

Chapter 1

INTRODUCTION

1.0 Objective

The thesis proposes a constraint-based mechanism for constructing Bangla¹ Compound Verbs within the framework of Head-Driven Phrase Structure Grammar (henceforth HPSG). Since the design of the constraint heavily relies on a model involving multiple inheritance hierarchy system in the lexicon, this work will devote some attention to the large-scale organization of the lexicon. Finally I will check the feasibility of the analysis developed for Bangla compound verbs in this thesis through implementation using the LKB (Language Knowledge Building) system. The LKB is a highly specialized program used extensively for building parser for natural languages.

1.1 Domain of Investigation

A wide variety of multi-verb constructions are attested in Bangla and other Indo-Aryan languages. Among them I have chosen a special kind of multi-verb construction that is commonly known as Compound Verb (henceforth CV) in the literature.

1.1.1 Definition of compound verbs

A Bangla CV sequence is a two-member construction in which the first member (V1) chooses between the usual conjunctive participial form *-e* and the (rare, for this construction) infinitive form *-te* while the second member (V2) bears the inflection. For simplicity it is normal to consider finite examples:

¹ Bangla (also known as Bengali) is an Indo-Aryan language spoken in Bangladesh and the Eastern Region of India.

1. *aša ækmaš-e æk-ṭa goṭa boḷ **likh-e phel-lo***
Asha one month-loc one-cl whole book write-cp drop-3 pt
‘Asha wrote a whole book within one month’

2. *mee-ṭa **heš-e uṭh-lo***
girl-cl laugh-cp rise-3 pt
‘The girl burst into laughter’

The boldfaced sequences of verbs are the CVs. I label these expressions as *periphrastic compounds*² in my work. The V2s are selected from a closed set of verbs and they are variously termed as vector, light verb, explicator and auxiliary³. The ‘V1-*te* V2’ sequences in the sentences (3a) and (3b) are akin to a CV in the following two senses:

- i. The semantics of the V2s is bleached off as that in case of CVs and
- ii. The V2 adds a semantic nuance to the whole cluster.

The following are examples of ‘V1-*te* V2’ type CV constructions:

3a. *buṛo lok-ṭa **mor-te boš-eche***
old man-cl die-inf sit-3 pr pft
‘The old man is about to die’

² Periphrastic compounds are, on the one hand, distinct from synthetic expressions such as *kor-echilam* (“do-1 pt pft”) ‘(I) had done (it)’ in which one word-form corresponds to one meaning and on the other hand, from those multi-word expressions in which the second verb plays the role of an auxiliary. The V2s within periphrastic compounds are not as completely grammaticalized as is the case for auxiliaries (for illustration see sentence (19a) in section 1.1.4).

³ Hook (1974) is one of those scholars who consider the V2 as an *auxiliary* because he assumes that it only contributes aspectual information to the CV construction and that it does not contain any information pertinent to the syntax. Abbi (1991, 1992) calls it an *explicator*. Butt (1995) chooses the term *light verb*. They argue that V2s do not get completely delexicalized. They have shown that the V2s in Hindi-Urdu sometimes determine the case marking on the subject of the CV sequence. Many scholars have used the term *vector* (Dasgupta (1989), Mohanty (1992), to name a few). In the present work, I will refer to this verb as V2 in general and as a *vector* in cases when a precise designation is required.

b. *tui̇ mitu-ke kɔ̃tha-ʈa bol-te ge-li kæno?*
 you Mitu-obj word-cl say-inf go-2 n-hon pt why
 ‘Why did you have to say this to Mitu?’

I shall not include these constructions in my discussion.

1.1.2 Homotactic sequences

Every Indo-Aryan language including Bangla contains verb sequences that are formally similar to the above CVs in the following respects:

1. The morphosyntactic forms of the first member of the sequences under consideration are identical. In Bangla, for example, the first member of these verbal sequences occurs in conjunctive participial form or in infinitive form. In Hindi, they are bare verb forms.
2. The second member always bears the inflection for tense and person and other grammatical categories such as number and gender.
3. The verbal sequence, as a whole, licenses the external argument that occupies the subject position in the sentence.

Thus the verbal sequences in the sentences in (4) are morphosyntactically similar to CV (‘V-e V’) constructions:

4a. *chele-ʈa amar dike chuʈ-e e-lo*
 boy-cl I-gen towards run-cp come-3 pt
 ‘The boy came running towards me’

b. *bacca-ʈa dudh khe-e ghumo-lo*
 child-cl milk eat-cp sleep-3 pt
 ‘After drinking milk the child slept’

Hook (1974) makes the following observation in relation to Hindi-Urdu CV constructions and it applies to verb sequences of other Indo-Aryan languages as well:

There exist homotactic sequences or Verb+Verb sequences, which are formally congruent to compound verbs, although they must be kept distinct from the compound verb sequence.

For example, the ‘V-*e* V’ sequences in the sentences in (4) are not CV constructions and they are to be separated from legitimate CVs. I presume that the following syntactic and semantic behaviors of the CVs secure a unique status for this construction in the language.

1.1.3 Characteristics of Indo-Aryan compound verbs

The following are some salient features of the Indo-Aryan CVs:

1.1.3.1 CVs are mono-clausal units

The participant-verbs in a CV sequence constitute one functional-semantic unit. For example, consider the sequence of Vs in (5). It does not express two separate events, ‘eating’ and ‘taking’. Instead the verbal sequence indicates that Ritu took her meal; the activity is completed and the result of the action is directed towards the actor Ritu in the sense that Ritu is a beneficiary:

5. *ritu khabar khe-e⁴ ni-lo*

Ritu meal eat-cp take-3 pt

‘Ritu took (her) meal (implying that Ritu is self-beneficiary)’

⁴ Under the influence of the suffix vowel that is originally derived from a high vowel (for example, the conjunctive participial form *-e* from the archaic form *-ia*), the root final vowel is heightened (a process known as vowel harmony). That is why the base *kha-* ‘eat’ is phonologically represented as *khe-* within the word *khee* ‘eat-cp’ in (5).

The following observations confirm that the participant verbs within CV sequences do not head separate clauses:

i) The event represented by a CV cannot be substituted by two conjoined independent events as shown in (6):

6. **ritu khabar khe-lo ebɔŋ ni-lo*

Ritu meal eat-pt and take-3 pt

‘Ritu ate meal and took (it)’

ii) The members of the sequence cannot be negated or modified separately. For example the CV sequence in (7) cannot be negated in the following manner with the first *na* ‘not’ taking scope over the verb *khee* ‘eat-cp’ and the second *na* ‘not’ over the second verb *nilo* ‘take-3 pt’:

7. **ritu bhat na khe-e ni-lo na*

Ritu rice not eat-cp take-3 pt not

1.1.3.2 Semantics of CV is an extension or modification of that of its V1 constituent

When a V1 participates in periphrastic compounding, the semantics of the V1 is extended or modified and the resultant CV acquires a meaning subtly distinct from that of the V1 itself. For example, let us consider the intransitive verbal root *ghumono*⁵ ‘sleep’. This verb selects the V2s *pɔra* ‘fall’, *neqa* ‘take’, and *jaqa* ‘go’. Column 3 in the following table represents the CVs. The semantic overtone that the V2s add to the meaning of *ghumono* ‘sleep’ is summarized in column 4:

⁵ I shall use the verbal noun form of verbs as the citation form

V1 'sleep'	V2	CV	Semantic overtone
<i>ghumono</i>	<i>pɔ̄ra</i> 'fall'	<i>ghumie pɔ̄ra</i>	A sense of immediateness ⁶
<i>ghumono</i>	<i>nēqa</i> 'take'	<i>ghumie nēqa</i>	Participant is self-beneficiary
<i>ghumono</i>	<i>jāqa</i> 'go'	<i>ghumie jāqa</i>	Extendedness of the event

Table: 1

The following sentences illustrate the use of CVs listed in column 3 of table 1:

8a. *mee-ṭa ghumi-e por-lo*

girl-cl sleep-cp fall-3 pt

'The girl fell asleep'

b. *mee-ṭa khanikkhon ghumi-e ni-lo*

girl-cl for a while sleep-cp take-3 pt

'The girl slept for a while (implying that the girl is self-beneficiary)'

c. *mee-ṭa šaradin šudhu ghumi-e gæ-lo*

girl-cl entire day only sleep-cp go-3 pt

'The girl kept on sleeping all day long'

These CVs demonstrate that the core meaning of *ghumono* 'sleep' is preserved in every case. The V2s, *pɔ̄ra* 'fall', *jāqa* 'go' and *nēqa* 'take' contribute specific semantic nuances to the meaning of *ghumono* 'sleep' as a result of which each resultant CV acquires a meaning subtly distinct from its V1 associate.

⁶ The connotation of these semantic overtones will be discussed in detail in chapter 3

1.1.3.3 Semantics of V2

Every V2 has an independent counterpart attested in the language. Scholars have unanimously maintained that a verb that occupies V2 position in a CV construction undergoes a semantic loss. Hook takes an extreme position and states that V2s become lexically empty. He describes the phenomenon as *grammaticalization*⁷ (Hook 1974, p. 94-97). Others confer a semi-lexical status on V2s⁸. They argue that a V2 has a different status from an auxiliary because it is not as grammaticalized as the auxiliaries. In chapter 3, I shall present a review of how different scholars have interpreted the nature of *semantic loss*. Nevertheless it is now generally agreed upon that the V2s add semantic nuances to the meaning of the construction as a whole. The thesis will attempt to present a precise semantic representation for every V2 in Bangla.

1.1.4 Tests for identifying CV sequences

Based on the behaviors of the CV sequences as demonstrated in the previous section, I design the following set of tests, which will presumably separate out CV sequences from other homotactic non-CV sequences.

⁷ Sarkar (1975) elucidates Porizka's perception of grammaticalization – a stripping of the main dictionary meaning from the vector verb in order to reduce it to the role of 'aspective'.

⁸There is a great deal of discussion available in the literature regarding the semantics of V2. The following are some references related to the works on Indo-Aryan compound verb structure: Hook (1974), Sarkar (1975), Dasgupta (1989), Abbi (1991,1992), Mohanty (1992), Bashir (1993) and Butt (1995).

1. Does the verbal sequence retain the core meaning of the V1 constituent?

If yes: The sequence might be a CV construction

If no: The sequence is most likely to be an idiomatic expression

2. Does the second constituent (V2) of the verbal sequence lose the *primary lexical meaning*?

If yes: Is the semantics associated with the V2 related in some way to the meaning of its full-verb counterpart?

If yes: The sequence is a CV construction.

If no: The sequence is a cluster of main verb and auxiliary.

If no: The sequence is not a CV construction and the second member independently expresses an event

3. Can the verbs in a sequence never be individually negated or modified by adverbs?

If yes: The sequence is a CV construction

If no: The sequence is not a CV construction

4. Can the verbs in a sequence never be separated by coordinating conjunction?

If yes: The sequence is a CV construction

If no: The sequence is not a CV construction

The first two tests are **semantic tests**. Test 3 and 4 check whether the two verbs together form one functional-semantic unit. I refer to them as **tests for predication**. Positive response to tests 2, 3 and 4 qualifies the candidature of a ‘V-e V’ sequence for CV construction. The second semantic test (i.e., test 2) requires careful examination. By the term **primary lexical meaning** I refer to the sense in which the verb is customarily used. For instance, the primary lexical meaning of the verb *mōra* is ‘die’, i.e., physical death.

However, we all know that a verb might have many uses. The verb *mɔra* ‘die’, for example, can occur in the following different contexts:

9a. *lok-ʔa šesporjonto pɔth-er dhare mor-lo*
 man-cl finally road-gen side die-3 pt
 ‘Finally the man died on the roadside’

b. *gach-ʔa mor-lo kikore?*
 tree-cl die-3 pt how
 ‘How did the tree wither away?’

c. *ga-er bætha kichute-i mor-che na*
 body-gen pain by any means-emp die-3 pr cont not
 ‘The pain just isn’t disappearing’

d. *ami mor-chi nijer jala-e*
 I die-1 pr cont own agony-loc
 ‘My own worries are killing me’

In chapter 3, I shall propose that the various uses of a verb share one common abstract concept, which I will term as **mould** or **core sense**. For instance the following is the core sense of the verb *mɔra* ‘die’: ‘the non-voluntary engagement of a participant in a situation (notionally ‘unpleasant’) over which the participant has no effective control’. In this case the verb’s primary lexical meaning, namely ‘physical death’, is one extreme manifestation of the above core sense. Semantic test 2 proposes that a verb does not retain the primary lexical meaning when it occurs in the V2 position within a CV. Nevertheless it shares the same **core sense** with its corresponding full-verbs. Thus the relation between the vector use of a verbal item and its use as a full verb is identified as a case of polysemy. An auxiliary, on the other hand, is stripped of the **core sense** altogether. Hence the auxiliary becomes entirely distinct from its independent counterparts.

I will apply the above set of tests to the following set of sentences. For instance, the two verbs in the following sentences express a modifier-modified relation:

10. *chele-ṭa amar dike chuṭ-e e-lo*

boy-cl I-gen toward run-cp come-3 pt

‘The boy came running towards me’

11. *balti theke jol upc-e por-che*

bucket from water overflow-cp fall-3 pr cont

‘Water is overflowing from the bucket’

12. *mee-ṭa cupcap boš-e ach-e / thak-e / rɔ-ɛ*

girl-cl quietly sit-cp be-3 pr / stay-3 pr / remain-3 pr

‘The girl sits quietly’

The second verb of the verbal sequence in (10) and (11) expresses the main event. The first member in both cases plays the role of a modifier only. The second verb within the verbal sequences in (12) conveys the meaning of ‘to be in a state’ and the first verb *boše* ‘sit-cp’ specifies the kind of the state the subject belongs to. For example, the V-sequence *boše thaka/rɔɔa* ‘sit-cp stay/ remain’ implies that the girl is ‘in a state of sitting’.

When the verbs *ach* ‘be’, *rɔɔa* ‘be, remain’ and *thaka* ‘stay’ occur independently in a sentence as shown in (12a) they express a similar meaning, ‘to be in a state’:

12a. *mee-ṭa cupcap thak-e / ach-e / rɔ-ɛ*

girl-cl quiet stay-3 pr / be-3 pr / remain-3 pr

‘The girl remains quiet’

Since the second member of ‘V-e V’ sequences, as illustrated in (10), (11) and (12) retain their primary lexical meaning they respond negatively to the semantic test 2: “Does the second member of the verbal sequence lose their primary lexical meaning?” Therefore they fail to meet the criterion necessary to qualify as a CV.

Sarkar (1975) and Dakshi (2000) consider the verb *dækha* ‘see’ to be a legitimate V2. Let us consider the use of the verb *dækha* ‘see’ in the following verbal sequences:

13a. *chobi-gulo-ḡ ækbar hat die dækh-o, tomar ki dōṣa kor-i*
 picture-cl-loc once hand give-cp see-2 pr imp you-gen what state do-1 pr
 ‘Touch the pictures once (and) see what I do to you’

14a. *kaj-ṭa kōdin kor-e dækh-o, bhalo na lag-le cher-e di-o*
 work-cl some days do-cp see-2 pr imp good not feel-cond leave-cp give-3 ft imp
 ‘Do the work for some days (and) see, (if) you do not like (it) give (it) up’

The verbal sequence (*hat die dækha*) “(hand) give-cp see” in (13a) intimates a sense of *threatening*. The speaker warns the addressee not to touch the pictures or else he/she will have to face the consequences. In (14a) the verb *kore dækha* “do-cp see” entails *wait and see*. The addressee is being advised to continue with his/her work for some days before he/she decides whether to quit it or not. The sentences in (13b) and (14b) express meanings similar to those in (13a) and (14a) respectively:

13b. *chobi-gulo-ḡ ækbar hat da-ḡ, dækh-o tomar ki dōṣa kor-i*
 picture-cl-loc once hand give-2 pr imp see-2 pr imp you-gen what state do-1 pr
 ‘Touch the pictures once and see what I do to you!’

14b. *kaj-ṭa kōdin kōr-o, dækh-o, bhalo na lag-le cher-e di-o*
 work-cl some days do-2 pr imp see-2 pr imp good not feel-cond leave-cp give-3ft imp
 ‘Do the work for some days (and) see, (if) you do not like (it) give (it) up’

The verbal sequences of the sentences in (13a) and (14a) are represented as conjoined sequences of two independent verbs in (13b) and (14b). Therefore, I do not consider ‘*V-e dækha*’ to be a CV construction. The verb *dækha* ‘see’ is not, however, used in its customary sense of physical perception in above sentences.

The second member of the verbal sequences in (15a), (16a), (17a) and (18a) also retains its primary lexical meaning in each case and the two verbs represent independent events as illustrated in (b) sentences:

15a. *ritu krititte-r šathe kaj-ṭa kor-e e-lo*

Ritu success-gen with work-cl do-cp come-3 pt

‘Ritu did the work successfully’

b. *ritu krititte-r šathe kaj-ṭa kor-e taratari baṛi e-lo*

Ritu success-gen with work-cl do-cp quickly house come-3 pt

‘Ritu came back home quickly after successfully completing (her) work’

The adverb *taratari* ‘quickly’ in (15b) modifies the event of coming back home. The event denoted by the verb *kore* ‘doing (the work)’ is modified by a separate adverbial phrase *krititter šathe* ‘successfully’. Thus the two events represented by the two verbs are independently modified by separate adverbial phrases. When the two verbs in a ‘V-e V’ sequence represent two separate and consecutive events the first verb form expresses perfectivity.

16a. *ritu ghoṛi-ṭa šari-e en-eche*

Ritu clock-cl repair-cp bring-3 pr pft

‘Ritu has repaired the watch and brought (it)’

b. *ritu ghoṛi-ṭa šari-e taratari baṛi an-lo*

Ritu clock-cl repair-cp quickly house bring-3 pt

‘Ritu repaired the watch and brought (it) home quickly’

17a. *nilu tɔktoɔš-e eš-e boš-lo*

Nilu bed-loc come-cp sit-3 pt

‘Nilu came and sat on the bed’

b. *nilu e-lo ebɔŋ tɔktoɔš-e boš-lo*

Nilu come-3 pt and bed-loc sit-3 pt

‘Nilu came and sat on the bed’

18a. *chele-ra šuto ŋen-e kākra-ŋa ɖaŋa-ɛ en-e tul-lo*

boy-cl thread pull-cp crab-cl ground-loc bring-cp lift-3 pt

‘The boys pulled the thread and bought the crab to the shore’

b. *chele-ra šuto ŋene kākra-ŋa an-lo ebɔŋ ɖaŋa-ɛ tul-lo*

boy-cl thread pull-cp crab-cl bring-3 pt and ground-loc lift-3 pt

‘The boys pulled the thread and bought the crab to the shore’

The sentences in (15-18) show that the verbal sequences under consideration respond negatively to the semantic test (Test 2) and the test for predication (as illustrated in the (b) sentences). As a consequence these ‘V-e V’ sequences fail to meet the criterion for being a legitimate CV sequence. Interestingly the second member of many of the verbal sequences illustrated in (10-18) also occupies the V2 position within CV sequences. The next section will demonstrate their role as V2.

Let us now consider the following sentence:

19. *chele-ŋa bɔŋobabu-ke kɔtha-ŋa bol-e thak-be*

boy-cl boss-obj word-cl tell-cp stay-3 ft

‘The boy might have said (this) to the boss’

Like a V2 in a CV sequence, the verb *thaka* ‘stay’ does not retain the lexical meaning in *bole thakbe* ‘might have said’. Now the second part of the second test is applied to the verbal sequence, namely, “Is the semantics associated to the verb related in some way to the meaning of its full-verb counterpart?” As the sentence in (19) shows, the verb *thaka* ‘stay’ has become fully grammaticalized. It expresses a sense of modality. Therefore I refer to *thaka* ‘stay’ in (19) as an auxiliary.

We shall see in the next section that the V2s in CV sequences, like auxiliaries, undergo semantic loss but they are not completely grammaticalized. From that perspective V2s remain distinct from auxiliaries in the language. There are quite a number of borderline cases where it becomes really difficult to determine whether the ‘V-e V’ construction under consideration is a CV sequence or not.

1.2 A Repertoire of Legitimate CV Sequences in Bangla

The verbs that occupy the V2 position within CV sequences are selected from a closed set of verbs. Scholars have noted that the frequency of occurrence of compound verb varies from one language to another within the Indo-Aryan subgroup of languages. The attestation of the compound verb is most frequent in Hindi-Urdu (Hook 1974), while it is very rare in Kashmiri (Kaul 1985). Bangla occupies the third position in the scale of frequency. I have identified the following frequently used V2s for the present work:

<i>aša</i> ‘come’	<i>rakha</i> ‘keep’	<i>ana</i> ‘bring’	<i>deḡa</i> ‘give’
<i>pṛa</i> ‘fall’	<i>paṭhano</i> ‘send’	<i>bæṛano</i> ‘roam’	<i>neḡa</i> ‘take’
<i>tola</i> ‘lift’	<i>bṛša</i> ‘sit’	<i>oṭha</i> ‘rise’	<i>jaḡa</i> ‘go’
<i>phæla</i> ‘drop’	<i>mṛa</i> ‘die’	<i>cṛla</i> ‘move’	

In chapter 3, I shall argue in favor of the polysemous existence of the V2s *deḡa* ‘give’, *jaḡa* ‘go’ and *aša* ‘come’ in Bangla. Each of the variants is given a separate lexical representation in our grammar. The use of each of the above V2s within a CV construction is illustrated in (20):

20a. *šondhe ho-e e-lo*
evening become-cp come-3 pt
'It is nearing evening'

The CV *hoe aša* 'coming to be' in (20a) conveys the sense: 'the evening is approaching'. The other variant of the V2 *aša* 'come' expresses extendedness of an event as shown in (20b):

20b. *tomake kɔbe theke kɔtha-ʔa bol-e aš-chi*
you-obj when since word-cl say-cp come-1 pr cont
'Since ages I have kept telling you this'

The V2 *deqa* 'give' as illustrated in (20c) entails that 'a causer acts upon an *affected entity* so that the effect of the action is 'directed' towards an *entity* other than the causer'.

20c. *ritu binu-ke æk-ʔa baʔi khūj-e di-lo*
Ritu Binu-obj one-cl house search-cp give-3 pt
'Ritu found a house for Binu'

Ritu, the actor in (20c), finds out a house for Binu. Here Binu is the beneficiary. The other variant of *deqa* 'give' implies that an actor is engaged in some activity. This variant of *deqa* 'give' does not entail the presence of an *affected entity* in the situation described. The actor Roghu in (20d) performs an activity.

20d. *roghu ašor-e gan ge-e di-lo*
Roghu concert-loc song sing-cp give-3 pt
'Roghu sang a song in the concert'

The V2 *neḡa* ‘take’ entails that the result of the action is directed toward the actor itself:

20e. *ami age bol-e ni-i tarpɔr tomar pala*
I first say-cp take-1 pr then you-gen turn
‘Let me say first, then you take your turn’

The actor *ami* ‘I’ is in focus in (20e).

The V2 *jaḡa* ‘go’ in the following sentence entails that the participant involved in the event undergoes a change of state and the whole construction expresses perfective aspect.

20f. *tak theke por-e glaš-ṭa bheṅ-e gæ-lo*
shelf from fall-cp glass-cl break-cp go-3 pt
‘The glass fell down from the shelf and broke’

The entity *glaš* ‘glass’ undergoes a change of state as it falls down from the shelf. The other variant of V2 *jaḡa* ‘go’ denotes extendedness of an event and it expresses an aspectual sense of imperfectivity as illustrated in the following sentence:

20g. *ritu šaradin ghɔr-er ækkon-e boš-e mon die chobi ěk-e gæ-lo*
Ritu whole day room-gen one corner-loc sit-cp mind with picture draw-cp go-3 pt
‘Sitting in a corner of the room Ritu continued to draw the picture attentively’

The actor Ritu is involved in an activity of drawing and the V2 implies that the event continues.

The CV sequence ‘V-e *pɔra*’ as illustrated in (20h) implies that Ritu attains an inactive state as she falls asleep. The event is resultative and the focus is on the immediate effect of the event.

20h. *ritu ghumi-e por-lo*

Ritu sleep-cp fall-3 pt

‘Ritu fell asleep’

The V2 *tola* ‘lift’ entails that the V1 event is accomplished and the V2 focuses that subpart of the V1 event, which precedes the culmination point. This is illustrated in (20i):

20i. *ribhu ašor jomi-e tul-lo*

Ribhu party warm up-cp lift-3 pt

‘Ribhu enlivened the party’

The V2 *rakha* ‘keep’ indicates that the action denoted by the V1 event is accomplished well in advance. The CV construction presents the semantics of resultative aspect.

20j. *ritu šob kaj kor-e rakh-lo*

Ritu all work do-cp keep-3 pt

‘Ritu completed all her work in advance’

The V2 *paṭhano* ‘send’ causativizes the event denoted by the V1 with which it occurs as exemplified in (20k).

20k. *ritu ribhu-ke die rōma-ke ḍeke paṭha-lo*

Ritu Ribhu-obj through Ramaa-obj call-cp send-3 pt

‘Ritu summoned Ramaa through Ribhu’

The CV *ḍeke paṭhano* ‘summon’ in (20k) implies that the actor Ritu entrusts Ribhu with the task of calling Ramaa.

The V2 *bōṣa* ‘sit’ indicates the accomplishment of the V1 event. It also implies the unwarrantedness of the event. For instance, the sentence in (20l) entails that the actor Ritu says something amidst all which is unexpected and unwarranted.

20l. *ritu ṣobar majhe kōtha-ṭa bol-e boṣ-lo*
 Ritu all-gen among word-cl say-cp sit-3 pt
 ‘Ritu went and said it in the middle of the crowd⁹’

The V2 *phæla* ‘drop’ in (20m) highlights the ability of the actor Ritu in completing her work within few hours. The V2 adds a sense of perfectivity to the meaning of the CV.

20m. *ritu kōek ghōṇṭa-r moddhe kaj-ṭa kor-e phel-lo*
 Ritu few hour-gen within work-cl do-cp drop-3 pt
 ‘Ritu finished her work within few hours’

The V2 *ana* ‘bring’, like the V2 *aṣa* ‘come’ (illustrated in (20a)), implies that the event denoted by the V1 is approaching to its end or completion. For instance, the actor Ritu in (20n) has almost arranged the house.

20n. *ritu ghōrdor praḡ guchi-e en-eche*
 Ritu house almost arrange-cp bring-3 pr pft
 ‘Ritu has almost arranged the house’

The V2 *bæṛano* ‘roam’ adds a sense of iterativity to the meaning of CV of which it forms a part. For instance, in (20o) the actor roams around showing this to be a trait of her habitual make-up.

⁹ The construction ‘go’ and VERB is a close idiomatic match for the CV V1-*e bōṣa* ‘sit’

20o. *ritu ghur-e bæra-e*

Ritu loiter-cp wander - 3 pr

‘Ritu roams idly’

The V2 *oṭha* ‘rise’ implies the commencement of the V1 event when the two verbs combine. The resultant CV expresses inceptive aspectual information.

20p. *ritu hoṭhat heš-e uṭh-lo*

Ritu suddenly laugh-cp rise-3 pt

‘Ritu suddenly burst into laughter’

The V2 *mōra* ‘die’ implies futility of the activity that the actor is involved in. For example, the following sentence entails that the actor *ma* ‘mother’ has no other choice but giving her all for her family.

20q. *ma šarajibon šonšar-er jonne kheṭ-e mor-len*

mother whole life family-gen for work hard-cp die-3 hon pt

‘Mother worked the whole of her life for (her) family (implying a sense of futility)’

The CV that has *cōla* ‘move’ as its V2 constituent implies that the event denoted by its V1 associate continues as illustrated in (20r):

20r. *lok-ṭa purodome kaj kor-e col-che*

man-cl in full swing work do-cp move-3 pr cont

‘The man continues to work in full swing’

Section 1 and 2 discuss the nature of CV sequences, their behavior and various identification tests that facilitate the task of isolating CV construction from the rest of ‘V-e V’ sequences.

1.3 Technical Background

In this subsection I will discuss some technical concepts and representational features that HPSG and other constraint based grammar formalisms utilize.

1.3.1 Feature structure

A feature structure is a structured way of representing information about an entity. It consists of a set of features (conventionally written in upper case), each of which is *mapped* (in the mathematician's sense of the word) to a particular value. The value can be a simple, structureless entity called atom, or it can be a feature structure, embedded in the larger one. For example, the categorical information associated with the pronoun *ami* '1 person singular pronoun' appears in the following feature structure description in figure 1. The description language used here for representing feature structure is similar to the attribute-value matrix (AVM) notation:

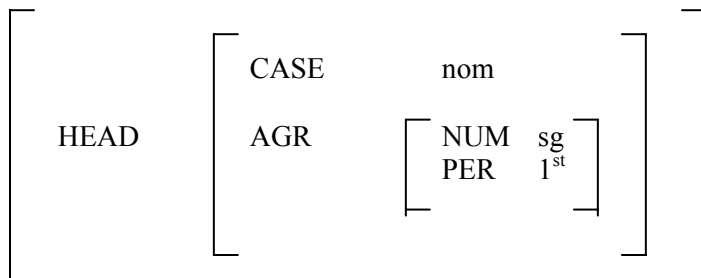


Figure: 1

As shown in figure 1, the value of HEAD is a complex feature structure that contains two features CASE and AGR(ement). The value of CASE is atomic while the value of AGR is a feature structure complex having specifications for two features NUM and PER.

Thus information about linguistic objects can be represented in terms of complex feature structures. We can imagine grouping these linguistic entities into classes to deterministically ensure which set of features is appropriate for a class. Unification-based natural language grammars have worked on this line and developed *typed feature structure* formalisms.

1.3.2 Typed feature Structure

Types are defined to have certain features that are appropriate only for entities designated by that type. This is to say that *types* specified in a linguistic domain tell us what kind of linguistic entities exist and the properties associated with these types tell us what general properties those entities exhibit. For example, we can ensure that the feature NUM can only be specified for certain kinds of linguistic entities or *types* such as those specifying nouns and pronouns and not, for instance, adverbs and prepositions. The syntactic description of the type *noun lexeme* is represented as a typed feature structure in figure 2:

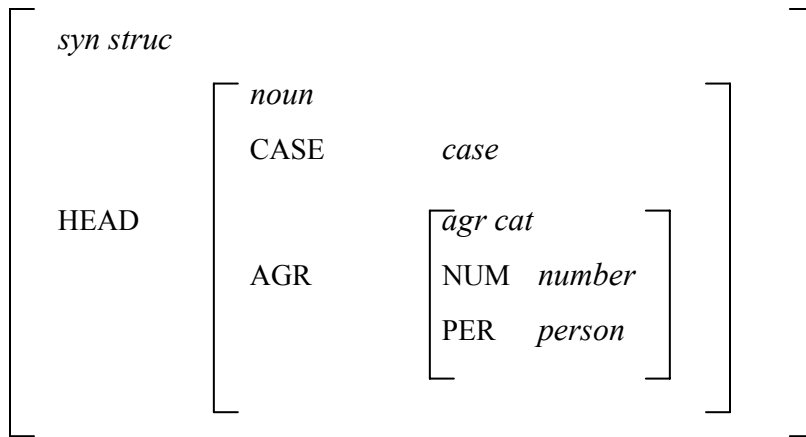


Figure: 2

The type name is customarily declared as a label, written in italics, on the top line within the feature structure. The value of HEAD for *noun lexeme* is declared to be of the type *noun*. Types are introduced in constraint-based grammar to define feature appropriateness and to avoid having to specify values for features that are irrelevant to particular type. Types are also used for stating constraints on feature structures. For example, the type *noun* is constrained to have two features, CASE and AGR. The feature AGR takes the value of type *agr cat* which itself designates a feature structure. Values of CASE, NUM and PER are described to be objects of specific types, *case*, *number* and *person* respectively. However these types are not maximally specific. Any individual member of the type *noun lexeme* will have more explicit feature structure representation. For instance, the noun *boj* ‘book’ is

described by the following feature structure that is type resolved and fully specified for the above features:

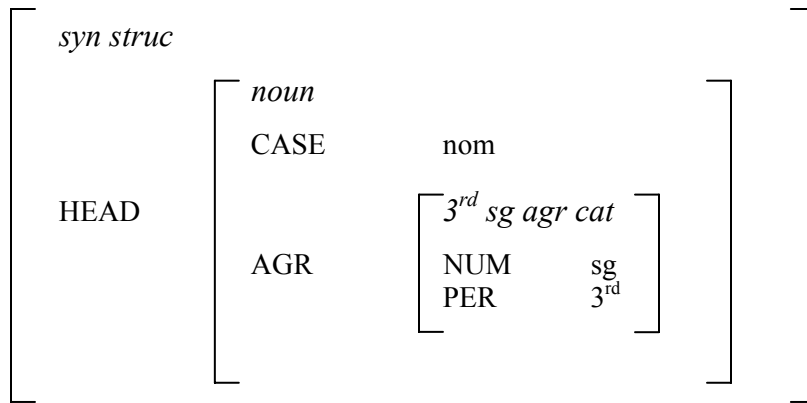


Figure: 3

In figure 3 the value of AGR is a more specific type that is a subtype of *agr cat*. This type constrains the value of NUM and PER to be ‘sg’ and ‘3rd’ respectively.

1.3.3 Type hierarchy

One advantage of classifying linguistic entities into types is that it enables the grammar writer to set the types up in a hierarchy structure so that properties, which hold for a type of objects, need only be declared once. If type *A* is a subtype of type *B*, all properties of *B* are inherited by default by *A*. Multiple inheritance hierarchy allows a type to be a subtype of more than one supertype, in which case the type inherits the non-conflicting properties of all its supertypes. The type hierarchy defines a *partial order* on the types and specifies which types are *consistent*. A set of types is said to be *consistent* if the members of the set share a common subtype. The unique greatest lower bound condition on *consistent types* allows feature structures to be typed deterministically. If two feature structures of types ‘A’ and ‘B’ are unified the type of the resultant feature structure will be ‘ $a \cap b$ ’, which must be unique if it exists. The significance of the greatest lower bound condition is that only feature structures with mutually consistent types can be *unified* (the concept of unification is discussed below).

1.3.4 Structure sharing

Structure sharing or reentrancy is a phenomenon by which values of two or more attributes are made token-identical, that is, they are declared to be of the same type. It is conventionally indicated by the identically numbered tags in AVM. This phenomenon is successfully utilized for defining constraints in the description of constraint-based grammar. I will present here a simple example case to illustrate structure sharing. In Bangla, the subject and the verb agree in person.

Subject	Suffix added to the verb in present tense
1 st person	/-i/
2 nd person	/-o/
3 rd person	/-e/

So one might postulate a constraint that states that the value of the PER feature of a verb and that of its argument that occupies the subject position is token-identical. One possible way for declaring this constraint on the description of the feature structure associated to a verb's lexical entry is the following:

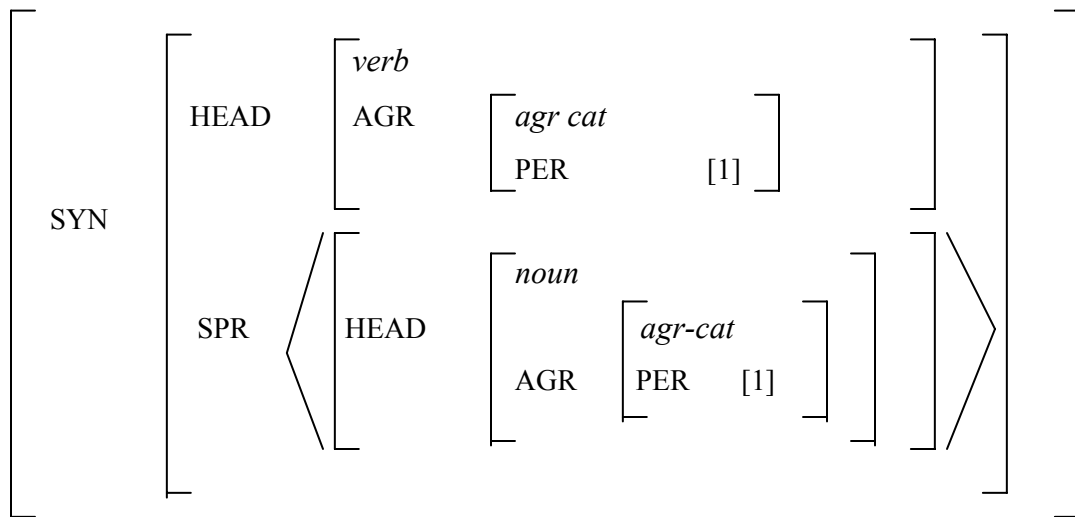


Figure: 4

The constraint requires that the PER value of a verb must be identified with that of the noun that satisfies the subject requirement of the verb.

1.3.5 Subsumption

The term subsumption refers to a relation among descriptions: a general description subsumes a more specific one. The more specific structure will always have all the features of the more general structure and may have additional features. For example, a singular noun is a specific type of noun. We can assume the following partial description for noun, which contains a feature NUM(ber) as shown below:

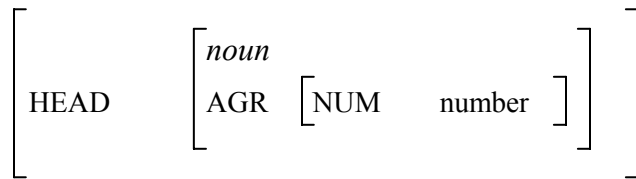


Figure: 5

This description subsumes the following one, which specifically mentions that the number is singular:

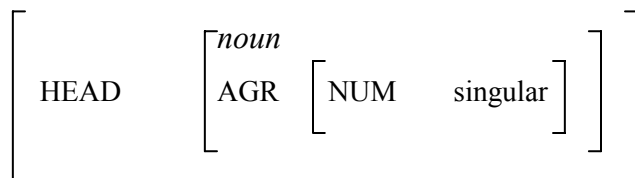


Figure: 6

The value ‘singular’ must be a subtype of ‘number’. Thus a feature structure ‘A’ *subsumes* another feature structure ‘B’ if for every feature F with a value *t* in ‘A’, there is a corresponding feature F in ‘B’ with a value which is either *t* or a subtype of *t*. And also every pair of features that are structure shared in ‘A’ must also structure-share in ‘B’.

We can have a type hierarchy in which the type *noun* is declared to subsume the type *singular noun*. The two types will be associated with the feature structures descriptions as illustrated in figure 5 and figure 6 respectively.

1.3.6 Unification

Unification is informally a function of combining two or more descriptions into a single coherent one.

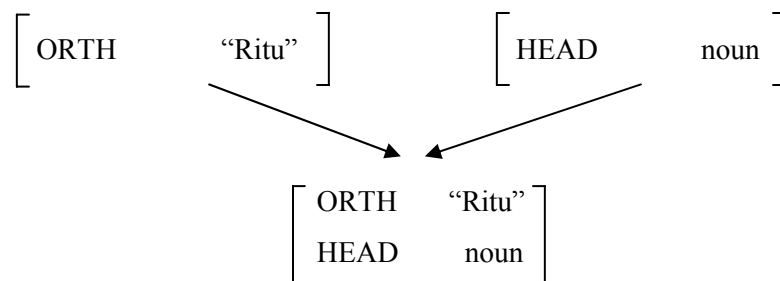


Figure: 7

The unification of two typed feature structures ‘A’ and ‘B’ results in a typed feature structure that contains the maximum information inherited from its supertypes. If there is a type ‘A’ and a type ‘B’ and both have a feature F whose value in A is *a* and in B is *b*, then A and B unify if and only if *a* and *b* are compatible types. For example the following two feature structures in figure 8 unify, while those in figure 9 do not:

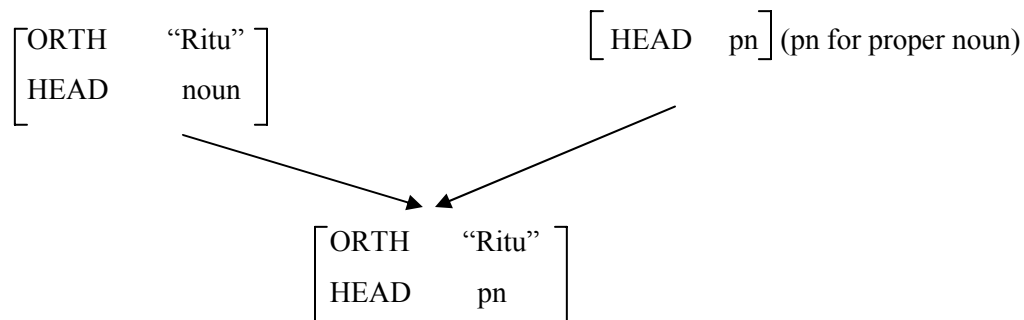


Figure: 8

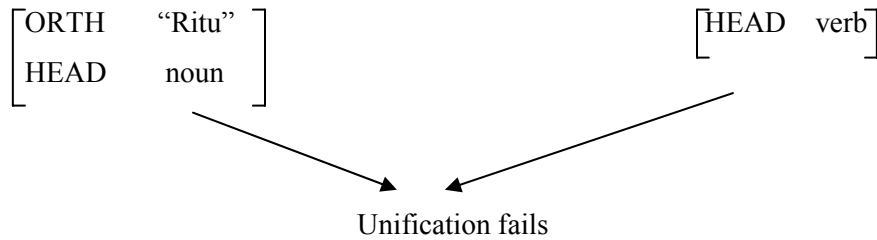


Figure: 9

For the two feature structures to be unified in figure 8, the value of HEAD, *noun* and *pn*, must be type compatible. In this case, *pn* is a subtype of *noun*. The two feature structures in figure 9 do not unify because the value of HEAD in the two cases are incompatible types. Thus unification of typed feature structure is always defined with respect to a particular type hierarchy.

This section discusses some major concepts related to the typed feature structure, type hierarchy and the unification function. In the next section we shall see that the formal device underlying the representations of linguistic description in HPSG employ a multiple inheritance hierarchy of typed feature structure and the system heavily relies on unification process.

1.4 A Brief Sketch of HPSG Framework

This section presents a brief sketch of relevant aspects of Head driven phrase structure grammar (HPSG) as presented in Sag and Wasow (1999) along with some modification proposed by Davis (2001). As noted by Sag et al., linguistic description in HPSG grammar is developed on the basis of the following three fundamental observations:

1. The grammar is strongly *Lexicalist* in the sense that most grammatical and semantic information have been localized within lexical entries. Rather than an unstructured set of lexical entries, HPSG treats the lexicon as a multiple inheritance hierarchy system.
2. It is constraint-based. The idea is that the grammar can be treated as sets of principles or constraints that determine what constitutes the legitimate linguistic sign in languages. These constraints apply simultaneously to a sign, so that the grammar is declarative rather than procedural.
3. It is surface oriented. The grammar is directly associated to the string of words that constitute sentences. No additional abstract structures are posited.

1.4.1 Representation of signs

HPSG is a linguistic theory based on *signs*. A *sign* in Saussurian sense is an arbitrary association of form with meaning. Linguistic signs such as lexical items, phrases, clauses and sentences are collections of multiple kinds of information, including phonological, syntactic, semantic and discourse constraints. The formal device underlying the representation of signs in HPSG are *typed feature structures*. They have at least the features PHON for phonology (or ORTH for orthographic representation) and SYNSEM for syntactic-semantic constraints as described in figure 10:

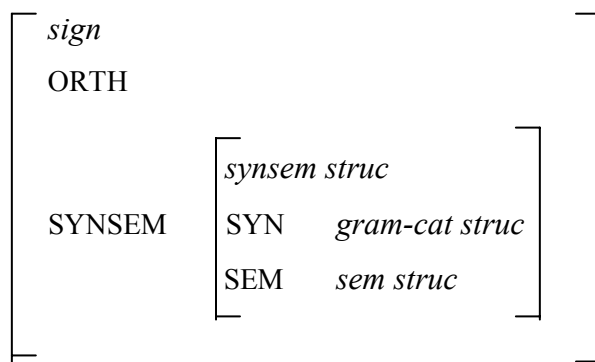


Figure: 10

The type *gram-cat struc*, the value of SYN, contains a feature HEAD. The feature HEAD always takes as its value a *part of speech (pos)* type. The major categories such as noun, verb, adjective, which can head a phrase are subtypes of the *pos* type. The HEAD of any lexical item identifies those properties of a sign, which are shared with the phrase it heads. Since we need to have agreement features specified on phrases as well as on their heads, AGR must be a head feature. The following figure represents a feature structure for the value of HEAD of a nominal entity:

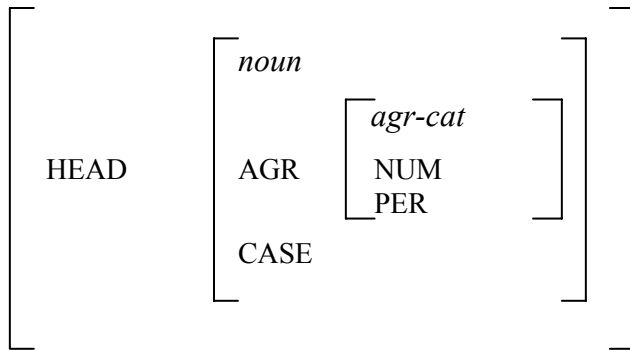


Figure: 11

Sag and Wasow (1999) maintain a fundamental distinction between the *lexeme* and *word* type at the representational level of lexical entries. The type definition for *lexeme* can be thought of as representing an abstract proto-word and the type *word* gives rise to the word structure. Thus ‘run’ is an abstract *lexeme* and their inflected word-forms such as ‘runs’, ‘ran’ and so on belong the type *word*. The word ‘runner’ can be taken as derivationally derived from the same *lexeme* type ‘run’. The following figure presents the type hierarchy for the basic representational levels in Sag et al. (1999):

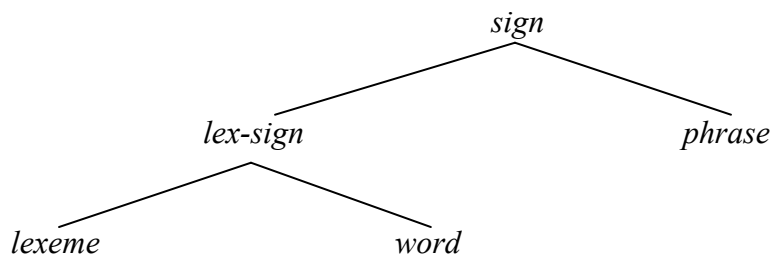


Figure: 12

The feature ARG-ST is introduced on the type *lex-sign*. Therefore every *lexeme* and *word* has an argument structure. The value of ARG-ST is an ordered list of *synsem struc* that denotes the *lex sign's* arguments. It represents the proper domain for stating several general constraints on argument structure. For instance, the theory of *binding* and also *linking* between the syntactic arguments and the semantic roles within SEM are defined on ARG-ST.

The Argument realization principle (ARP), which will be discussed in detail in the next section, relates the ARG-ST value to the values of the valence features SPR and COMPS on the *word structure*. The value of SPR is a one-member list. In canonical lexical entries of verbs, the concatenation of the SPR and COMPS lists yields the value of ARG-ST. Since I am not dealing with long distance dependency in the present work, the summation of the value of SPR and COMPS is taken to be value of the ARG-ST list.

I will introduce in this work one more feature VEC within the SYN of a verb's *lex sign*. The value of VEC will be a list that contains the *synsem struc* of V2s that a verb selects for in order to construct CV sequences.

The feature SEM represents the meaning component. Following Davis I model *semantic relations* as types themselves and not as value of an attribute RELATION as most of Pollard et al. works have treated them. For example, the value of RESTR for the lexical entry *kamṛano* 'bite' is presented in Figure 13 using the RELATION attribute and in Figure 14 as a subtype of the type *cause rel*, which is again a subtype of *act-und-rel*:

RELATION	<i>kamṛano</i>
BITER	
BITTEN	

Figure: 13

<i>kamṛano rel</i>	
ACT	biter
UND	bitten

Figure: 14

The advantage of modeling *semantic relations* as types is the following:

1. In a hierarchy model, types representing various *semantic relations* can be arranged in hierarchy (see Figure 15):

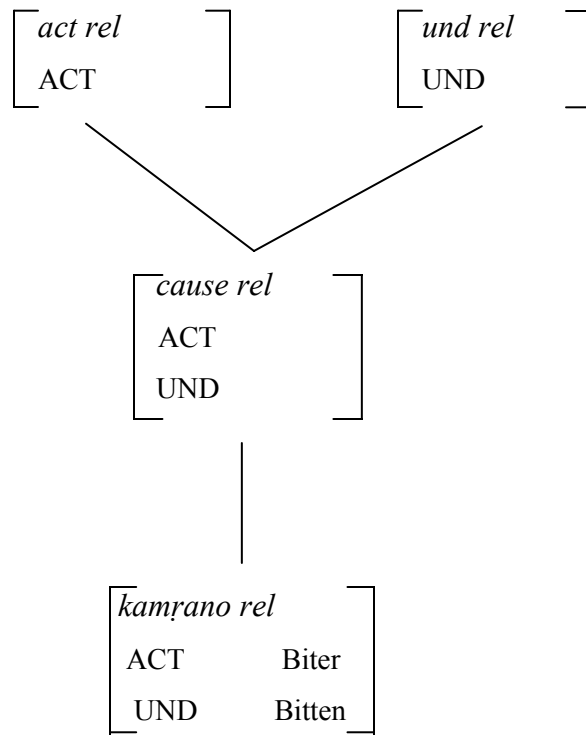


Figure: 15

The regularities captured in the semantic relations hierarchy are crucial because restrictions on more general types apply to more specific types as well.

2. The attributes present within *semantic relations* are specifically licensed by that type or the types that it inherits from. Thus in Figure 15, the classification of *kamṛano rel* as a subtype of *cause rel* entails the presence of ACT and UND attributes. And no other attribute will be present in *kamṛano rel* unless it is entitled to inherit from some other supertype. No such guarantee is, however, ensured for the presentation in Figure 13.

Thus the value of RESTR can be represented as a specific semantic type as shown below:

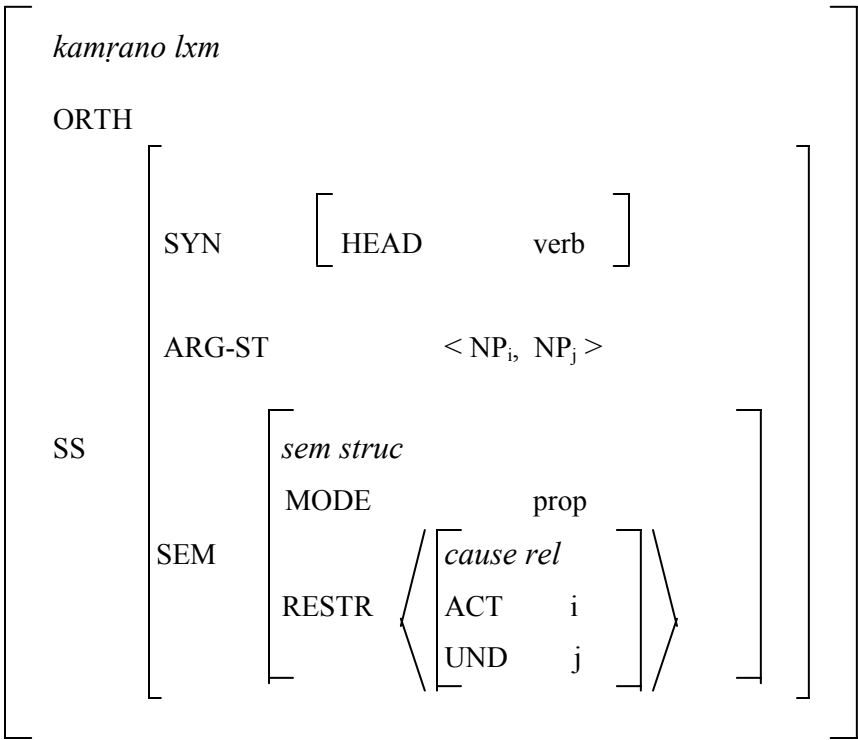


Figure: 16

I will introduce one more feature PREDS within the *sem struc*. The value of PREDS will represent the meaning component for the verb and the value of RESTR will be built from that of PREDS at the phrasal level. The need for having an additional feature for representing semantics of verbs will become evident in chapter 6.

The semantic types for nominal objects will have a feature INDEX inside the SEM, whose value corresponds to the individual referred to. Thus the semantic representation of the proper noun ‘Binu’ is the following:

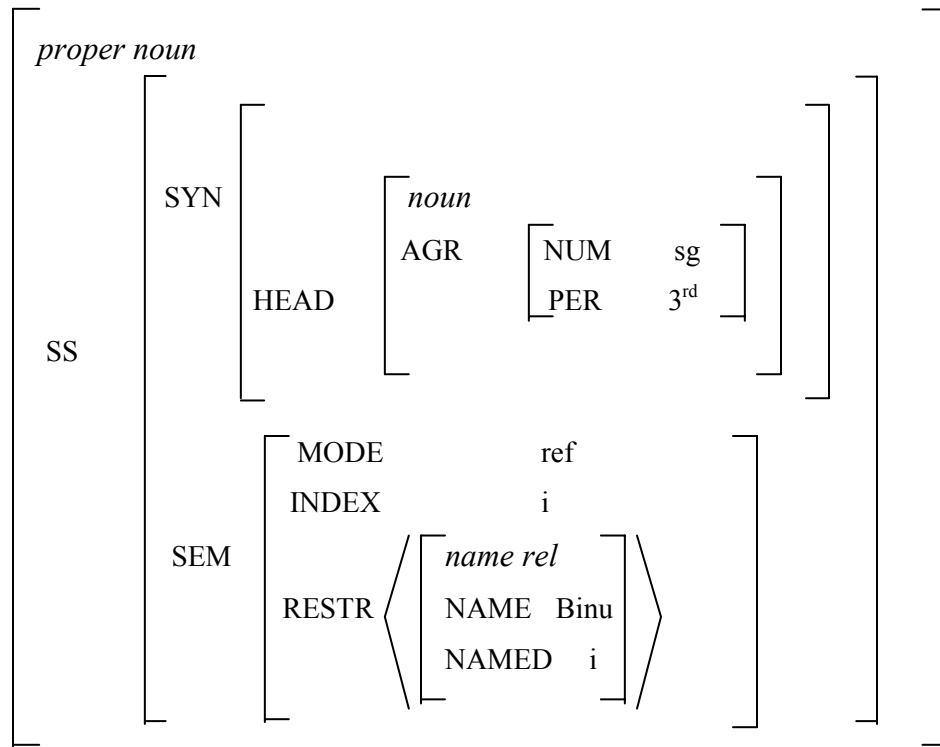


Figure: 17

As described in figure 16, the mapping between the semantic roles, ACT and UND, and the two syntactic arguments is accomplished through co-indexation, i.e., structure sharing of indices. The INDEX value of NP is structure shared with the role indices specified inside the *semantic relation* type.

Phrases are also modeled as a kind of sign, i.e., a feature structure which, like the lexical signs, specify values for the feature ORTH and SYNSEM. In addition, a phrasal type introduces attributes for its immediate daughters, each of which is a sign (either word or phrase). For instance, the Bangla sentence *binu ghumog* ‘Binu sleeps’ is modeled by the following phrasal type. The values of HD-DTR and NHD-DTR are the structural descriptions of the head daughter (here the verb *ghumog* ‘sleep’) and the subject daughter (here *binu*) respectively:

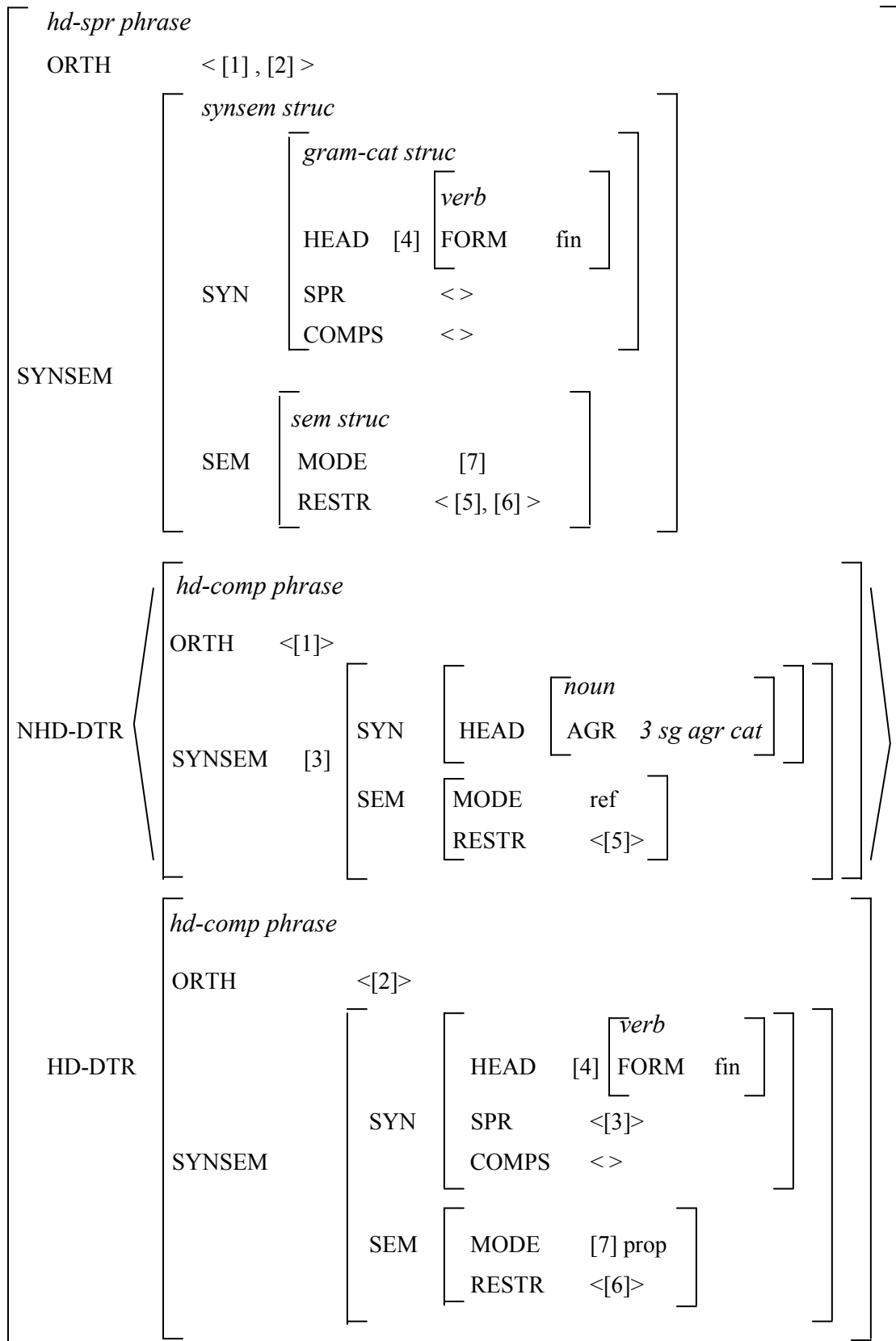


Figure: 18

The ORTH values of the daughters are added together to yield the ORTH value of the higher phrase. The syntactic and the semantic properties of the phrasal type are specified by relating the values of its SYN and SEM features to those of its daughters (the value of HD-DTR and NHD-DTR) mainly through structure sharing. The constraint on the type *hd-spr phrase* requires the following:

- a. The SYNSEM of the NHD-DTR is token-identical with the SPR value of the HD-DTR and the list for the phrasal type is empty,
- b. The head daughter picks up all of its complements before combining with the specifier.

It also incorporates the *head feature principle* and the *semantic compositionality principle* that I will be discussing in the next section.

1.4.2 Principles and constraints

Grammar rules and principles (Sag and Wasow 1999) are described here as constraints on different types of phrases. Let us first examine the universal *head feature principle* that states that the value of the HEAD of the HD-DTR is structure-shared with its mother's HEAD feature.

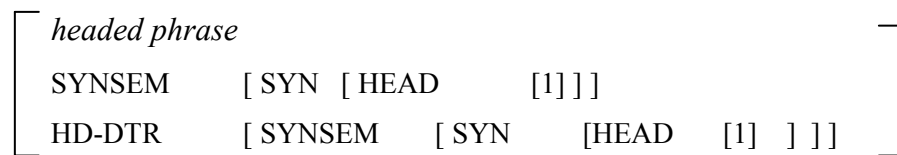


Figure: 19

The *semantic compositionality principle* constrains the value of RESTR of a phrasal type to be the summation of that of its daughters. This constraint mainly builds the semantic component of the phrases as depicted below:

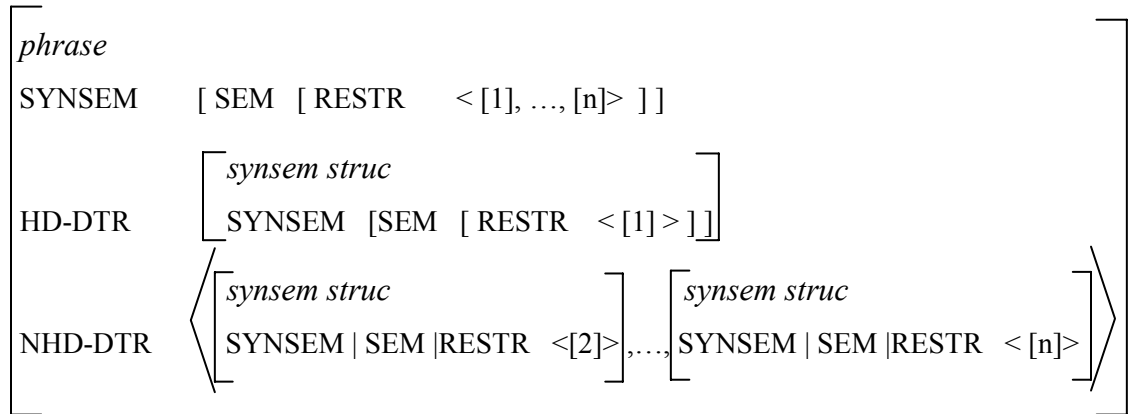


Figure: 20

The type for the feature structure depicted in figure 21 is declared to be *word*. The constraint states that the SPR and COMPS features are specified for the *word* type. The *word* is built from a *lexeme* type, the value of its STEM. The default mapping between the members of ARG-ST of the stem *lexeme* and that of SPR and COMPS of the correspondent *word* type is formalized in the following constraint declaration.

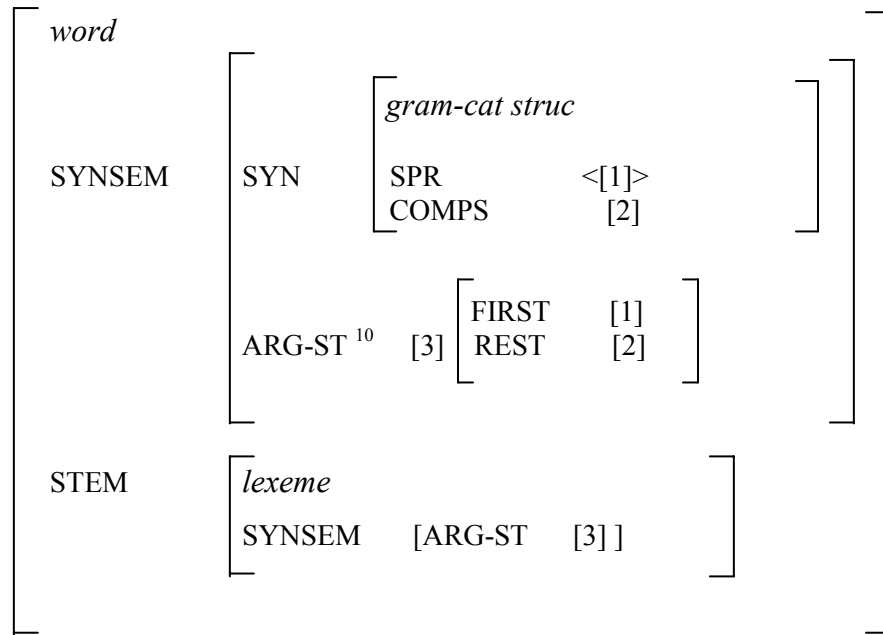


Figure: 21

¹⁰ The list value of ARG-ST is represented in terms of the features FIRST and REST

Argument satisfaction principles such as *head complement constraint* and *head specifier constraint* are imposed on the *word* and *phrasal* structure and not on *lexeme* type. The constraint on *hd-comp phrase*, as represented in figure 22, declares that the value of COMPS within the HD-DTR to be unified with values of NHD-DTR, as a result of which the COMPS feature of the mother *phrase* type will have an empty list.

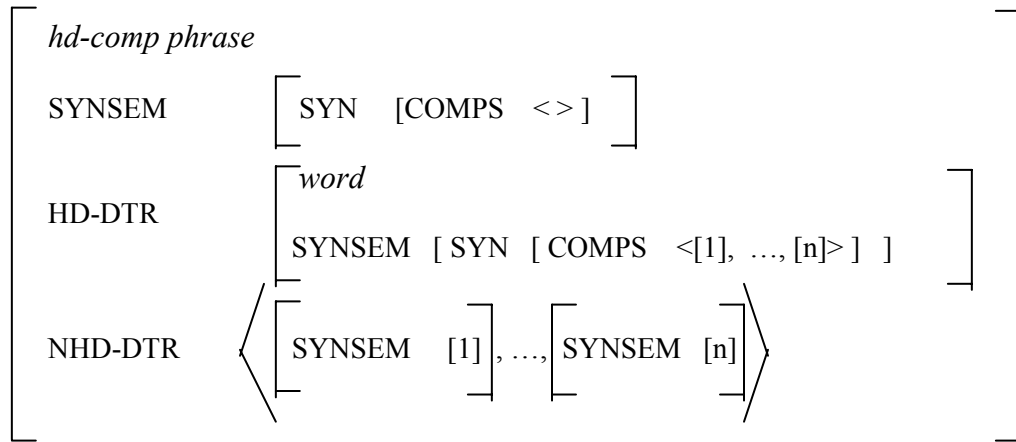


Figure: 22

Thus the constraint stated within *hd-comp phrase* actually satisfies the complement requirement of the head daughter. Similarly the constraint on *hd-spr phrase* declares that the value of the SPR within HD-DTR and the value of SYNSEM within NHD-DTR unify and the list of SPR for the *phrase* will be empty:

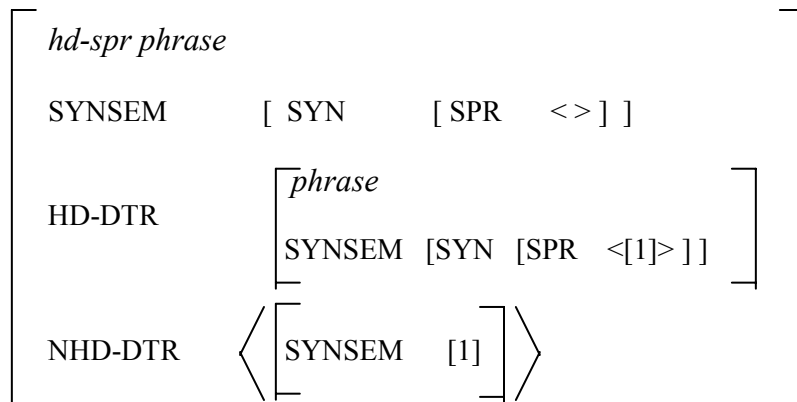


Figure: 23

I discussed here some fundamental principles and constraints that determine the well-formedness of the word and phrasal structures. In chapter 6, I will propose an additional syntactic schema, which will specify the well-formedness of CV structures for Indo-Aryan languages.

1.4.3 Elimination of redundancy

A central challenge in a lexicalist framework like HPSG is to develop strategies for eliminating the massive redundancies present in fully specified lexical entries. To capture the lexical regularities found in inflection, derivation and valence alternation, one of the simplest technical devices within the constraint-based frameworks is known as *underspecification*. An attribute is given as its value a less specific type T, to indicate that any of the more specific subtypes of T can unify with that attribute's value. For example, the lexical entry for noun can be left underspecified for the value of the feature CASE. A verb whose subcategorization (or more precisely complement) requirement declares its complement argument to be of [CASE acc] in the grammar will unify with such noun object and the unification will resolve the CASE value of the noun to 'acc'. In addition to underspecification, standard HPSG offers two primary means of expressing the sharing properties and constraints in the lexicon: (a) arrangement of types in multiple inheritance hierarchy and (b) specification of lexical rules which relate pairs of lexical entries.

1.4.3.1 Multiple inheritance hierarchy

It is fundamental to the HPSG approach that linguistic information is organized into a hierarchy of types, which includes types such as *word* and *phrase*, *part of speech* (such as noun and verb and so on) types, subcategorization based types (e.g. intransitive or transitive verbs), declension and conjugation types and also semantic types such as a type denoting a situation in which an actor is involved, a situation in which both an actor and an undergoer are involved and so on. The type system in HPSG is defined hierarchically in terms of subsumption. In such a set up, the lexical entry for a specific lexical item needs only specify the idiosyncratic information and the supertypes from which it inherits. The hierarchy of types therefore serves to minimize redundancy in the lexicon. Figure 24 illustrates a multiple

inheritance hierarchy network in which types based on part-of-speech distinction are distinguished from types based on valence differences and also from the semantic types. The hierarchy predicts the cross-cutting generalizations:

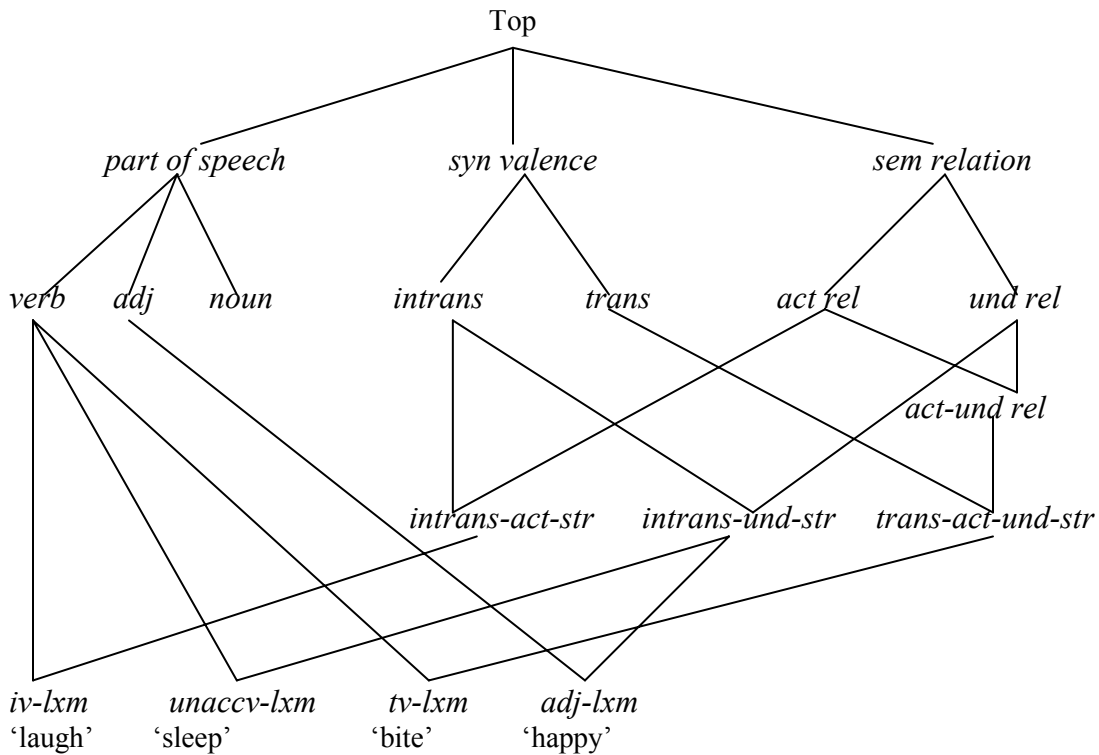


Figure: 24

1.4.3.2 Lexical rules

The lexical rule is a mechanism which can be used for further reducing redundancy by using information in one lexical entry as the basis for generating related lexical entries. Lexical rules have been proposed in HPSG to account for inflection and derivation of various kinds. The 3rd-singular verb inflectional rule, for example, adds the following stipulation to verb lexemes:

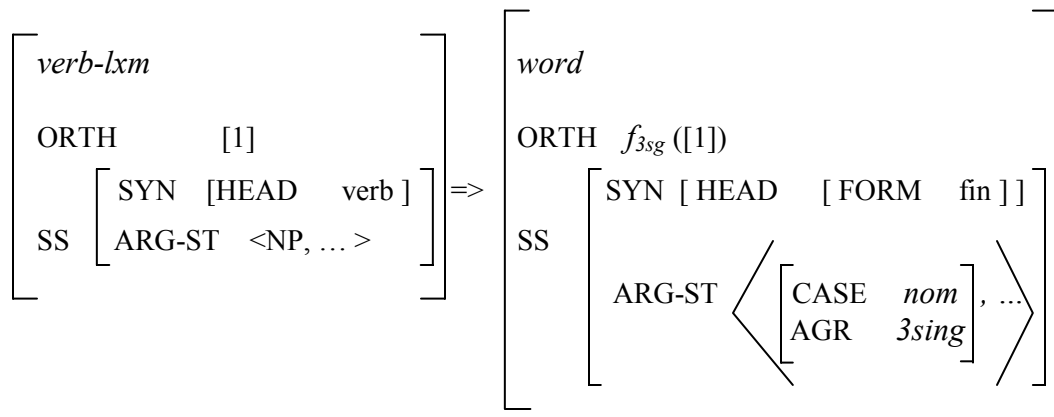


Figure: 25

The function f_{3sg} changes the orthography of the verbal lexeme and the word-form represents the 3rd singular form of the verb. The rule imposes the requirement that the value of FORM within the HEAD feature of the verbal lexeme will be fin(ite). The rule also states that the derived verb word type takes a nominative third-person singular subject. Inflectional rules generally map lexeme type to word type for inflected words. There can be another kind of lexical rule, which maps a lexeme to another lexeme. This type of rule defines derivation. One common type of English derivational rule generates *agentive noun lexeme* from *verb lexeme*: ‘give’ vs. ‘giver’, ‘drive’ vs. ‘driver’ and so on.

In this section I have discussed the salient features of HPSG grammar: the concept of sign, the arrangement of types in multiple inheritance hierarchy and the principles and constraints imposed on the phrases and lexical signs which determine the well-formedness of linguistic structure. The next section briefly introduces the features of the LKB system.

1.5 The LKB System

The LKB system (Copestake 2002) is a “high-level specialized programming language”, which provides a grammar and lexicon development environment for use with constraint-based formalisms such as HPSG. This can be used as a tool for building parsing systems for natural languages. The LKB system is implemented in Common Lisp.

LKB uses one type of data structure called *typed feature structures* and one operation, the *unification*. Both lexical descriptions and grammar rules are represented as *typed feature structures*. The type system consists of a *type hierarchy*, which indicates consistency of types, plus a set of *constraints* on types, which determine which typed feature structures are *well-formed*. As discussed earlier (see 1.3.3) a set of types is said to be consistent if and only if the members of the set share a common subtype. The significance of this is that only feature structures with mutually consistent types can be unified. Two types, which are unordered in the hierarchy, are assumed to be inconsistent unless a common subtype has been explicitly specified. I will exploit the feature *type consistency* in order to define the constraint on unification of the two Vs in a CV in the following manner: two Vs can unify to form a legitimate CV sequence if and only if their semantic types are declared consistent. Grammar rules are actually partial descriptions of phrases. The daughters of phrases are contained in the feature ARGs. This takes as its value a list where the order of list elements corresponds to the linear order of the daughters.

The type system is defined in one or more type files. It is necessary to have a valid type system before anything else can work. The system loads the type files and then carries out a series of checks and expands the type constraints.

1.6 Hierarchy of Semantic Types Representing Bangla Verbal Predicates

I will postulate a set of *semantic relations* (represented as typed feature structure) each of which corresponds to a situation¹¹ in the world outside and the value of the proto-role attributes within these semantic relations denote participants involved in the situation. These *semantic relations* will be arranged in a hierarchy with a more specific type being subtype of a more general one. Similarly, the *valence predicates* and *semantic predicates* will also be

¹¹ For example, the sentence ‘John is running’ corresponds to some running situation in which the participant John is involved as an actor. The term ‘event’ will often be used in this thesis in a general sense in place of the term ‘situation’, even though I understand that situation include non-event like circumstances such as ‘John is tall’ or ‘John has a car’. I will also use the term ‘eventuality’ (Bach 1986) which is more neutral at some occasions.

arranged in a hierarchy system. Their greatest lower bounds (which actually define the mapping between the syntactic arguments and semantic roles) in the lower part of the hierarchy will subsume the types for lexical entries. *Semantic relations* will be the value of SEM within *semantic predicates*. I will impose one universal condition on the hierarchy of the lexicon, which requires the semantic class or semantic predicate hierarchy to mirror the semantic relation hierarchy. This is an adaptation of Davis's semantic condition of correspondence between the semantic relation hierarchy and the lexical hierarchy (Davis 2001, p. 189). Informally we can illustrate the correspondence in the following manner: If there is a *act-rel* in the semantic relation hierarchy, then there will be a semantic predicate, *act-sem*, in the lexical hierarchy whose SEM feature will have the value *act-rel*. Similarly the existence of *und-rel* in the hierarchy of semantic relation will correspond to a semantic predicate, *und-sem*, in the lexical hierarchy with *und-rel* being the value of its SEM feature and so on.

I will treat the composition of well-formed CV sequence to be a matter of unification between semantically compatible V1s and V2s. The unification will be constrained by what I call the *semantic constraint on periphrastic compounding*. For a V1 and V2 to be semantically compatible in a type hierarchy, I propose that the semantic relation representing the meaning of one of the verbs (either V1 or V2) must be a subtype of that of the other, or there must be a unique common subtype, which will be the greatest lower bound of the two semantic relations representing the semantic types of V1 and V2. The grammar that I will attempt to develop in this thesis will utilize the aforementioned *semantic constraint* in association with a multiple inheritance network of lexical hierarchy in order to ensure that only well-formed CV constructions are licensed by the grammar.

1.7 Overview of the Thesis

The goal of this thesis is to present a constraint-based and semantically-grounded account of the composition of Bangla CV sequences within the framework of the lexicalist theory HPSG. The system of grammar is constructed using a monotonic multiple-inheritance hierarchy of typed feature structures. In a monotonic hierarchy, constraints on supertypes

affect all instances of subtypes without exception. The grammar is constraint based and semantically grounded in the sense that any constraint that controls the unification of a V1 and a V2 is imposed on the semantic types of verbs. This constraint will ensure the unification of semantically compatible V1s and V2s as a result of which only well-formed CV sequences will be licensed by the grammar. The analysis is finally implemented on the LKB system.

The constituents of CV sequences, on the one hand, enjoy considerable amount of freedom of movement as either of the participant verbs can independently move to various positions within a sentence. Also words of different categories such as adverbs and relative pronouns can intervene the constituent verbs. On the other hand, these constructions behave like a close-knit mono-clausal unit in various morphosyntactic operations such as passivization, agreement and so on. Even when adverbs and negative particles occur between the two constituent verbs they always take scope over the whole construction and never modify the meaning of individual participant verbs. In chapter 2, I will argue for shifting our focus from the syntactic representations of these constructions to the notion of *predicate*. A predicate is viewed as a functional-semantic unit that can either be represented by a single word form or by multiple word expressions in the language. I will mainly adopt Ackerman and Webelhuth's (1998) argument for the construct predicate. My contention will be that the multiple-word representation of CV often obscures the fact that these constructions actually represent one predicate.

The semantics of V2s and the semantic contribution that these linguistic elements make to the overall meaning of CVs will be examined in chapter 3. In this connection I will attempt to address the following issues:

1. Why does a V1 participate in *periphrastic compounding*?
2. Given that V2s are not complete predicates and that they cannot occur independently in the language what are the sources of their meaning?

Speaker's choice and context in the discourse sometimes require an emphasis to be laid on part of the event denoted by a simple verb and also add some semantic nuances to the original meaning of the verb. In my opinion *periphrastic compounds* are the lexical means to express the situation in focus. In chapter 3, I will propose that the semantics of V2s *profile* the base-events of V1s by highlighting the manner of involvement of the participants in the event and/or by imposing temporal and aspectual focus on the event. Thus the resultant CV will represent a new event type. The chapter also claims that V2s and their full-verb associates stand in polysemous relation. There exists a set of *core sense* or abstract concepts independent of linguistic expression. The meaning of V2s and their full-counterparts express same core sense.

I will discuss three models of semantic representation in chapter 4: the system of thematic roles, the decomposition model and the entailment approach. I will adopt here a partially configurational lexical semantic structure, which will represent information about the lexical entailments that hold of the participants in a given event. Following Davis (2001) I will call these semantic structures *semantic relations*. The semantic relations will be arranged in a multiple hierarchy model. The significance of organizing the semantic relations in a hierarchy network will become evident in relation to my proposal for the *semantic constraint on periphrastic compounding* in chapter 6 that will constrain the unification of a V1 and a V2.

In chapter 5 I will examine the argument structure of CV constructions. Although CVs generally inherit the arguments of their V1 constituent, this chapter records those instances where this is not the case. The chapter will propose a semantic account for argument structure modification. I will argue that V2s add semantic overtones to the meaning of V1s and consequently the resultant CVs acquire a meaning distinct from their V1 associate. The semantic structure of the derived verb might contain a set of semantic roles that requires a different set of arguments on its argument structure list than that of its V1 constituent in order to satisfy the linking constraints. Consequently the argument structure modification takes

place formally. In the latter part of the chapter, I will present a brief note on *linking mechanism*. Linking constraints will be declared as types on the lexical hierarchy.

In chapter 6, I will specify a *semantic constraint* that will control the unification of a V1 with a V2. The constraint is designed to apply on the semantic component of the verbs. In this chapter, I will also propose a syntactic schema that will build the phrase structure of CVs. Even though the V2 bears the categorial information for the CV construction, I will argue in favor of conferring the status of head on the V1 constituent. My argument will be based on the following two observations:

1. CVs are lexical variants of their V1 constituents and
2. There are languages where the categorial information is distributed between the two constituents of an analytic expression representing one predicate

Just like a simple verb, CVs project the sentences they head. I will lay emphasis on the observation that morphosyntactic representation of predicates, either a one-word form or multi-word expressions, constitutes the head of a syntactic phrase. Finally I will present an LKB-based implementation, mainly designed to test the analysis for the CV constructions proposed in this thesis.