build a parsing system for Bangla simple sentences with special focus on CV constructions. The following are the major characteristic features of the LKB grammar that I implement in the present project:

1. The *vector* (V2) and *non-vector lexeme* types are distinguished on the hierarchy of the *lexeme* type.

2. The *valence* and *semantic predicators* are independently arranged on a hierarchy. Their common subtypes are constrained to be the *linking* types that determine the mapping between the semantic roles and the syntactic arguments. Lexical representation of a verb instantiates one of these linking types (see chapter 5, section 5.6.1).

3. The meaning component of a verb is represented at two levels as the value of the features THEM and GRAM (see section 6.2.2).

4. In addition to the feature RESTR (Sag and Wasow (1999)), I introduce one more feature PREDS within the SEM of the verb lexeme type. The value of PREDS is a complex feature structure having the two features THEM and GRAM. I will justify the postulation of the additional feature PREDS in the subsequent section.

5. *Semantic relations* that represent the value of THEM are organized into a multiple inheritance hierarchy network. Following Davis's (2001) proposal I establish homomorphism between the *semantic relations* and *lexical* hierarchy in chapter 5 (see section 5.6.1).

6. Words are derived from the stem lexeme. The hierarchy of word type specifies two subtypes of the *verb word* type: the *pole verb* and the *non-pole verb* (or *simple verb*)

type. The subtypes of *pole verb* participate in *periphrastic compounding*, while those of *simple verb* type are realized as full verbs on the surface.

7. A syntactic schema that builds the phrase structure of CV by saturating the VEC list on the *pole-lrule-infl* type is introduced in the grammar. The *pole-lrule-infl* type instantiates the morphosyntactic representation of verbs that are derived from the *pole verb word* type by applying an inflectional rule which I will be discussing below. I refer to the syntactic schema as *compound rule*.

I will discuss these features in detail in the following subsections. All lexical signs in the LKB system are defined as typed feature structures. Grammar and lexical rules are represented as typed feature structures, which represent relationships between two or more signs. I make a three-way distinction of linguistic signs: *lexeme, word* and *phrase* as shown in the following figure:

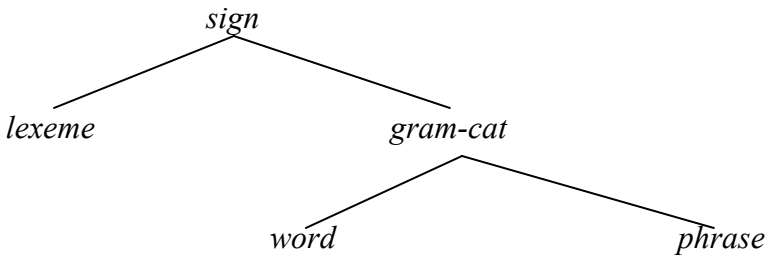


Figure: 14

The members of the type *lexeme* are the basic underived lexical items. In theory, however, a lexical type can be derivationally related to another lexical type. In those cases both the input and the output of the derivation will be of the *lexeme* type. One common instance of this kind of derivation involves valence alternation phenomenon. (When a verb occurs both as intransitive and transitive form, one of the forms can be taken to be the base and the other form can be derived by applying a lexical rule on the base form). The present version of grammar takes the members of the *lexeme* type to be the basic underived lexical entities. The subtypes of the *gram-cat* type are constrained to have one or more than one daughters. For

example, every *word type* has one daughter, the value of STEM that specifies the *lexeme* type from which the word is derived. The members of the *phrase* type, on the other hand, have one or more daughter of *word* type or *phrase* type. The significance of postulating a common supertype *gram-cat* for the types *phrase* and *word* is the following: all grammar rules are so defined that their daughters are of type *gram-cat*. This ensures that in the present grammar rules do not apply to the *lexeme* type.

6.4.1 A distinction between *vector lexeme type* and *non-vector lexeme* type

I identify two types of verb lexemes: *non-vector verb lexeme* (I will call it *verb lexeme* henceforth) and *vector verb lexeme*. The type *verb lexeme* is the STEM value of the *verb word* types. As noted earlier, the subtype of *verb word* is realized on the surface either as a simple (inflected) verb form or as the first constituent of a CV. The SYN feature of the type *verb lexeme* contains three attributes HEAD and VEC, while the SYN of the *vector lexeme* type has only one attribute, the HEAD. The syntactic arguments appear in a list that constitutes the value of the feature ARG-ST for *verb lexeme*. The VEC list consists of the lexical description of a V2 that the *pole verb* can unify with. A vector verb never stands alone in a sentence and it does not unify with another V2 on its own. Therefore I have chosen to assign only the HEAD feature within the syntactic description of the type *vector lexeme*. Since different *pos* types (see 6.2.1 and the discussion below) represent the value of HEAD of the types *verb lexeme* and *vector verb lexeme*, I do not see any need to introduce any specific feature within the HEAD value in order to differentiate these two types of lexemes. As discussed in section 6.2, the value of HEAD of a *non-vector lexeme* type will be a complex feature structure having two features HD and CR. The HEAD of the *vector lexeme*, on the other hand, is assigned only the HD feature in order to ensure that a V2 always occurs as functional head. This verb bears the inflection for tense, person and aspect within a CV sequence. The categorial information about a verbal category is presented in the following figure:

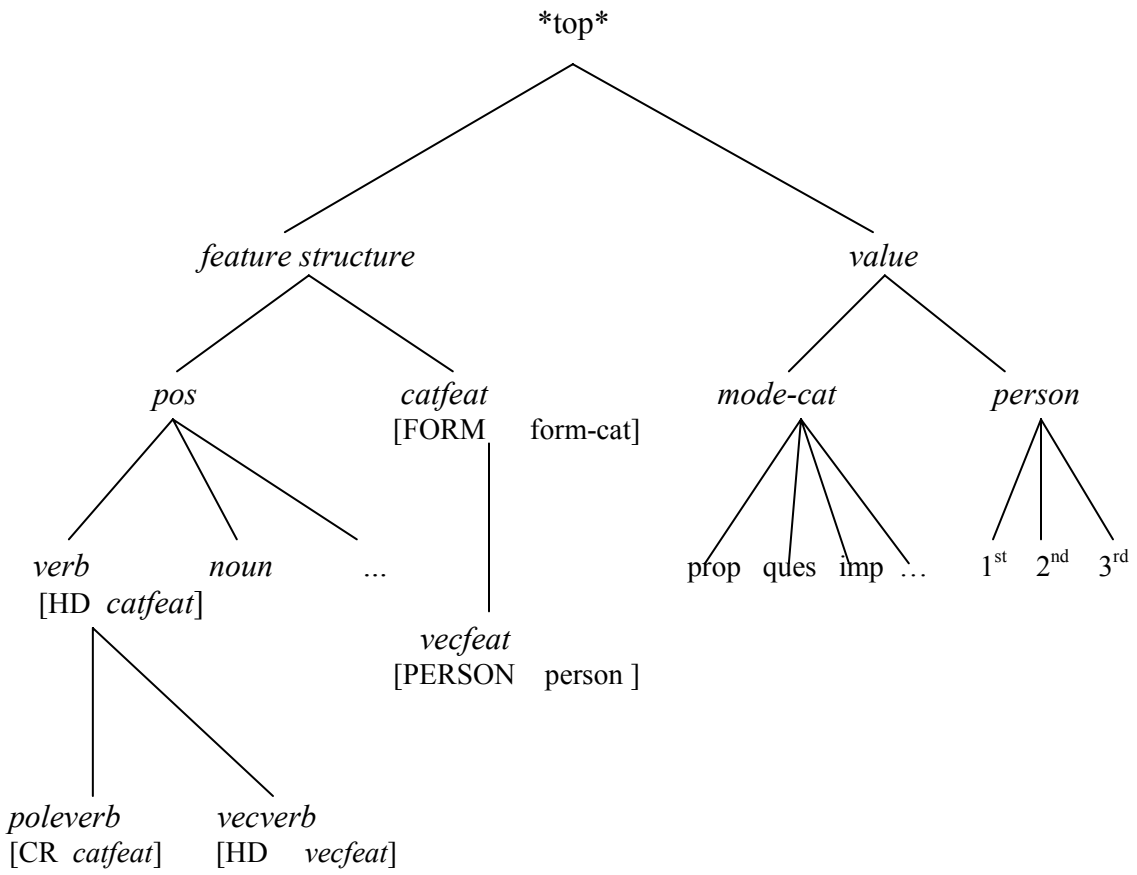


Figure: 15

As discussed in 6.2.1, the type *pos* represents the categorial information about various linguistic categories such as noun, verb, adjective and so on. All type hierarchies have a unique most general type, which is referred to as *top* in LKB. The types *feature structure* and *value* are the two most general types in the hierarchy structure. The members of *value* represent atomic entities, which is assigned as value to various features inside the typed feature structures. The type *feature structure*, on the other hand, subsumes various typed feature structures.

The hierarchy in figure 15 shows that the HD value of the *vector lexeme* type, which is referred to as the *vecfeat* type in the hierarchy, contains a feature PERSON. Its value is

specified when the vector verb is inflected for finite verb forms. For instance the PERSON feature for a 3rd person finite form of a vector verb will be "3rd".

6.4.2 A hierarchy of *verb word* type

Words are derived from their corresponding lexemes. Every member of the type *word* contains a feature STEM whose value describes the *lexeme* type from which the word is derived. Depending on whether the derivation involves affixation, a word will be the member of either the *lrule-no-aff* or the *lrule-infl* type. Figure 16 presents the hierarchy of *verb word* types:

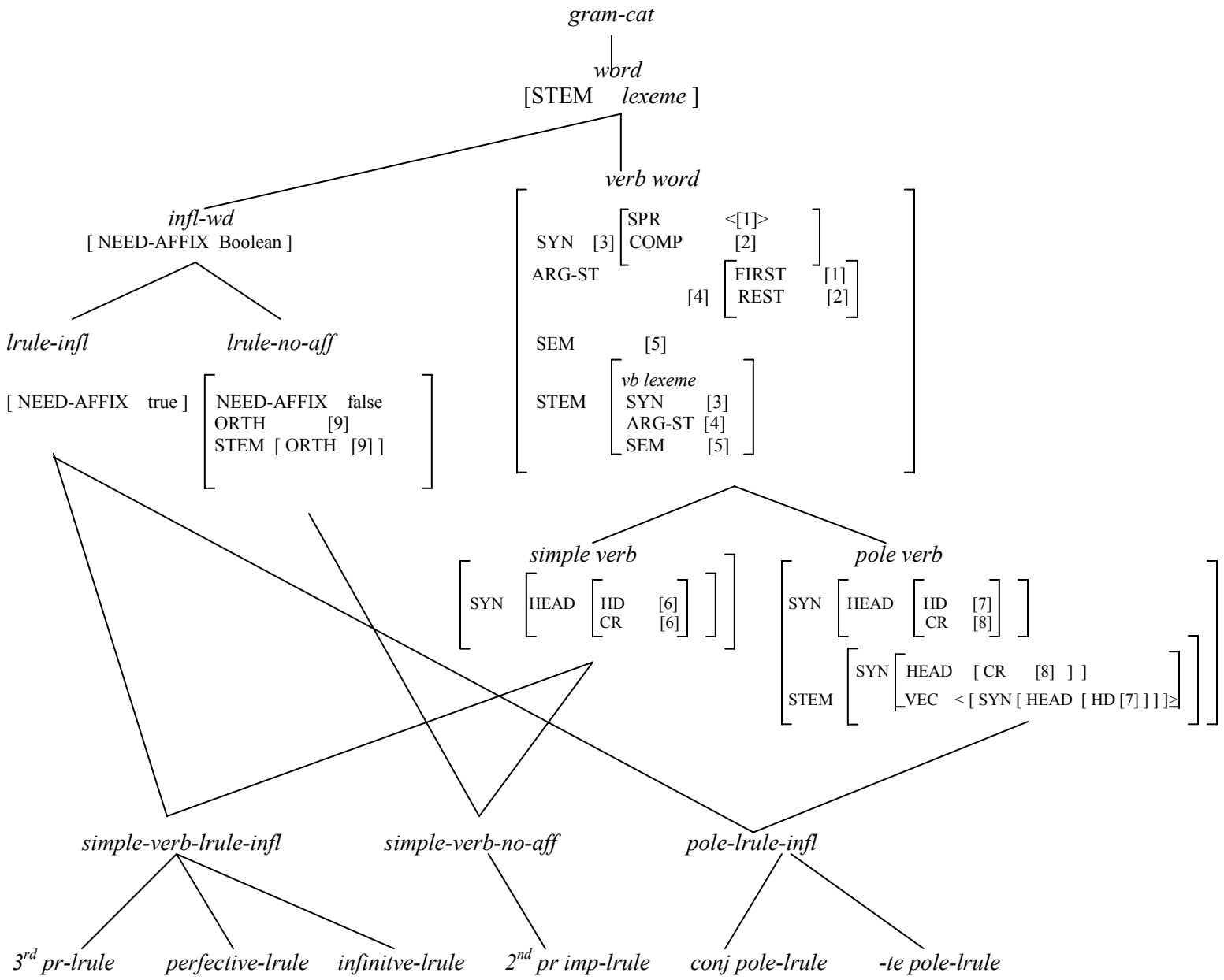

gram-cat
word
[STEM lexeme]
infl-wd
[NEED-AFFIX Boolean]
verb word
SYN [3] SPR <[1]>
COMP [2]
ARG-ST [4] FIRST [1]
REST [2]
SEM [5]
STEM vb lexeme
SYN [3]
ARG-ST [4]
SEM [5]
lrule-infl
lrule-no-aff
[NEED-AFFIX true]
NEED-AFFIX false
ORTH [9]
STEM [ORTH [9]]
simple verb
SYN HEAD HD [6]
CR [6]
pole verb
SYN HEAD HD [7]
CR [8]
STEM SYN HEAD [CR [8]]
VEC < [SYN [HEAD [HD [7]]]]≥
simple-verb-lrule-infl
simple-verb-no-aff
pole-lrule-infl
3rd pr-lrule
perfective-lrule
infinitve-lrule
2nd pr imp-lrule
conj pole-lrule
-te pole-lrule

Figure: 16

I declare the constraint of the *argument realization principle* on the type *word* as indicated by the word structure description of the type *verb word* in figure 16. The *word* type inherits the value of ARG-ST of its stem and the ARG-ST list is split into two subparts. Since I am not dealing with long distance dependency in the present grammar, the summation of the value of SPR and COMPS is taken to be value of the ARG-ST list. The value of SPR is always a one-member list. I postulate two subtypes of the type *verb word*. The members of the *simple verb* type are realized as one word verbal expression on the surface. We have discussed this type earlier in the chapter. It is to be noted that the HD and CR value are made token-identical. On the other hand, the value of HD inside the *pole-verb* type structure-shares with that of its V2 dependent, the value of the feature VEC.

The subtypes of the type *infl-wd* specify the orthographic representation of the derived word. Words that instantiate the *lrule-no-aff* type do not involve affixation. The values of the top-level and the embedded ORTH features are declared token-identical on the *lrule-no-aff* type in figure 16 implying that the orthographic form of the stem lexeme is copied to the value of the ORTH feature on the *word* type. The value of NEED-AFFIX is declared 'false'. Bangla 2nd person present imperative non-honorific verb form belongs to this type. On the other hand, the subtypes of *lrule-infl* undergo orthographic changes. The surface spell-out of all the members of the *simple verb* type belong either the *simple-verb-lrule-infl* type or the *simple-verb-no-aff* type. These two types are declared as the supertypes of all inflectional rules to be applied on the *non-vector verb lexeme* type in the present grammar. Both lexical rules and grammar rules are represented as typed feature structures in The LKB system. An example of 3rd person simple present inflected form of the verb lexeme *ken* 'buy' appears in figure 17:

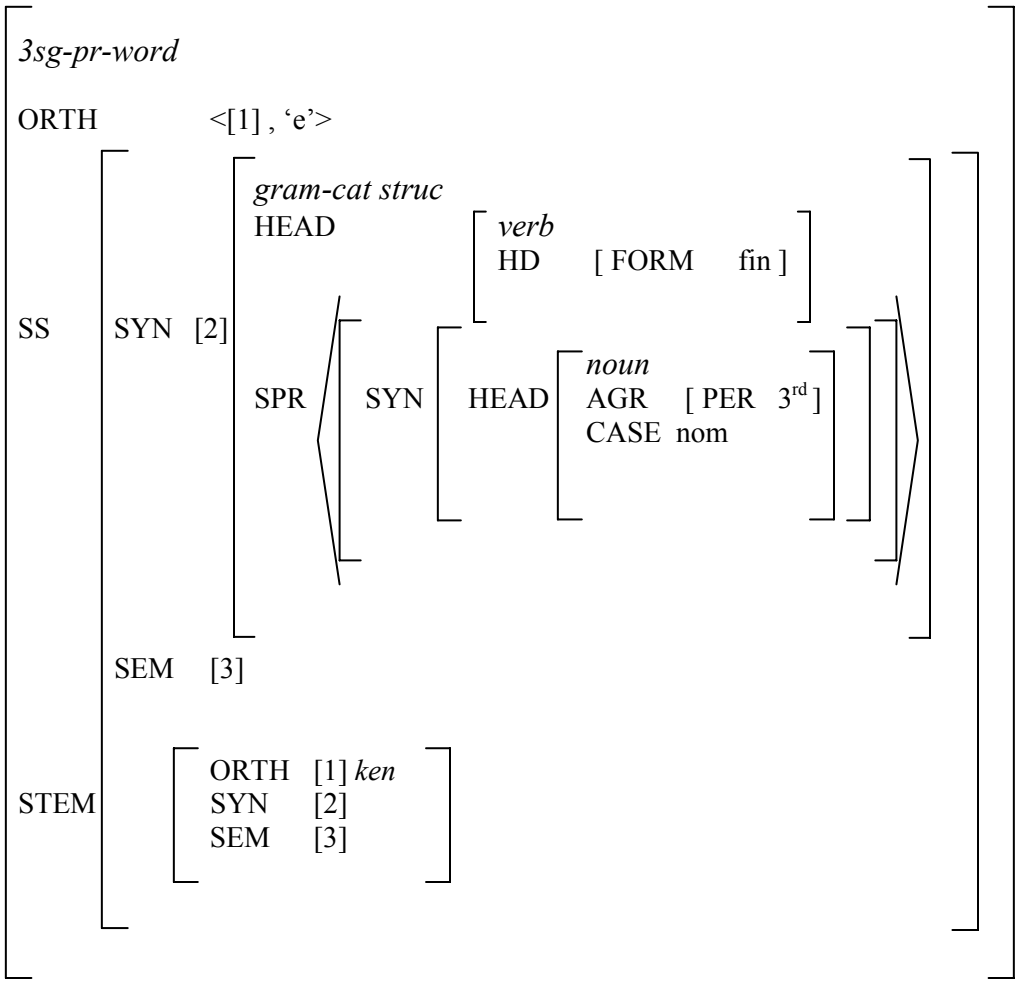


Figure: 17

The LKB system uses a simple string unification approach for encoding orthographic information. The spelling rule in LKB is read in two passes. The information regarding string unification is interpreted as part of the orthographic specification, and everything else being treated as feature structure description. For instance, the rule specified for the above feature structure description in figure 17 also specifies that (i) the value of FORM within HD of the verb will be 'fin' (for finite) and (ii) the argument within SPR list will have a case value 'nominative' and person value '3rd'.

The hierarchy presented in figure 16 shows that the type *pole verb* always undergoes affixation. Therefore no common subtype of *pole verb* type and *lrule-no-aff* type is declared in the grammar. The orthographic output resulted from applying the *perfective-lrule* and the *conj pole-lrule* on the verb lexeme is identical. For example, the verb form *khe-e* in the following sentence express the perfective aspectual information:

9. *ritu bhat* <u>*khe-e*</u> *kOlej-e gæ-lo*
   Ritu rice eat-cp college-loc go-3 pt
   'Ritu took (her) meal ("rice") and went to the college"

The V1 appears in the same orthographic form within a CV sequence as shown in the following sentence:

10. *ritu khabar* ***<u>khe-e</u> ni-lo***
    Ritu meal eat-cp take-3 pt
    'Ritu took (her) meal'

The verb form *khe-e* "eat-cp" is derived from the lexical entry *khaọa* 'eat'. I have designed two separate inflectional rules for deriving the perfective and the conjunctive participial verb form. The *perfective-lrule* appends the perfective morph */-e/* to the morphosyntactic representation of the verb lexeme. The *conj pole-lrule* builds the orthographic form of the *pole verbs* (V1s). The grammar stipulates the conjunctive verb form to participate in *periphrastic compounding*. The need for having two separate inflectional rules is the following:

The syntactic and semantic structure of the perfective non-finite form is distinct from that of the conjunctive pole verb (V1) form. The former is a subtype of *simple verb* type and it is realized as one word expression on the surface. On the other hand the conjunctive verb form is a subtype of the type *pole verb*. It always occurs as the V1 participant within a CV.

The inflectional rule determines the value of FORM for a verb in the present grammar. The value of FORM indicates a verb's functional categories. *Perfective-lrule* declares the FORM value to be 'pft'. This entails that the verb represents the categorial function of perfective aspect. The categorial information for a CV is determined from its V2 participant. The *conj pole-lrule* declares the value of FORM of the V1 to be 'conj'.

6.4.3 Semantics of the CV sequence at the phrasal level is not constructed compositionally

The building of a phrasal structure involves constructing the syntactic and semantic description of the phrasal sign from constituent daughters. All phrases share one common property: they have at least one daughter. The well-formed structural description of phrases is guaranteed through the satisfaction of a set of principles and constraints, which set up a correlation between the syntactic and semantic feature of the mother and the constituent head and non-head daughters. In this section and the following section I will attempt to construct the phrasal structure of the CV sequence. This amounts to building the syntactic structure including the HEAD feature and the semantic structure description for the CV sequence. The present section concentrates in constructing the semantic structure of the CV sequence. The *semantic compositionality principle* (Pollard and Sag (1994) and Sag and Wasow (1999)), which builds the meaning complex for the phrasal structure, states that:

> In any well-formed phrase structure, the mother's RESTR value is the sum[8]
> of the RESTR values of its daughters
>
> (Sag and Wasow 1999, p. 116)

In the grammar of Sag et al. the value of RESTR for a lexical expression denotes semantically relevant restrictions that "specify which of the properties must hold of individuals and

[8] "The term 'sum' has a straightforward meaning: the sum of the RESTR values of the daughters is the list whose members are those values, taken in order. That is, the sum of lists <A>, <B,C> and <D> is the list <A, B, C, D >" (Sag and Wasow 1999 pg 116)

situations, and which relations must hold among them, for an expression to be applicable" (Sag and Wasow 1999, p. 108). That is, *semantic relations* as defined in the present thesis could be taken as constituting the value of RESTR in Sag et al' grammar. The *semantic compositionality principle* as stated above appends the value of the RESTR of the daughters and passes the appended list on to the higher phrase. Thus the meaning of the whole sentence is taken to be the summation of the meanings of the individual linguistic elements that the sentence is made up of. I have, however, noted that the semantic description of a CV is not a summation of the semantics of its constituent Vs. The value of GRAM for the CV sequence is inherited from that of its V2 participant and the value of the THEM is the more specific type, which is consistent with the types specified as the value of the THEM feature for the V1 and V2 constituent. The compositionality principle will, therefore, fail to build a well-formed semantic structure for the CV sequence. Instead, I propose here a *semantic compounding principle* that imposes the following constraint on the semantic structure of the CV phrasal structure:

In the well-formed phrase structure of CV, the GRAM value of the mother structure-shares with the value of the GRAM feature of the V2 constituent. The value of the mother's THEM feature is made token-identical with that of its V1 constituent. The mother's PREDS value thus constructed is copied to its RESTR feature.

Thus the *semantic compounding principle* imposes the following constraint on the phrase structure:

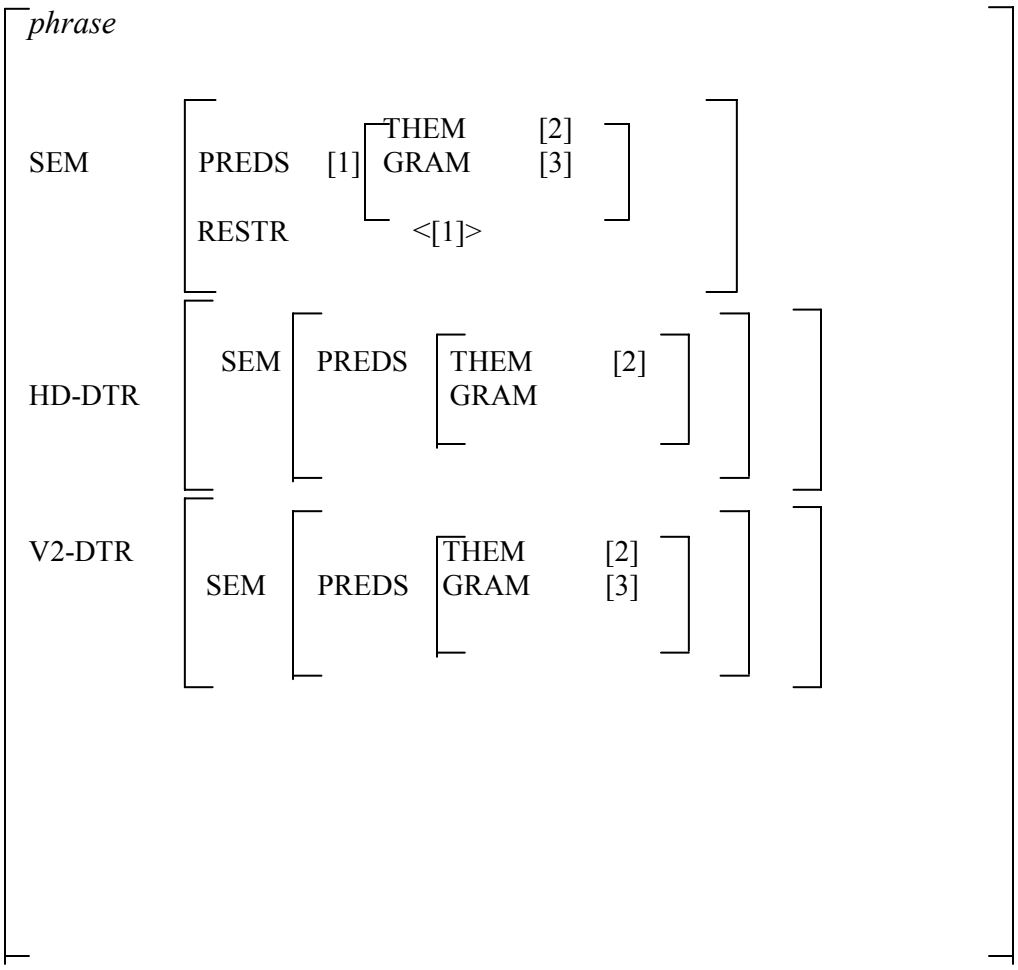


Figure: 18

6.4.4 Syntactic schema for building phrase structure of CV sequence

In the overview section I argued in favor of taking V1 as the head daughter of the CV phrase structure. I postulate a *pole verb word* type that represents the head daughter (V1) of the phrasal structure of CVs. I proposed that the categorial contribution that the V2 has to make to the syntactic description of the CV phrase structure is copied to the HEAD value of the *pole verb word* type at the word level. This approach has the following two significances:

1. The constraint of the *head feature principle* on the phrasal category is maintained
2. At the syntactic level, the V2 only satisfies the *vector* requirement of the *pole verb word*

The *head feature principle* requires the categorial information associated with the HEAD feature of the head daughter is shared with its mother in a headed-phrase[9]. This can be stated in terms of the unification of HEAD specifications as shown in the following representation of a headed phrase:

$$\begin{bmatrix} \textit{headed phrase} & \\ \text{SYN} & [\ \text{HEAD} \quad [1]\ ]\ ] \\ \text{HD-DTR} & [\ \text{SYN} \quad [\ \text{HEAD} \quad [1]\ ]\ ] \end{bmatrix}$$

Figure: 19

The categorial information is borne by the V2 constituent of a CV. By copying this information on the head feature of the V1 participant through the feature HD at the word level (see 6.2.1) I ensure that the HEAD feature for the CV phrasal structure can be constructed from that of the head daughter alone without any direct head feature contribution from the V2, which is supposedly the non-head participant, at the phrase structure level. In this way we maintain the *head feature principle* even while passing information to the HEAD of a phrasal sign from a non-head daughter's HEAD structure.

The *head feature principle* together with the *semantic compounding principle* stated in section 6.4.3 and the *principle of orthography* that specifies that the ORTH value of a phrasal structure will be the listing in a linear order the ORTH value of constituent daughters, yield a

[9] A headed phrase is a certain type of phrase in which "one daughter is assigned a special status as the head daughter" (Sag et al. 1999, p. 60)

well-formed headed phrase which is used as an input to a grammar rule, which accomplishes the following task:

1. The SPR and COMPS value of the lexical head is passed up to its mother.
2. The head daughter's VEC list is cancelled off.

I refer to the new grammar rule as *compound rule*. This grammar rule actually builds the phrase structure of the CV. The rule corresponds to the following constraint on the phrasal structure:

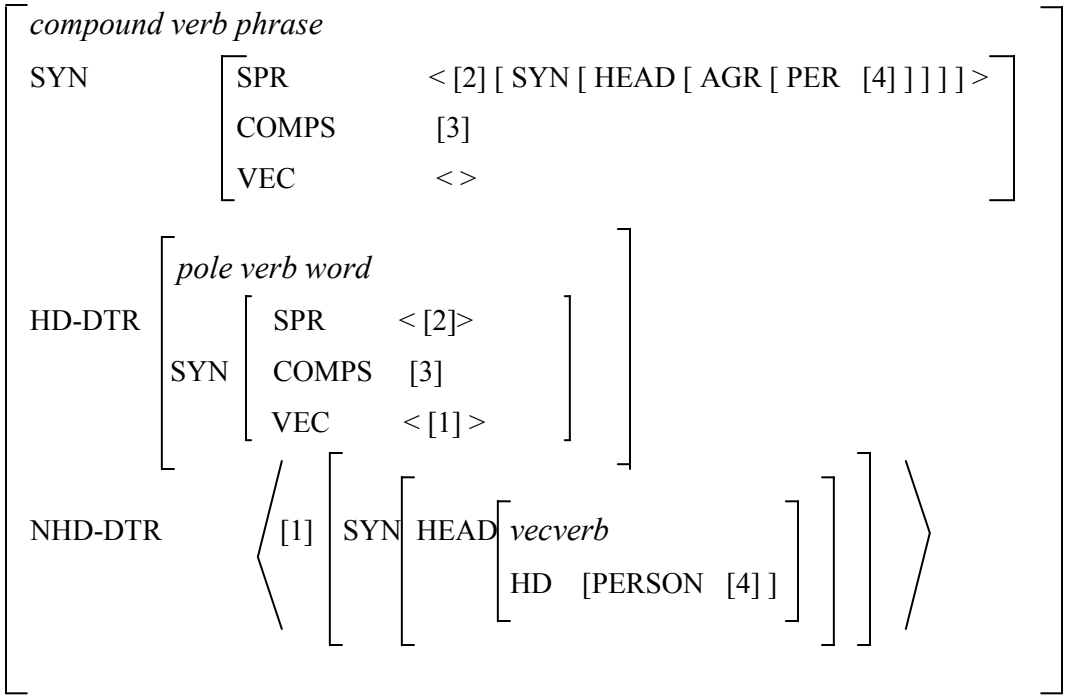


Figure: 20

This rule also imposes a constraint on the agreement feature of the SPR value of the mother. As shown in figure 20, the SPR argument of the CV must agree in person with its V2 constituent.

6.5 Conclusion

In this chapter, I proposed a semantically grounded *periphrastic compounding* mechanism that composes the compound verb constructions by unifying the two semantically compatible Vs. The system works on the basis of a semantic constraint that is imposed on semantic *predicators* arranged in a multiple inheritance hierarchy network. I chose the option of making V1 the head daughter of the phrase structure of CV and the categorial and semantic information that the V2 contributes to the structural description of CV is passed up through V1's lexical description. In this context I discussed the alternative approaches in the overview section. Finally I presented a description of the implementation of this grammar in LKB platform that has been undertaken in the present project.