

LANGUAGE IN INDIA
Strength for Today and Bright Hope for Tomorrow
Volume 6 : 7 July 2006

Managing Editor: M. S. Thirumalai, Ph.D.
Editors: B. Mallikarjun, Ph.D.
Sam Mohanlal, Ph.D.
B. A. Sharada, Ph.D.
A. R. Fatihi, Ph.D.
Lakhan Gusain, Ph.D.
K. Karunakaran, Ph.D.
Jennifer Marie Bayer, Ph.D.

**THE MORPHODYNAMICS OF BENGALI COMPOUNDS –
DECOMPOSING THEM FOR LEXICAL PROCESSING**

Niladri Sekhar Dash, Ph.D.

THE MORPHODYNAMICS OF BENGALI COMPOUNDS: DECOMPOSING THEM FOR LEXICAL PROCESSING

Niladri Sekhar Dash, Ph.D.

ABSTRACT

Compounding is a highly fertile process. It is quite often used in various innovative ways for generating new words in most of the languages. At the time of compounding the participating members often undergo a process of morphosyntactic change that forces them to lose much of their lexicosemantic information.

In this paper we make an attempt to capture lexicosemantic properties, which are lost in this process, and try to identify the factors that play active roles behind such metamorphosis of compounds. Our investigation is based on Bengali compounds as the central area of study with occasional references to the English compounds for understanding the phenomenon in a systematic way. The present study has direct applicational relevance in the area of applied linguistics, mainstream linguistics and language technology.

1. Introduction

Compounding is one of the most fertile processes of new word formation in Bengali in which new words are generated by combining two (rarely more) words (or stems). Prototypically, it is a process of lexical concatenation that generates new words to increase the load of the lexical stock (i.e. lexicon) of a language as well as to provide new lexical devices for expressing novel ideals and objects entering into a language form time to time for various linguistic and extralinguistic reasons. Compounding is a comparatively free process that involves a large set of open-class lexical items enriched with specific meanings. The resultant compound generally contains at least two bases, which are either free words, or at any rate, root morphemes.^[1]

The presence of small or large number of compound words in a natural language is a widely approved phenomenon. But the modus operandi of their formation and the nature of their lexicosyntactic and lexicosemantic functions differ from language to language. These underlying differences necessitate an intimate analysis of forms and functions of compound words in Bengali before they are fit to be placed within a generalised framework of compound classification and processing.^[2]

The process of compound formation and their functional types differ from language to language because compounding involves combination of more than one stem/affix either in free or in bound forms, which is often integrated lexically to function as single lexical units. The basic strength of compound words lies in their indivisibility of structure, since no external element is usually allowed to occupy a place between the two or more members of a compound. Moreover, the specific sequential order of the members is so rigidly fixed that they are hardly allowed to occur in reverse order.^[3] Besides, from semantic perspective, the original meanings of compound words cannot be generally, but not exclusively, derived from sum total of meanings of the formative members.

Within empirical frame of corpus-based analysis of compounds for lexical processing our main concern is to find out whether the scheme of compound word classification

proposed by earlier scholars is equally effective for analysing the compound words found in Bengali corpus. While dealing with the lexical stock of Bengali compounds we note some peculiarities, which are uniquely interrelated to the formal and semantic characteristics of the Bengali compounds.

In the present study we intend to explore how linguistic as well as extralinguistic information embedded within surface forms of compounds are interfaced for normal cognition of the phenomenon. We find that lexicosemantic analysis of compound words supply valuable information which we may use for developing study materials for language teaching, compiling dictionaries for compounds, designing sophisticated systems for compound decomposition for machine learning, and developing electronic lexical databases for machine translation, lexical processing and information retrieval. Moreover, in this study we look into the process of compound generation in Bengali to contribute towards the understanding of the process of linguistic generativity of normal Bengali speakers. However, before we explore the phenomenon in some details, it is sensible to have a look at the past history of the studies related to both English and Bengali compounds.

2. The Past History: A General View

Compounding is not a process universally used by all languages. Scholars generally argue that compound, which is formed with combination of at least two lexical items, is a privilege of a few languages. Some languages like French and Spanish have little or no compounds while languages like German, Dutch, Sanskrit, Hindi, Bengali, Telugu, etc. have extensive use of compounds. Bloomfield (1933), Jespersen (1929), Murchand (1959), Hockett (1963) and others have presented descriptive studies on the phenomenon in English without paying much attention to the generative aspect of compounds. Adams (1973), on the other hand, has tried to construct a taxonomic list of the types of semantic relation underlying compounding members. He has also used terms such as ‘appositional’, ‘associative’, ‘instrumental’, ‘locative’, ‘resemblance’, ‘composition/form/content’, ‘adjectives-noun’, ‘names’, etc. to identify the classes of non-verbal compounds.

Among the members of generative school, Lees (1960) has argued that compounds have their source in (a plethora of) underlying sentence structures and are derived by the process of transformation. Similarly, Roeper and Siegel (1978) have argued that a subset of compounds that contains deverbal second elements (such as *school teacher*, *teacher training*, *mind-boggling*, etc.) is easily derivable by simple process of lexical transformation. On the other hand, Selkirk (1983) has probed into the structure and headedness of nominal compounds, verbal compounds, and category type of English compounds. She has presented an explicit analysis of English compound constituents with a claim that transformations have no role to play in English compound formation. Therefore, it is not necessary to introduce rules into grammar of compound formation. In order to understand the system of compounding Leiber (1983) has provided an elegant account to analyse the compounds by introducing the terms like:

- (a) **Argument structure of words** – an element that bears a thematic relation such as agent, theme, goal, source, instrument, etc. to the head of a compound.
- (b) **Feature percolation** – the features of morphemes that make up a word are interpreted as features of word as a whole.
- (c) **Argument-linking principle** – in compound governing it is a vital element that claims that a compound must be able to link all its obligatory internal arguments.

Within the frame of derivational morphology, Bybee (1985), with some examples obtained from English and Tiwi (an Australian language) has shown that compounding and incorporation are related to derivational morphology and there is a diachronic relation between the two. Anderson (1988) on the other hand, has raised theoretical objections in acceptance of compounds as single words. Jensen (1990) also has related conventional division of compounds from generative perspective with a provision for classifying compounding constituents into synthetic compounds, gerundial compounds, participial compounds, and exocentric compounds. He has also applied the argument structure and feature percolation proposed by Leiber (1983) to analyse the English compounds. Spencer (1991) has examined the compounding processes with emphasis on English synthetic compounds as well as on Turkish compounds. On the other hand Katamba (1993) has discussed the phonological factors and syntactic aspects related to English compounds.

The history of analysing Bengali compounds is a long one. Perhaps, William Carey is first scholar who has made a serious attempt to study the Bengali compounds from analytical perspective (1805).

In recent times, Suniti Kumar Chatterji (1926) has presented detailed classification of compounds along with information about their formative and semantic properties. Bamandev Chakravarty (1974) has shown the differences between *Tatsama* and *non-Tatsama* compounds while Sukumar Sen (1993) has traced the differences between Bengali and Sanskrit compounds.

Pabitra Sarkar (1998) has discussed the formation patterns of compounds in the context of word formation in Bengali. Satya Ranjan Banerji (1997) has analysed Bengali compounds with close references to Sanskrit and Greek compounds. Sumita Bhattacharya (1983, 1997) has presented descriptive study of Bengali compounds while Sukumar Sen and Subhadra Kumar Sen (1994) have classified Bengali compounds in two groups according to the pattern of change of meaning of formative constituents.

Jyoti Bhusan Chaki (1996) has also tried to find answers to the questions related to actual identity of the compounds as well as the significance of respective head names. Punya Sloka Ray (1997) has tried to draw a line of distinction between compounds and phrases on the basis of some negative potentialities.

3. Purpose Behind Compound Formation

We are highly curious to know why people have a tendency to generate compound words. According to some scholars, behind the phenomenon of compound formation in a language there are at least four purposes, which may be summarised in the following way:

- (a) The basic idea behind formation of compounds lies in an intention to abridge the expanded words.
- (b) Since the advantage of compounding mostly lies in making a lengthy sentence into a single complete and composite form, many sentences may be reduced into single word by way of applying this method.

- (c) Process of compounding brings several words into a single system of accentuation. Thus, the uniformity in accentuation becomes responsible for bringing ease in pronunciation of compounds, which is hardly available in sentence.
- (d) Process of compounding also brings in a series of syntactically related words into a single frame of case-termination.

The arguments furnished above clearly show that there are several cognitive as well as functional advantages in formation of compounds in a language.

This observation, however, contradicts the statement made by Mukhopadhyay (1999) who has argued that use of compound words in Bengali is gradually diminishing in number because the first member of a compound word shows a tendency to be separated with an inflection or case marker.

For instance, some compounds like *pAkhiniR* “nest”, *cokhjal* “tears”, *khabarkAgaj* “newspaper”, etc. are found to be rewritten as *pAkhir nIR*, *cokher jal* and *khabarar kAgaj*, respectively. In these cases, the first members of the compounds are tagged with case markers while the second members remain unchanged in their own forms. Whether this observation is true will be verified with compound words obtained from the Bengali text corpus (Dash 2005, Chapter 2) in the next section of the present discussion.

4. The Nature of Bengali Compounds

In the present discussion we have tried to analyse the Bengali compounds to find out their patterns of formation as well as their functional roles in the language. The derived structure as well as meaning of compounds is given equal importance because without reference to these issues analysis of compounds is bound to be skewed and incomplete.

Moreover, due to genealogical, typological and functional affinities with Sanskrit compounds, Bengali compounds sometimes behave like phrasal units in the text where they denote a sense, which is more often syntactic than lexical. However, before we probe in details, it is needed to explore the general features of compound words. These features might possess some universal traits underlying their surface composition:

- (a) Compound words are often lexicalised and stored in dictionary as entry words with distinct lexical entity. Here, they resemble with unique lexical expressions because the resulting units are words, meanings of which are not often predictable from the summation of meaning of the parts.
- (b) Compounds are often subjected to semantic drift normally associated with stored words of a language. This implies that meaning of compounds may become non-compositional or even totally idiosyncratic in respect to meaning of constituents.
- (c) Morphologically complex words created by compounding differ from those created by derivation and inflection in that such words cannot be analysed as consisting of a stem or root plus affixes. Rather, they contain more than one stem or root. The elements that combine to form compounds are not lexical plus grammatical, rather two or more lexical elements.

- (d) Process of compounding is productive in the sense that new compound words are freely created. The results of compounding are lexicalised and they gradually tend to lose their semantic and phonological transparency.

The use of compound words in Bengali texts is a regular phenomenon. The process of compounding is mostly used for generating new words. Compounds, in most cases, are considered as single lexical units. The striking aspect is that we can generate a derived word from a compound word by way of adding derivational affixes. Moreover, like simple words, addition of derivational affix to compounds may cause morphophonemic change of final forms.

Form a simple statistical count run on the Bengali corpus of written texts, we have found that nearly twenty percent (20%) of the total vocabulary used in Bengali belongs to compounds. It includes not only nominal and adjectival compounds but also verbal compounds derived by way of combining a participial form of a verb with another verb either in finite or non-finite form (e.g., *dharApaRA* “caught”, *balAkathA* “spoken word”, *jAnAshonA* “known”, *pheleAsA* “left behind”, etc.).

From structural point of view, most of the nominal compounds are formed by adding two or more noun words (e.g., *lokjan* “people”, *rAjbARi* “palace”, *gaNashakti* “public force”, etc.), or by adding a prefix to a noun word (e.g., *prajanma* “progeny”, *svajan* “kin”, *sudin* “good day”, etc.). The adjectival compounds are formed by adding two adjectives (e.g. *himshital* “ice-cold”, *shvetashubhra* “white”, *krSNakAlo* “black”, etc.), or by adding specific adjectival suffix with noun words (e.g., *bhAtrpratik* “brother-like”, *yuddhabAj* “war-monger”, *kharAprabaN* “draught-prone”, *parbatpramAN* “mountain-like”, *bacanbAgIsh* “talkative”, *patangabhuk* “insect-eating”, etc.).

Form the count of the corpus we have found that the number of compound words in Bengali is much higher in *SAdhu* or chaste version of texts than in *Calit* or colloquial version. It may be due to the fact that a *SAdhu* version of text is more archaic in nature and therefore, it is akin to preserve large number of *Tatsama* words, which are mostly compound in form and meaning. ^[4]

In the corpus texts we have also found that there is a strong tendency to form new compound words in Bengali for technical and scientific terms or phrases borrowed from foreign languages, particularly English. For example, several English phrases like *difficult to be performed* is converted into a compound like *dUrabhisambhab*, *expelled from family* is compounded as *kUlacyuta*, *ministry of education* is compounded as *shikSAmantrak*, *son of the soil* is compounded as *bhUmiputra*, *person who lacks balance* is formed as *tAlkAnA*, *locked in house* is compounded as *grhabandI*, *run by air* is coined as *bAyucAlita* in Bengali.

There are many such examples scattered in the corpus the analysis of which easily establishes the fact that the number of compound words in Bengali is increasing over time. Even regular Bengali phrasal forms are changed in compound words by simple lexical contraction. For instance, *hAte gaRA ruTi* is changed into *hAtRuTi* “capati”, *leker dhAre bARi* becomes *lekbARi* “lake house”, *anks shekhAnor mASTAr* changes into *ankamASTAr* “maths teacher”, *rAte parAr poshaAk* is converted into *rAtposhAk* “night dress”, etc. These observations, therefore, strongly defies the statement made by Mukhopadhyay (1999). ^[5]

4.1 Types of Bengali Compounds

In general, four types of compound words are found to be used in Bengali text corpus. These are:

- (a) Tatsama compounds, which are directly inherited from Sanskrit.
- (b) Indigenously formed compounds by way of derivation and sandhi.
- (c) Compound words borrowed from foreign languages such as Arabic, Persian, English, and others.
- (d) Analogically formed compound words (i.e. loan translations) following the word formation rules allowed in Bengali and using various native and non-native lexical elements.

There is considerable amount of inconsistency in case of orthographic representation of these compounds. In fact, orthographic convention is not at all a good guide for proper identification of compound words in Bengali. Because while some very well established compound words are often written as single word units, with and without a hyphen mark between the members (e.g. *rAjbARi* “palace”, *rAjpath* “main road”, *rAja-bAdshA* “kings and monarchs”, *rAjputra* “prince”, *jIban-maraN* “life and death”, *ghar-duyAr* “house and courtyards”, etc.), some other compound words are found to be formed with two isolated words located with a space in between (e.g., *rAm chAgal* “ram”, *rAj kArya* “royal work”, *rAtcara* “nocturnal”, *bAjkhAi* “eagle-like”, etc.).

Thus, from orthographic point of view some compound words appear as single word units but some other compound words appear as two separated words. Therefore, orthographic convention offers little help to distinguish compound words in Bengali – neither structurally nor orthographically.

Euphonic combination plays a vital role in formation of Bengali compounds. Here two different words are merged together to generate a compound by morphophonemic metamorphosis where a character (i.e. grapheme, allograph, etc.) belonging to a word is lost to generate a new character (e.g. *vidyAlaya* “school”, *nIlotpal* “blue lotus”, *dampati* “couple”, *shabdArtha* “word meaning”, *nirukta* “unsspoken”, *mahodaya* “great man”, etc.). In the examples mentioned here, the word-final character of W_1 and the word-initial character of W_2 are combined together to generate a different character represented in the derived surface form.

Since the process of compounding is highly recursive in nature in Bengali, generation of more complex structures of compound is possible by way of using the recursive elements.

Generally, Bengali compounds are made up with two formative words, each one of which belongs to a particular lexical category. But, at certain times, three or more words are combined together to generate either a single word unit (e.g., *janamanasancArI* “creeping into public mind”, *abAnmanasagocara* “beyond mind and vision”, etc.) or multiword unit (e.g., *hAt-pA-nAk-kAn* “hand-leg-nose-ear”, *hAjAr-hAt-kAlI* “goddess Kali with a garland of thousand cut-off hands”, etc.) or a syntactically loose compound where each words is linked with its neighbouring word in syntacticosemantic relation (e.g., *kolkAtA nAgarik unnayan samiti sammelan* “Kolkata citizen welfare society meeting”, etc.).

Prefix (particularly negative prefix) has an important role in compound formation in Bengali. It is a well-known fact that addition of prefix can affect the semantic load of a compound. In case of Bengali compounds, addition of prefix does not only affect the meaning of final compounds but also changes their lexical categories. For instance, a noun by way of addition of a prefix to it changes into a compound adjective. Some instances are presented in the following table (Table 1):

Underlying Form	Surface form	Gloss
[A-] _{Prefix} + [maraN] _N	[AmaraN] _{Adj}	till death
[nir-] _{Prefix} + [jal] _N	[nirjal] _{Adj}	waterless
[phi-] _{Prefix} + [bachar] _N	[phibachar] _{Adv}	every year
[phul-] _{Prefix} + [hAtA] _N	[phulhAtA] _{Adj}	full sleeve
[be-] _{Prefix} + [imAn] _N	[beimAn] _{Adj}	shameless
[sa-] _{Prefix} + [bAndhab] _N	[sabAndhab] _{Adj}	with friend
[har-] _{Prefix} + [roj] _N	[harroj] _{Adj}	every day
[hA-] _{Prefix} + [ghar] _N	[hAghar] _{Adj}	homeless
[hAph] _{Prefix} + [hAtA] _N	[hAphhAtA] _{Adj}	half sleeve

Table 1: Change of lexical category due to addition of prefix

Since there is no well-defined system for distinguishing compound words from phrases in Bengali, it is difficult to identify the mark of distinction between the two. However, we are able to identify some clues that may contribute towards this goal:

- (a) Morphological integrity noted in compounds distinguishes them from phrases.
- (b) Phonological processes like Sandhi are freely applied to compound word formation but are never used to form phrases.
- (c) Structural expansion or modification never occurs in compounds to interrupt the sequence of the formative members.
- (d) Juncture never occurs to interrupt the sequence of the formative constituents of compounds. This signifies that formative constituents of compound words cannot be split up to insert other words or items.
- (e) Accentuation is applied either on first or last syllable of a compound word. This rule does not apply for phrases.

Therefore, for proper understanding of Bengali compounds as well as for identifying their separate entity from phrases we can probably consider the following conditions:

- (a) A compound word must have at least two formative constituents, which may be or may not be used as individual lexical items in the language. In other words, although bases are combined together to form compounds, they are essentially autonomous words. Therefore, possibility of occurring as independent words is not a pre-requisite condition for bases to be qualified for compounds. Numerically, only a few compounds contain bound bases.
- (b) Formative constituents may stay separated with a space or hyphen in between, but taken together they will definitely provide some extra information or meaning, which is not possible to derive from the totality of meanings of each participating constituents.
- (c) Compounds may resemble to single word units because they are often open for lexicalisation. Moreover, they are often put to semantic gradience (Leech, Francis, and Xu 1994), which is normally associated with stored words. That

means the meaning of formative members becomes non-compositional or totally idiosyncratic for compounds. For instance, compounds like *shulapANi*, *vInApANi* no longer indicate a person with a spear/lyre in his/her hand. Rather they denote god or goddess, respectively, associated with these compounds from time immemorial by some extralinguistic events or beliefs.

- (d) There are often some lexical restrictions on which the formation of compounds is permitted. It results into a kind of ‘paradigmatic gap’, which strongly resembles with derivational and inflectional affixation. For instance, we can write *shabdAha* “burning dead body” and *shabsAdhana* “meditation over a dead body” but cannot say **maRAAdAha* and **maRAasAdhana* although respective meanings of these two hypothetically formed compounds are identical with earlier two examples. In a similar fashion, we can form *brSTipAt* “rain fall”, *bAripAt* “rain fall”, *tuSArpAt* “snow fall”, *shishirpAt* “dew fall”, *ashrupAt* “tear-shed” or even *raktapAt* “blood-shed”, but cannot say **jalpAt* or **nirpAt* or **pAnipAt* or **kAnnApAt*, etc. although in all the cases, the meaning is nearly similar denoting the fall or shedding of rain, water, tear, etc.
- (e) Morphological integrity is an important aspect of compounds. The constituents of compounds cannot be split up or replaced by other words or phrases. It is not an uncommon feature that the formative elements of compounds, because of their frequent co-occurrence in compound formation, become so lexicalised that they lose their status as independent words and becomes an enclitic or an affix.
- (f) Several highly generative morphophonemic processes operate in formation of new compounds. Generally, there are sets of euphonic rules, which are applied for this purpose. For instance, compounds like *mahauSadhi* “great medicine”, *yAbajjIban* “whole life”, *Adyanta* “from beginning to end”, *rAjarSi* “hermit-king”, *yatheccha* “as one likes”, etc. are formed by applying sandhi rules.

4.2 Formal Aspects of Bengali Compounds

The process of compound formation in English, as supporting evidence, is referred to for understanding the complex phenomenon of Bengali compounds. In English, from structural point of view, compounds with a combination of N + N (e.g. *school bus*), Adj + N (e.g. *black board*), N + Fv (e.g. *babysit*) occurs quite frequently. Further lexical analysis shows that (Selkirk 1983: 14):

- (a) Compound noun consists of a noun, adjective, preposition or verb on the left and a noun on the right,
- (b) Compound adjective consists of a noun, adjective or preposition followed by an adjective, and
- (c) Compound verb consists of a preposition followed by a verb.

Besides such lexical categorisation of formative constituents, English compounds are also analysed structurally from two different angles. From structural point of view, compounds are divided into two broad perspectives: morphological perspective and syntactical perspective. From morphological perspective again, compounds are divided into two types:

- (a) Synthetic compounds: In case of these compounds the second member is derived by adding a verbal suffix or a past participle form of a verb while the first member is usually a noun. E.g. *breath taking*, *watch maker*, *time killed*, etc., and

- (b) Primary compounds. This group encompasses all other types of compound available in the language.

Compounds in English

From constructional or syntactical point of view English compounds are identified as:

- (a) Syntactic compounds: Participating members force the final compound to act like a phrase to bring in an extra dimension in the meaning of the final form (e.g. *black bird, white cap*, etc.),
- (b) Asyntactic compounds: Participating members hardly develop any grammatical or semantic relation among themselves to denote novel sense in the final form. E.g. *door-knob* or *green field* etc.), and
- (c) Semi-syntactic compounds: Participating members develop some kind of syntactic relations but not like phrases. E.g. *house keep* \Leftrightarrow *keep house* etc.

If we analyse the syntactical perspective behind the formation of compounds, we may easily find that these compounds have at least two sets of characteristic property. The first set makes compounds resemble to normal syntactic processes in that there are is typically recursive. The elements of compounds have syntactic relations to each other, which resemble to the relations holding between the constituents of sentences. For example, let us consider the following compounds.

- (a) mass literacy...
- (b) mass literacy campaign...
- (c) mass literacy campaign programme...
- (d) mass literacy campaign programme committee...

The second set brings compounding closer to simple word formation. It points out that compounds have a constituent structure, which in general, depends on the way the final forms are built up. For example, compounds like *mass literacy campaign* can easily be analysed as:

- (a) [mass [literacy campaign]]
- (b) [[mass literacy] campaign]

Complexity of Compounds in Bengali

The scheme of analysis used for English compounds can also be applied for Bengali compounds with necessary modifications.

However, the nature of Bengali compounds is far more complicated because of the presence of a few unique factors in their formation. Formally, the components, which enter in compound formation, combine phonetically and/or orthographically.

The combination is further characterised by the use of stress, intonation, spelling, or hyphenation, etc. To account for these factors, the following aspects of Bengali compounds have to be taken into consideration while we try to explore their forms.

- (a) Some formative constituents do not undergo morphophonemic change of any kind while entering into compound formation. E.g. *banaphul* “wild flower”, *shaharbaSI* “city dweller”, *nIlkamal* “blue lotus”, *kAlratri* “ominous night”, *jAtashatru* “born enemy”, *nAcghar* “dance room”, *mAjhdariyA* “mid-stream”, *raktacakSu* “red eyes”, *shvetabastra* “white cloth”, etc.
- (b) Some formative members undergo morphophonemic change while they participate in formation of compounds. E.g. *vidyAlaya* “school”, *nilotpal* “blue lotus”, *mahodaya* “great man”, *debesh* “lord of gods”, *naresh* “king of men”, *narendra* “king of men”, *jalocchvAsh* “deluge of water”, etc.

For some compounds the morphophonemic change becomes so significant that the formative components lose their entire formal identities, e.g. *dampati* “couple”, etc. Taking these aspects into consideration we summarise the structural processes of compound formation in Bengali in the following way (Table 2) with examples collected from a Bengali text corpus.

Process	Final form	Glossary
$W_{1+\emptyset} + W_{2+\emptyset}$	manpaban	mental-air
$W_{1+Case} + W_{2+Case}$	patheghATe	in road and fields
$W_{1+Case} + W_{2+\emptyset}$	hAtekhaRi	holding chalk in hand
$W_{1+\emptyset} + W_{2+Case}$	lAlpeRe	with red border

Table 2: Structural formation of Bengali compounds

4.3 Semantic Aspects of Bengali Compounds

From functional or semantic point of view the English compounds are divided into two types:

- (a) Endocentric compounds: One of the formative members stands for the whole form or works as a head for the entire form. E.g. *mailman*, *blackbird*, etc., and
- (c) Exocentric compounds: None of the formative members is entitled to substitute the whole compound. E.g. *pickpocket*, *lazybones*, *sweetheart*, etc.

It is generally argued that most English compounds are of endocentric type because the right-hand constituent carries the basic meaning of the compound as a whole and the modifier element of the compound has a function of attributing a property to the head, much like the function of an attributive adjective. In case of exocentric type neither of the constituents is called the head of the final forms. The compounds like *sweetheart* do not imply a kind of *heart*, which is *sweet* in taste but refers to *fiancé* (Jensen 1990: 99). In these compounds we isolate a predicate-type element (e.g. *pick*, *lazy*, *sweet* etc.) and an argument-type element (e.g. *pocket*, *bones*, *heart* etc.).

Possessive compounds refer usually to a compound noun with no head (Spencer 1991: 311). It is a type of compound that denotes a thing, which is not a part of the compound. For example, *low-life* and *block-head* are possessive compounds, since a *low-life* is not a kind of life, and a *block-head* is not a kind of head. The concept will become clear if these forms are compared with more common and headed compounds like *fly-ball* (a kind of ball), *alley cat* (a kind of cat), etc. Infact, possessive compounds are often translated by “possessing...” or “-ed” (e.g. *possessing much butter* = *much buttered*, etc.).

In case of determinative compounds, the first member is in a case relationship with the following member. For example, a compound like *doghouse* is tagged as dative or determinative compound because it means “a house for dog”^[6].

Finally, in case of descriptive compounds, relation of the first member with the last one is appositional, attributive or adverbial. E.g. *bulldog* “a dog in the shape of a bull”, *notebook* “a book for writing notes”, etc.

Analysis of Bengali compound should be based on Sanskrit compounds because Bengali compounds are virtually modelled after Sanskrit compounds. According to the *MahAbhASya* (2/1/6) of Patanjali, there are six types of compound words in Sanskrit:

- (a) dvandva samAsa (i.e. copulative compounds),
- (b) tatpuruSa samAsa (i.e. determinative compounds),
- (c) dvigu samAsa (i.e. numeral compounds),
- (d) karmadhAray samAsa (i.e. descriptive compounds),
- (e) abyayIbhAba samAsa (i.e. adverbial compounds), and
- (f) bahubrihi samAsa (i.e. reciprocal compounds).

From semantic perspective it has been argued that in case of copulative compounds both the formative members carry equal semantic importance^[7]. In case of adverbial and numeral compounds the first member is important and it carries the entire semantic load of the final form. In case of determinative and descriptive compounds, the last constituent carries all the semantic relevance of the final form. And finally, in case of reciprocal compounds something other than the participating members are important semantically.

Bengali has a long history of analysing compounds from semantic point of view. Following the age-old tradition of Sanskrit compound analysis, Bengali compounds are analysed from semantic point of view rather than their structural point. Within a broad semantic frame Bengali compounds are divided into three main types: collective, determinative and descriptive. The determinative includes the compounds with one element governing another, appositional determinative and numeral determinative. Collective compound includes copulative compounds and similar other words, while the descriptive includes reciprocal compounds (Chatterji 1995: 176). From the point of meaning these are also divided into eight types: determinative descriptive, reciprocal, correlative, numerical, copulative, adverbial, verbal, and phrasal (syntactical) compounds (Sen 1993: 261-263).

From the point of view of affect on meaning of the formative constituents the Bengali compounds are classified into two types (Sen 1993: 264):

- (a) Semantically unaffected (svArthak) compounds where meaning of the participating members remain polarised and unchanged, and
- (b) Semantically affected (anyarthak) compounds where no participating member retains its original meaning unchanged.

Taking all the observations of earlier scholars into account we consider the following semantic aspects of Bengali compounds at the time of processing them at lexical level, since the feature of lexical generativity affects heavily the internal semantic

content as well as the external orthographic texture of the constitutive members while deriving the final forms.

- (a) In most of the cases none of the formative constituents loses its own meaning. As a result of this the whole meaning of the finally derived compound becomes exactly equal in sense to the total meaning of the formative elements. E.g. *khAbarkAgaj* “newspaper”, *mAchbhAjA* “fish fry”, *lAlgolAp* “red rose”, *sAdAkAgaj* “white paper”, *loksangit* “folk song”, *gaNashatru* “public enemy”, etc.
- (b) The meaning of some compounds is not derivable from the combined meaning of their constituents. That means the meaning of the whole is not the sum total of the meanings of the constituents. Usually, one or both of the constituents lose their meaning either partially or totally. The compounds collectively refer to another object or item having a different sense or meaning. The whole group of *BahuvrIhi* compounds falls under this class. E.g. *dashaAnan* “Ravana”, *balipuSTa* “crow”, *basantadut* “cuckoo”, *shulapANi* “Lord Shiva”, *bInApANi* “Goddess Sarasvati”, *choTalok* “ignominious person”, *nishikuTumba* “thief”, etc.

The compounds, the meanings of which are derivable from meanings of the formative members, do not posit much complex problem for automatic lexical processing. But those compounds in which the meanings of the formative members are not possible to predict from the meanings of the derived forms are real challenge for automatic lexical processing.

The first set of compounds may be treated in the same way as single lexical items are processed.

The second set of compounds, however, ask for sophisticated treatment supported with information derived from context of word use, native language efficiency, cognition of discourse, and world-knowledge (Dash 2005).

That means the lexical processor has to depend on some extralinguistic knowledgebase for automatic decipherment of meanings of the compounds. However, severe complexities will arise in the formulation of algorithms of processing, since these compounds often act as single lexical units in language. Also, these forms are treated as single words in derivation, inflection as well as in dictionary entry.

5. Decomposing Bengali Compounds

Since Bengali compounds are mostly semantically controlled, the method of positional analysis proposed in RHHRs (Right Hand Head Rules), LHHRs (Left Hand Head Rules) for English compounds, is of little help. However, we can use certain principles for decomposing Bengali compounds because the grammar of a language must include some systems of rules explicitly demarcating the range of possible compound types of the language. In the following subsections we try to decompose Bengali compounds to investigate the morphodynamic patterns embedded within their derived outputs.

5.1. Root compounds

In Bengali, root compounds are simple concatenated form of words where none of the formative constituents is inflected – only the bare stems are combined. At the time of structural analysis of root compounds, it is found that only the words belonging to the major lexical classes (i.e. noun, adjective, and verb) are productively involved in root compound formation. However, not all possible configurations of these lexical classes are attested. Generally, nouns and adjectives are most productive classes of possible heads that compound with nouns, adjectives, and verbs. Some examples are:

- (a) **Noun + Noun > Noun:** pitA + mAAtA > pitAmAtA “parents”, path + ghAT > pathghAT “roads”, din + kAl > dinkAl “days”, hAt + pA > hAtpA “hand-feet”, rAj + besh > rAjbesh “king-cloak”, gangA + jal > gangAjal “Ganges water”, shishir + bindu > shishirbindu “dew drop”, rAj + prAsAd > rAjprAsAd “palace”, etc.
- (b) **Noun + Verb > Noun:** ban + bhojan > banbhojan “picnic”, pashu + pAlAn > pashupAlAn “animal-rearing”, kanyA + dAn > kanyAdAn “daughter-giving”, rakta + pAn > raktapAn “blood sucking”, ashru + mocan > ashrumocan “tears shading”, bastra + dhAraN > bastradhAraN “dress wearing”, etc.
- (c) **Noun + Adjective > Adjective:** taila + sikta > tailasikta “oil-soaked”, rakta + snAta > raktasnAta “blood-bathed”, bastra + sajjita > bastrasajjita “adorned with clothes”, banyA + plAbita > banyAplAbita “flooded”, kharA + pIRita > kharApIRita “draught stricken”, etc.
- (d) **Adjective + Adjective > Adjective:** sAdA + kAla > sAdAkAla “white-black”, choTa + baRa > choTabaRa “small-big”, bhAla + manda > bhAlamanda “good-bad”, moTA + saru > moTAsaru “thick-thin”, sojA + bnAkA > sojAbnAkA “straight-twist”, hAlkA + bhAri > hAlkAbhAri “light-heavy”, etc.
- (e) **Adjective + Noun > Noun:** banya + prANI > banyaprANI “wild animal”, bhadra + lok > bhadrAlok “gentleman”, shubha + kSaN > shubhakSaN “auspicious hour”, shunya + pad > shunyapad “vacant post”, him + ghar > himghar “cold store”, nll + kuThi > nllkuThi “blue bungalow”, kholA + ciThi > kholAciThi “open letter”, etc.
- (f) **Verb + Verb > Noun:** AnA + gonA > AnAgonA “coming and going”, mAra + dharA > mAraAdharA “hitting and catching”, dharA + bnAdhA > dharAbnAdhA “catching and binding”, marA + bnAcA > marAbnAcA “dying and living”, oThA + basA > oThAbasA “standing and sitting”, dekhA + shonA > dekhAshonA “seeing and hearing”, lekhA + paRA > lekhApaRA “writing and reading”, etc.

If we analyse the above examples we can find that the formative constituents of some of the compounds are potentially productive for generating new compounds by way of simple lexical association, since each constituent carries its own individual semantic load in the formation of the final form. But this rule does not equally apply for some other examples given above, since the final form is not equal in meaning with the total meanings of the formative constituents.

This phenomenon, however, shows that the constituents used in the formation of compounds belong to noun, adjective, and verb; and that most of the root compounds found in Bengali are productive in nature. But the question of identification of the underlying generative rules in the lexical productivity still remains unanswered.

Also, the question whether it is at all necessary to distinguish the productive type of compounds from the non-productive compounding types is not yet resolved.

5.2 Synthetic compounds

In contrast to the root compounds, Bengali synthetic compounds are formed by way of adding inflectional markers to one or both the formative constituents. The constituents, which are physically available within compounds, comprise with inflections, gerundial markers, participial forms, passive participles, and prefixes, to generate final forms. The patterns of formation of synthetic compounds with elements may be summarised in the following way:

- (a) Both the formative constituents are nouns. They are inflected with a locative case marker *-e*. E.g. [hAT]_{N-e}Case + [bAjAr]_{N-e}Case > [hATebAjAre]_N “in markets and other places”, [path]_{N-e}Case + [ghAT]_{N-e}Case > [patheghATe]_N “in roads and fields”, [kol]_{N-e}Case + [piTh]_{N-e}Case > [kolepiThe]_N “in laps and backs”, [dudh]_{N-e}Case + [bhAt]_{N-e}Case > [dudhebhAte]_N “with milk and rice”, [mA]_{N-ye}Case + [jhi]_{N-ye}Case > [mAyejhiye]_N “by mother and daughter”, [hAt]_{N-e}Case + [pA]_{N-ye}Case > [hAtepAye]_N “in hands and legs”, etc.
- (b) The first formative constituent is invariably a noun and it is inflected with a locative case marker *-e*. The last member is either a noun or an adjective or verbal noun which is intact or unaffected without any case or suffix. E.g. [hAt]_{N-e}Case + [khaRi]_N > [hAtekhaRi]_N “first education”, [gA]_{N-ye}Case + [halud]_N > [gAyehalud]_N “smearing body with turmeric paste”, [tel]_{N-e}Case + [bhAjA]_N > [telebhaAjA]_N “fried in oil”, [hAt]_{N-e}Case + [garam]_{Adj} > [hAtegaram]_N “served hot in hand”, etc.
- (c) The first formative constituent is either an adjective or a noun, which is without any suffix or case marker while the last member is invariably a noun, which is inflected with a locative case marker *-e*. E.g., [lAl]_{Adj} + [pAR]_{N-e}Case > [lAlpeRe]_{Adj} “with red border”, [chaRi]_N + [hAt]_{N-e}Case > [chaRihAte]_{Adj} “with a cane in hand”, [ek]_{Adj} + [ghar]_{N-e}Case > [ekghare]_{Adj} “to one house”, [cashmA]_N + [nAk]_{N-e}Case > [cashmAnAke]_{Adj} “glass-on-nose”, etc.
- (d) The following compounds are either gerundial or adverbial in sense. The formative constituents are either derived from verb or noun. The first constituent carries the gerundial or adverbial infix *-A*, while second member carries the suffix *-i* denoting same grammatical functions^[8]. E.g., [kAn]_{N-A}Inf. + [kAn]_{N-i}Suffix > [kAnAkAni]_N “whispering”, [dal]_{N-A}Inf. + [dal]_{N-i}Suffix > [dalAdali]_N “groupism”, [rAt]_{N-A}Inf. + [rAt]_{N-i}Suffix > [rAtArAti]_{Adv.} “within a nights”, [pAsh]_{N-A}Inf. + [pAsh]_{N-i}Suffix > [pAshaApAshi]_{Adv.} “side by side”, etc.
- (e) The following compounds are gerundial forms, which are derived from verbs denoting similar sense of action. Here both the formative members are tagged with a gerundial suffix *-A* or *-Ano* to denote nominal sense. E.g., [cal]_{FV-A}Inf. + [pher]_{FV-A}Suffix > [calApherA]_N “movement”, [sAj]_{FV-Ano}Inf. + [goch]_{FV-Ano}Suffix > [sAjAnogochAno]_N “arrangement”, [lekh]_{FV-A}Inf. + [paR]_{FV-A}Suffix > [lekhApaRA]_N “education”, [dekh]_{FV-Ano}Inf. + [shon]_{FV-Ano}Suffix > [dekhAnoshonAno]_N “exhibition”, etc.
- (f) The following compounds are adjectival compounds, where the second member contains a (non)verbal passive participles while the first one is free form any

kind of addition. The first member is invariably a noun while the second one is a (non)verbal passive participle with a suffix *-A*, *-Ano* or an adjectival form derived from noun. The final form of the compound denotes a sense of completeness of action mostly carried by the second member^[9]. E.g., [man]_N + [maj]_{FV-Ano}Suffix > [manmajAno]_{Adj} “mind-soothing”, [sukh]_N + [jAg]_{FV-Ano}Suffix > [sukhjAgAno]_{Adj} “pleasure-arousing”, [jal]_N + [to]_{FV}- A]Suffix > [jaltolA]_{Adj} “water-pulling”, [hAt]_N + [dhar]_{FV-A}Suffix. > [hAtdharA]_{Adj} “hands holding”, etc.

- (g) The following compounds are formed after adding a prefix with the root word. They show that the first formative member is invariably a prefix and the second member is either a verb, or a noun, or an adjective. The final lexical class of the compound is mostly dependent on the prefix although at certain cases, the second member can hold its lexical class. E.g., [a-]_{Prefix} + [cal]_{FV} > [acal]_{Adj} “obsolete”, [a-]_{Prefix} + [mAnuS]_N > [amAnuS]_{Adj} “inhuman”, [bi-]_{Prefix} + [mAtA]_N > [bimAtA]_N “stepmother”, [nir-]_{Prefix} + [dhan]_N > [nirdhan]_{Adj} “poor”, [A-]_{Prefix} + [kAl]_N > [AkAl]_N “famine”, [nIh-]_{Prefix} + [rab]_N > [nIrirab]_{Adj} “silent”, etc.

To conclude, the range of structural types of the synthetic compounds in Bengali is not vastly different from the range of root compounds. Since root compounds are generated lexically, we assumed that synthetic compounds are also generated in similar lexical process. Although we argue that both synthetic and root compounds are the results of syntactic processes, we feel that certain aspects of synthetic compound formation are lexical while other aspects are the results of pure syntactic process.

5.3 Bengali Compounds: Syntactic or Lexical?

There are confusions regarding the question of satisfying the argument structure in compounds. Scholars are in fact grouped into two different camps with regard to the question of analysing compounds in this way. The scholars who consider a compound word as a condensed form of a sentence (Fabb 1984, Roeper and Siegel 1978, Spoot 1985) claim that compound formation takes place largely or solely at the sentence level. Therefore, syntactic rules can manipulate the innermost structure not only of compounds but also of idioms and simple word units.^[10] At the other hand, the scholars who are willing to consider compound as single lexical unit (Selkirk 1983, Lieber 1983, Katamba 1993) argue that syntactic rules applied to the level of a phrase cannot take into account the internal structure of compounds and words. Therefore, compounding is a type of word formation process, and hence, it is a matter of simple lexical organization. The syntactic effects noted in compounds are the result of essentially morphological rules and principles permitting argument structure to be accessible outside the domain of the word itself.

According to the later group, although the internal structure of compounds is formed by using the syntactic rules of a language that generate phrases and sentences, they cannot access the internal structure of compounds. Whatever the internal structure a compound has, it is inaccessible to the rules of syntax. Words are *syntactic atoms* while compounds are affixed words, which are meant to be treated as indivisible units to be used in syntactic rules. Syntactic rules are able to manipulate the elements inside phrasal units but they cannot manipulate the elements inside the words, no matter whether they are affixed words or compounds.^[11]

However, to dissolve the ongoing debate we argue that it is probably better to propose that compound is a distinct form of linguistic expression, which contains

more than one combinatory process in it to exhibit characteristics of both *syntactic* and *lexical* expressions. It resembles to a sentence because the units, which are combined to generate the final form, may also exist as independent words with their complete phonetic, orthographic, and semantic entities. Compounding is a generative process where both morphology and syntax (also something else) are intermixed.

Therefore, compound gives us clues to look at the synthetic aspects embedded in the rules of syntax as well as to look at the problem of how to define the notion of *word* in a language (Spencer 1991: 309). When two different words (as opposed to roots) are compounded each one is a *minimal free form* by its entity and definition. They are combined in a way where both syntactic and lexical rules are equally operational to address specific linguistic needs. Therefore, the question ‘whether a compound is a word’ may be dissolved in the following manner: if we regard compounding process as essentially syntactic, then the answer is ‘no’; and if we consider compounding as a morphological process then the answer is ‘yes’ (Spencer 1991: 43).

Bengali compounds are normally considered as single lexical units although these are derived either from concatenation of two different lexical items or phrases. For instance, copulative and numeral compounds are formed from the concatenation of two lexical units while determinative, descriptive, adverbial and reciprocal compounds are formed out of phrasal units. If we consider the following examples analytically (Table 3) we can probably understand their single lexical entity.

Surface form	Literal meaning	Lexical meaning
shItgrISma	winter-summer	round the year
sAtsamudra	seven-seas	around the world
padadalita	foot-trodden	poor or destitute
madhupa	honey-sucker	honey bee
AkAshbAnI	oracle	radiogram
gajAnan	elephant-face	Lord Ganesha

Table 3: The lexical entity of Bengali compounds

The compound words given above (Table 3) are formed by way of simple side-by-side placement of two lexical units either belonging to noun or adjective. The final form is a single lexical unit where paradigmatic replacement of any member with another will destroy the balance of the structure and meaning of the final form. Moreover, following the standard rules of inflection applicable in Bengali, the second member carries the suffix or inflectional properties while the first member remains untouched. Therefore, they may be considered as examples of single lexical units in which morphological rules applied to single word units are equally applicable to these forms.

However, a closer look into their mode of formation (with reference to their literal and connotative meanings) shows that the literal meanings of the members fail to generate the meaning of the final form we understand. Moreover, they behave like phrasal forms since they are derivable from phrases by a simple process of lexical generalisation.

That means, in case of *shItgrISma*, the two months stand for the whole year; for *sAtsamudra*, seven oceans imply the whole world; in *padadalita*, the second member symbolises a class of people oppressed by changing the semantic load of the

members; in *madhup*, a token stand for the type by way of particularisation of meaning; in *AkAshbAnI*, the semantic load of the members is changed by figurative connotation, and in *gajAnan*, a picture of conceptual similarity is drawn between the first and last the members, which act as simile rather than simple reference. In essence, the members have gone through a kind of semantic shift at the time of final form generation. They act more like phrasal units where syntactic rules are allowed to operate in the surface form. Also, paradigmatic substitution of the constituents destroys their embedded connotative meanings. In support of the syntactic nature of compounds we may analyse the examples given below to find Bengali compounds to be syntactic in nature:

- (a) bAyu duSaN
- (b) bAyu duSaN niyantraN
- (c) bAyu duSaN niyantraN parSad
- (d) rAjya bAyu duSaN niyantraN parSad
- (e) pashimbanga rAjya bAyu duSaN niyantraN parSad

It is now evident that in many respects Bengali compounds represent a close and non-definable interface between morphology and syntax. We can probably assume that a compound is one of the ways by which a word is formed by adding another word(s), which are supposed to be syntactically connected.

Therefore, mere congregation of words into one form will not make any compound; there must be some amount of syntactical and semantic connections between the formative constituents. If words are not syntactically and semantically related to each other, they cannot form compounds. This implies that compound words should be discussed in between morphology and syntax because compounds are transformed versions of sentences of which at least two lexical items are juxtaposed in such a way that their semantic relationships are never understood from their apparent surface structures.

6. Application of 'Head' in Bengali Compounds

The notion of *head* in word analysis plays a crucial role in generative morphology. Within this frame of lexical analysis it is argued that the head plays important role in description of the distribution of diacritic features related to both inflectional and derivational morphology (Selkirk 1983: 21).

This observation considered relevant for single word units is equally applicable to compounds because 'head' becomes crucial in characterising the semantic properties of compounds. The class of elements denoted by compounds is a subset of the class denoted by their heads.

The non-head constituent of compounds, in some way or other, shows that the head is the nucleus of compound, although the semantic relation obtaining between the head constituent and the non-head constituent vary considerably.

With regard to the 'headness' principle, all compound words are normally classified using the following three criteria:

- (a) Whether a compound has a head,
- (b) If it has a head then what is the word class of the head, and
- (c) Whether the head appears at the left or right of the compound.

Using the model of X-bar syntax, scholars highlight that just as phrases in syntax, words also have heads. Syntactically, head is a dominant constituent of a compound word. As an outcome of this feature, inflectional (grammatical) properties of the head (i.e. tense, number, case, etc.) percolate to the entire compound. This entails that the head of a nominal compound will be a noun, the head of a verbal compound will be a verb, and the head of an adjectival compound will be an adjective.

If we subscribe to this principle we can find that for endocentric compounds the head element normally appears as the right-most member, whose lexical class (e.g. *noun*, *verb*, *adjective*, etc.) percolates the entire compound (e.g. *greenfly*, *bookcase*, *textbook*, *mailbox*, etc.). In other words, the head of the compound actually determines the lexical class of the final form. This generalisation is captioned as *Right Hand Head Rule* (RHHR) for compounds (Williams 1981: 248).

In the field of generative morphology we may define the head of a morphologically complex word to be the right-hand member of that word. In simple word structure the head is defined in terms of the position of a constituent, not in terms of a relation between categories based on their respective types (i.e. levels) and feature complexes. In case of inflected compounds the RHHR applies to affixes i.e. the rightmost suffix in a word assigns its property by percolation to the entire word.^[12] This, in fact, requires revision of the traditional assumption, which proposes that stems and bases select their affixes based on their natures and lexical classes. Thus, while nouns select nominal suffixes (e.g., *-er*, *-ion*, *-ment*, etc.), verbs select verbal affixes (e.g. *-ed*, *-ing*, etc.), and adjectives select adjectival suffixes (e.g. *-er*, *-est*, *-ful*, etc.), and so on.

The RHHR is certainly not the universal principle that regulates compounding in all languages. There are languages like Italian, French, and Vietnamese where the head of a compound is normally the left-hand member. In French, for instance, the plural inflection is mostly attached to the first noun in the compound where the RHHR does not apply.

On the other hand, although the head of a compound is (normally) on the right in English, there are small numbers of endocentric compound where left-hand heads control the final form. They include nouns which form their plural by adding the plural morphemes to the noun in first position (e.g., *passer-by* : *passers-by*, *son-in-law* : *sons-in-law*, etc.)

These exceptional examples force scholars to observe that genuine generalisation, which one should attempt to capture, is that inflectional affixes are (normally) attached to the head, regardless of whether it is on the left or on the right of a compound.

The formative members of Bengali compounds belong either to noun, verb, adjective, or affix. Although there are compounds made up without an affix, the use of affix in formation of compounds is not an event to be ignored. A closer analytical look can easily find that Bengali compounds generally contain the following types of lexical units most of which are potentially competent to be inflected:

- (a) $W_{[N]} + W_{[N]}$,
- (b) $W_{[N]} + W_{[Adj]}$,
- (c) $W_{[NV]} + W_{[FV]}$,

- (d) $W_{[Prefix]} + W_{[N]}$,
 (e) $W_{[Prefix]} + W_{[Adj]}$,
 (f) $W_{[Adj]} + W_{[Adj]}$, and
 (g) $W_{[Adj]} + W_{[N]}$.

In case of adverbial, numeral, determinative, and descriptive compounds, the right hand member is functionally important since it carries the semantic load to be the head of the final form. Even at the time of inflection it carries the suffix and case properties, which percolate to the entire compound.^[13] E.g., *pratidiner* “of everyday”, *upakule* “on coast”, *yathAshaktite* “with all strength”, *shatAbdIr* “of century”, *kAlidAser* “of Kalidas”, *jIbanbimAy* “in life insurance”, *prANpAkhike* “to life bird”, etc.

There is another type of compound in Bengali, namely the copulative compound, where both the heads carry equal amount of morphological importance and semantic mileage. E.g., *kAgajpatra* “papers”, *janamAnab* “people”, *pAtramitra* “courtiers”, *rAtdin* “day and night”, *chelemeye* “boys and girls”, *sAdAkAlo* “black and white”, *choTabaRa* “small and big”, *sukhdukSa* “joy and woe”, etc.

From semantic point of view, since the coupled members are of equal status neither of the members should be regarded as a head to dominate the final form. Moreover, these are semantically transparent, since each member characterises a separate and distinct aspect of meaning of the final form. From syntactic point of view, these compounds are however, right-headed compounds, because the right-hand member abides by RHRs when it uses affixes or inflections. E.g., *kAgajpatr-er* “of papers”, *janamAnab-er* “of people”, *pAtramitra-ke* “to courtiers”, *rAtdin-er* “of day and night”, *chelemeye-gulike* “to the boys and girls”, *sAdAkAlo-y* “in black and white”, *choTabaRa-r* “of small and big”, *sukhdukS-er* “of joy and woe”, etc.

There is yet another type of compound, which may be labelled as ‘headless compound’ (e.g. *greenhouse*, *darkroom*, *redcap*, *sweetheart*, *black-box*, *turncoat*, etc.). These compounds do not have any member, which can act as a head to control the non-head element.

In Bengali such compounds are called *bahubrIhi* (reciprocal) compounds where something other than the formative members becomes important semantically. E.g. *muralIdhar* “Lord Krishna”, *mahAjan* “moneylender”, *pANigrahaN* “marriage”, *mirjaphar* “traitor”, *nIlkhaNTha* “Lord Shiva”, *ratnAkar* “ocean”, etc. From syntactic point of view these compounds are generated by standard rule that generates regular endocentric compounds in the patterns like $[[muralI]_N [dhar]_N]_N$, $[[mahA]_{Adj} [jan]_N]_N$, following RHRs. For these, the rightmost member determines the lexical class of the final form. Moreover, affixes and inflectional markers are tagged to the right hand member following the standard methods of inflection.

Form semantic point of view, however, these are derived in the patterns like $W_1 + W_2 > W_3$, where the internal meanings of W_1 and W_2 are entirely lost and a new meaning is assigned to W_3 . In these cases at least, the final forms are semantically opaque because it is impossible to derive what a final form means from the sum total of meanings of its constituents. For instance, *pANigrahaN* is not an act of holding somebody’s hands but to marry a girl where holding hands of the bride is a ritualistic act performed by the groom.^[14]

7. Conclusion

In this paper we have made an attempt analyse the structure, the meaning, and the modes of formation of Bengali compounds to process them automatically by computer for various works of language technology. We have also made an effort to observe the variations in their orthographic structure whenever these compounds undergo any kind of morphophonemic reformation or alternation. The investigation is necessary to know how the compounds behave at the time of inflection, derivation, sandhi or at the time of using case markers. We have tried to explore to know what types of change take place in the structure of the participating members at the time compound formation; and whether the lexical class of the formative members are at all affected in final outputs. Also, the syntactic and semantic aspects of the members are taken into consideration with a goal for using this information at the time of automatic processing by computer.

It is noted that the structure of compounds is not formed by the syntactic rules that are applied to construct phrases. Compound words and phrases are two different linguistic properties, which are not subject to the same set of syntactic principles. Also, they differ from idioms in the sense that while idiomatic expressions are subject to the same syntactic rules applied for the phrases, compounds are different because syntactic operations are hardly applicable on the constituents of compounds.

It is also demonstrated that compounds differ from phrases and sentences in the way their meanings are accessed. The meanings of exocentric or *bahubrIhi* compounds are idiosyncratic and non-compositional, since they cannot be computed from the meanings of their constituents. The semantic idiosyncrasy and unpredictability of a large set of compounds makes it necessary to list their meanings in the lexical database separately. On the contrary, the regularity noted in their formation makes us argue that most of the compounds should not be listed in the lexicon because they can be generated in a rule-governed manner. Probably, we require a semantico-syntactic interface to dissolve the dichotomy arises from structural and semantic interpretations of compounds – a unique lexical class of its own in a natural language.

Notes

- [1] A notable feature of nominal system of a natural language is the very common use of nominal compounds, which may be huge (10+ words) as in some modern languages such as, German. Nominal compounds occur with various structures, however morphologically speaking they are essentially the same. Each noun (or adjective) is in its (weak) stem form, with only the final element receiving case inflection.
- [2] This inspires some scholars (Bloomfield 1933: 233) to argue that, “Linguists often make the mistake of taking for granted the universal existence of whatever types of compound words are current in their own language. It is true that the main types of compound words in various languages are somewhat similar, but this similarity is worthy of notice; moreover, the details, and especially the restrictions, vary in different language. The differences are great enough to prevent our setting up any scheme of classification that would fit all languages”.

- [3] If we do so, the original intended meaning of the compound words will be either reversed or lost entirely. For instance, in Bengali, the order of the members (W_1 and W_2) in a compound like *chelemeye* means “sons and daughters”. But if we reverse the order of the members (W_2 followed by W_1) we will generate a form like *meyechele*, the meaning of which is “a kept”.
- [4] Further analysis of the text helps us to argue that in the colloquial versions of the Bengali texts the number of compound words is drastically reduced by way of lexical decomposition for their better comprehension and access.
- [5] Since this observation goes against the observation made by Mukhopadhyay (1999), it is sensible to have a separate and independent investigation across the text of various types with empirical support from a few large, multidisciplinary, diachronic, and monolingual corpora before we come to trust-worthy conclusion about the increasing or decreasing scenario of Bengali compounds.
- [6] It could be called a ‘caturthi tatpuruSa’ since caturthi refers to the fourth case, i.e. the dative case. Incidentally, tatpuruSa implies to someone’s agent, while caturthi tatpuruAa is a karmadhArya, being both dative and tatpuruSa. An easy way to understand it is to look at English examples of tatpuruSa such as *battlefield*, where there is a genitive relationship between *field* and *battle* “a field of battle”. Other examples include instrumental relationship (e.g., *thunderstruck* etc.) and locative relationship (e.g., *citydwelling*, etc.).
- [7] Dvandva (copulative) compounds consist of two or more noun stems, connected in sense with ‘and’ (e.g. *mAtApitA* “mother and father”, *bhAibon* “brother and sister”, etc.). Due to these reasons, dvanda compounds having more than one noun in them are treated as dual or plural forms.
- [8] Although these forms are generally considered as compound forms in standard Bengali grammars, scholars have argued that these should be better considered as reduplicated words rather than compounds (Chaki 1996: 231).
- [9] With examples from English (e.g. *sun-baked*, *moth-eaten*, *bed-ridden*, *home-spun*, *hand-made*, *god-gifted*, *horse driven*, *famine-stricken*, etc.) it is argued that the interpretation of these compounds is ultimately based on the lexical form of the participle, and not on that of the (active) verb from which it is derived. Therefore, these compounds are assigned to the structure such as: $A[NA]_A$ where the adjective is based on the passive participle of a verb (Selkirk 1983: 41).
- [10] Even some of the scholars explicitly questioned whether there is any organization of the lexicon beyond its function as a list of idiosyncrasies (Spencer 1991: 344). Moreover, they claimed that many words, including the morphophonemic ones (e.g. *drop*, *kill* etc.) are not listed in the lexicon but are created as a result of simple syntactic manipulations.
- [11] The nucleus of this argument is summarised as: “... syntactic rules apply to words regardless of their internal structure. The morphological and syntactic characteristics of a word are independent of each other. So, in syntax, morphologically complex words do not behave differently from words containing un-derived simple roots” (Katamba 1993: 299). Therefore, “we shall reject the speculations of the generative semanticists that syntax can access the internal structure of words and maintain the lexicalist position” (Katamba 1993: 301).
- [12] It is argued that for an affix to determine the properties of its word, it must appear in the ‘ultimate’ head position.
- [13] This observation goes against the traditional observation where it is claimed that in case of adverbial and numerical compounds the first constituent,

occurring at left-hand position, controls the semantic load of the compound as proposed by Patanjali.

- [14] This particular lexico-semantic feature of compounds leads us to argue that the creation of exocentric compounds is less frequent than endocentric compounds in case of new compound formation in a language like Bengali.
-
-

Acknowledgement

I acknowledge the help I have received from some of my students as well as from some foreign scholars who used to send me emails quite often asking about various aspects of Bengali compounds, in particular, and Sanskrit compounds, in general. In fact, they have given me an opportunity to shoehorn my ideas about this particular phenomenon of Bengali while answering their queries.

References

- Adams, V. (1973) *An Introduction to Modern English Word Formation*. Longman: London.
- Anderson, S.R. (1988) "Morphological Theory", in Newmeyer, F.J. (Ed.) *Linguistics: The Cambridge Survey*. Cambridge: Cambridge University Press. Pp. 169-191.
- Banerji, Satya Ranjan (1997) "Sanskrit and Greek Compounds". *Bulletin of the Dept. of Linguistics*. Calcutta University, Kolkata. C. Dutt Memorial Number 13. Pp. 5-14.
- Banerji, Satya Ranjan (2000) *A Handbook of Sanskrit Philology*. Kolkata: Sanskrit Pustak Bhandar.
- Bhattacharya, Sumita (1983) *The Study of Bengali Compounds*. Unpublished Doctoral Dissertation. Calcutta University, Kolkata.
- Bhattacharya, Sumita (1997) "Compounding or derivation?", *Bulletin of the Dept. of Linguistics*. Calcutta University, Kolkata. C. Dutt Memorial Number. 13. Pp. 53-55.
- Bloomfield, Leonard (1933) *Language*. New York: Holt, Rinehart & Winston.
- Bybee, Joan L. (1985) *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.
- Chaki, Jyoti Bhusan (1996) *Bangla Bhasar Byakaran*. Kolkata: Ananda Publishers.
- Chakravarty, Bamandev (1974) *Ucchatara Bangla Vyakaran*. Kolkata: Sarkar and Co.
- Chatterji, Suniti Kumar (1926) *The Origin and Development of the Bengali Language* (Reprinted in 1993 by Rupa and Co.)
- Chatterji, Suniti Kumar (1995) *Bhasa Prakash Bangala Byakaran*. Kolkata: Rupa & Co.

- Dash, Niladri Sekhar (2005) *Corpus Linguistics and Language Technology: With Reference to Indian Languages*. New Delhi: Mittal Publications.
- Dash, Niladri Sekhar (2005) "The role of context in sense variation - introducing corpus linguistics in Indian contexts". *Language In India*. 5(6): 1-32.
- Fabb, Nigel (1984) *Syntactic Affixation*. Unpublished Doctoral Dissertation, MIT.
- Hockett, Charles F. (1963) *A Course in Modern Linguistics*. New York: Macmillan.
- Jensen, John T. (1990) *Morphology: Word Structure in Generative Grammar*. Amsterdam: John Benjamins.
- Jespersen, Otto (1929) *A Modern English Grammar on Historical Principles*. Copenhagen: Ejnar Munksgaard.
- Katamba, Francis (1993) *Morphology*. London: The Macmillan Press Ltd.
- Leech, Geoffrey, B. Francis, and X. Xu 1994. "The use of computer corpora in the textual demonstrability of gradience in linguistic categories". In, Fuchs, C. and B. Vitorri (Eds.) *Continuity in Linguistic Semantics*. Amsterdam: John Benjamins. Pp. 31-47.
- Lees, Robert B. (1960) *The Grammar of English Nominalizations*. Bloomington, IN: Indiana University Press.
- Lieber, R. (1980) *On the Organization of the Lexicon*. Cambridge, Mass.: MIT Press.
- Lieber, R. (1983) "Argument linking and compounding in English". *Linguistic Inquiry*. 14(2): 251-286.
- Lieber, R. (1988) "Phrasal compounds in English and the morphology-syntax interface". *Papers of the Chicago Linguistic Society*. 24: 202-222.
- McCawley, J. (1968) "Lexical insertion in generative grammar without deep structure", *Papers of the Chicago Linguistic Society*. 4(1): 71-80.
- Mukhopadhyay, Ashok (1999) "Artha banam samartha". *Korak*. Dictionary No. Pp. 32-40.
- Murchand, Hans (1969) *The Categories and Types of Present-Day English Word Formation*. Munchen: C. H. Beck.
- Ray, Punya Sloka (1997) *Bengali Language Handbook*. Kolkata: Bangla Akademi.
- Roeper, T and M.E.A. Siegel (1978) "A lexical transformation for verbal compound". *Linguistic Inquiry*. 9(2): 199-260.
- Sarkar, Pabitra (1998) "Bangla rupatattver bhumika". *Bahubacan*. 1(1): 81-106.
- Sarkar, Pabitra and Ganesh Basu (1994) *Bhasa-jijnasa*. Kolkata: Vidyasagar Pustak Mandir.

- Scalise, S. (1984) *Generative Morphology*. Dordrecht: Foris.
- Selkirk, Elizabeth O. (1983) *The Syntax of words*. Cambridge, Mass.: MIT Press.
- Sen, Sukumar (1993) *Bhasar Itivritta*. Kolkata: Ananda Publishers.
- Sen, Sukumar and Subhadra Kumar Sen (1994) *Bangalir Bhasa*. Kolkata: Bangla Akademi.
- Spencer, Andrew (1991) *Morphological Theory*. Oxford: Basil Blackwell.
- Spoart, Richard (1985) *On Deriving the Lexicon*. PhD dissertation.
- Thompson, L. (1965) *A Vietnamese Grammar*. Seattle: University of Washington Press.
- Williams, E. (1981) "On the notions 'lexically related' and 'head of a word'".
Linguistic Inquiry.12(2): 234-274.
-

Niladri Sekhar Dash, Ph.D.
Linguistic Research Unit
Indian Statistical Institute
Kolkata, India
niladri@isical.ac.in