

## Exocentric Compounds in Classical Sanskrit

Brendan Gillon

► **To cite this version:**

Brendan Gillon. Exocentric Compounds in Classical Sanskrit. First International Sanskrit Computational Linguistics Symposium, INRIA Paris-Rocquencourt, Oct 2007, Rocquencourt, France. inria-00202860

**HAL Id: inria-00202860**

**<https://hal.inria.fr/inria-00202860>**

Submitted on 8 Jan 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# EXOCENTRIC (BAHUVRĪHI) COMPOUNDS IN CLASSICAL SANSKRIT

Brendan S. Gillon  
McGill University

## 1 Introduction

Constituency grammars originated with Leonard Bloomfield (1933) and were developed during the nineteen forties and nineteen fifties by a number of American structuralist linguists, including Harris (1946) and Wells (1947) — to mention just two. In the late nineteen fifties, Chomsky (1957) suggested that constituency grammars could be formalized as context free grammars. It is now clear that context free grammars fall short of properly formalizing the constituency grammars of the American Structuralists (Manaster-Ramer and Kac 1990). Indeed, in the nineteen sixties, Chomsky himself formalized a number of other important aspects of constituency grammars, introducing more complexity to the labels, both terminal and non-terminal, and permitting the use of null elements.

Though constituency grammars were initially conceived of as applying to phrases, work in the nineteen eighties (Di Sciullo and Williams 1987; Selkirk 1982; Williams 1981) showed that such rules could be used to analyze profitably compounds and derivational morphology of English. Gillon (1995) showed that the same analysis extended to the analysis of compounds and derivational morphology in Classical Sanskrit.

The aim of this paper is to look more carefully at the application of constituency rules to the analysis of exocentric (*bahuvrīhi*) compounds. In particular, using a conjecture by Bhartṛhari pertaining to non-constituent (*a-samārtha*) compounds, I wish to show that both Sanskrit non-constituent (*a-samārtha*) compounds and Sanskrit exocentric (*bahuvrīhi*) compounds can be analyzed in terms of constituency grammar, suitably augmented with argument frames, a generalization of subcategorization frames. Argument frames, as I shall show, play a key role in providing a satisfactory analysis of both kinds of compounds.

## 2 Exocentric (*Bahuvrīhi*) Compounds

The classical Indian grammatical tradition identifies a number of different kinds of exocentric (*bahuvrīhi*) compounds. They include: privative exocentric (*nañ-bahuvrīhi*) compounds, comitative exocentric (*saha-bahuvrīhi*) compounds, prepositional exocentric (*prādi-bahuvrīhi*) compounds, homo-denotative exocentric (*samānādhikaraṇa-bahuvrīhi*) compounds, and hetero-denotative exocentric (*vyadhikaraṇa-bahuvrīhi*) compounds. Our attention will be confined to homo-denotative ones (*samānādhikaraṇa-bahuvrīhi*).

Homo-denotative exocentric (*samānādhikaraṇa-bahuvrīhi*)<sup>1</sup> compounds are compounds whose canonical phrasal paraphrase is a relative clause in which the first constituent of the compound is predicated of the second, and they thereby share the first, or nominative, case – and if the first constituent is an adjective, they agree in number and gender as well. Putting these observations together, one arrives at the natural hypothesis that an exocentric compound is a descriptive compound suffixed with a phonetically null suffix which converts the descriptive compound into an adjective.

- (1) Exocentric (*bahuvrīhi*) Compound:  
Compound: *samacittaḥ* (even-minded)  
Analysis:  $((_A \text{ sama}) \prec (_N \text{ cittaḥ}))_1$   
 $((_A \text{ even}) \prec (_N \text{ mind}))$   
Paraphrase:  $[_{RC} [_{VP} \text{ (asti)} [_{AP1} \text{ samam} ] [_{NP1} \text{ cittaṃ} ] \text{ yasya} ]$   
 $(\text{is}) \quad \text{even} \quad \text{mind} \quad \text{whose}$   
whose mind is even

The evidence that homo-denotative exocentric (*samānādhikaraṇa-bahuvrīhi*) compounds are adjectives is that they have all the properties adjectives in Sanskrit have. First, adjectives in Sanskrit, like those in Latin, agree with the nouns they modify in case, number and gender. Consider the adjective *tīkṣṇa* (*sharp*). If it modifies a noun in the nominative, singular, masculine, say *asiḥ* (*sword*), then it has the form *tīkṣṇaḥ*; and if it modifies a noun in the nominative, singular, feminine, say *churī* (*knife*), then it has the form *tīkṣṇā*; and

---

<sup>1</sup>Homo-denotation (*samānādhikaraṇa*) is the counterpart in the Indian grammatical tradition of the Western technical notion of concord or agreement. As will be elaborated below, adjectives which modify nouns in Sanskrit agree with the nouns in case, number and gender. The concord is seen by the traditional Indian grammarian as accruing to the fact that a noun and an adjective modifying it have the same denotation (*samānādhikaraṇa*).

if it modifies a noun in the nominative, singular, neuter, say *patram* (*blade*), then it has the form *tīkṣṇam*. Now consider the compound *dīrgha-kaṇṭha*. If it is to be construed with a masculine, singular noun in the nominative case, say *puruṣaḥ* (*man*), to yield the sense *long-necked man*, then the compound must have the nominative, masculine, singular form, namely, *dīrgha-kaṇṭhaḥ*. If it is to be construed with a feminine, nominative, singular noun, say *strī* (*woman*), to yield the sense *long-necked woman*, then the compound must have the feminine, nominative, singular form, *dīrgha-kaṇṭhā*. And finally, if it is to be construed with a neuter, nominative, singular noun, say *mitram* (*friend*), to yield the sense *long-necked friend*, then the compound must have the neuter, nominative, singular form, *dīrgha-kaṇṭham*.

Next, adjectives in Sanskrit can be turned into abstract nouns by the affixation of the suffix *-tva* (*-ness*): for example, the adjective *kṛśa* (*thin*) may be converted into the abstract noun, *kṛśa-tva* (*thin-ness*). Exocentric *bahuvrīhi* compounds are susceptible of the same conversion: for example, *dīrgha-kaṇṭha* (*long-neck-ed*; cf., *level-head-ed*) be turned into *dīrgha-kaṇṭha-tva* (*long-neck-ed-ness*; cf., *level-head-ed-ness*).

Moreover, just as an adjective such as *kṛśaḥ* (*thin*) can function, as its English translation can, as a common noun, meaning the same thing as its English nominal counterpart, *the thin*, so too should an exocentric (*bahuvrīhi*) compound be liable to function as a common noun. And this too is true, as observed by Speijer (1886, §222, fn. 1) and as exemplified by following compound and its commentarial gloss.

(2.1) NBT 48.4  
(vyutpanna- $\leftarrow$ saṁketasya)

(2.2) NBTP 49.1-2  
vyutpannaḥ jñātaḥ saṁketaḥ yena saḥ  
arisen known convention by whom he  
One by whom the conventions of language are known  
(*jñāta* (*known*) glosses *vyutpanna* (*arisen*).)

There is independent confirmation that homo-denotative exocentric (*sa-māna-adhikaraṇa-bahuvrīhi*) compounds are best treated as descriptive (*kar-madhāraya*) compounds to which a phonetically null, possessive, adjectival suffix (symbolized hence forth with ‘B’) is affixed. Sanskrit has a phonet-

ically overt, possessive, adjectival suffix *-ka* which is virtually synonymous with the phonetically null one just hypothesized. Though their distributions are somewhat different (A 5.4.151 ff.), nonetheless, they overlap to such an extent that commentators to a text in which an exocentric (*bahuvrīhi*) compound occurs frequently repeat the compound, adding the *-ka* suffix to signal the fact that the compound in question is to be construed, not as a descriptive (*karmadhāraya*) compound, but as an exocentric (*bahuvrīhi*) compound (Boose and Tubb 1981, ch. 5, sec. 15).

English too has homo-denotative exocentric compounds. By and large, they are marked by the adjectival, possessive suffix, *-ed*. These English compounds, exemplified by such compounds as *longlegged*, *literal-minded*, and *two-footed*, have a distribution narrower than that of its counterpart in Sanskrit — a fact which will be dilated on below.

Not every English homo-denotative exocentric compound has the *-ed* suffix. In particular, English homo-denotative exocentric compounds which serve as proper names or epithets seem to require a phonetically null counterpart to the *-ed* suffix.<sup>2</sup> Examples of proper names are particularly common in children’s stories. For example, in the children’s movie, *Land Before Time*, the two dinosaurs which are the main characters are named, *big foot* and *long neck*, instead of *big-footed* and *long-necked*. Examples of epithets are such compounds as *red-head*, *dim-wit*, *hard-back*, etc., to which there correspond *red-headed*, *dim-witted*, *hard-backed*, and so forth. (See Marchand 1969, ch. 2, sec. 18 for other examples.) Moreover, it seems that the *-ed* suffix and its phonetically null counterpart are in free variation in exocentric compounds which are initial constituents in larger compounds: *long-necked bottle plant* and *long neck bottle plant* both denote plants for bottles whose necks are long.

Another parallel between English and Sanskrit homo-denotative exocentric (*samāna-adhikaraṇa-bahuvrīhi*) compounds is the predication relation in the canonical paraphrase may be metaphorical, instead of literal. Thus, in the compounds *candra-mukha* (*moon-faced*), *sthūla-caraṇa* (*club-footed*), and *ayo-muṣṭi* (*iron-fisted*), a face (*mukha*) is likened unto a moon (*candra*), a foot (*caraṇa*) unto a club (*sthūla*), and a fist (*muṣṭi*) unto iron (*ayas*).

---

<sup>2</sup>It is interesting to note in this connection that the Sanskrit suffix *-ka*, used to mark phonetically a *bahuvrīhi* compound, is said by Pāṇini (A 5.4.155) to be prohibited from affixation to *bahuvrīhi* compounds which serve as names.

As noted by Di Sciullo and Williams (1987, p. 30), in English, constituents outside of a compound cannot be construed with constituents subordinate within a compound. This generalization is undoubtedly true of English compound formation, as illustrated by the contrast in the interpretability of the expressions in (3).

- (3.1) ((man <eating) <shark)  
 (3.2) \*(eating <shark) of men

Interestingly, this generalization was presupposed as true of Sanskrit by Pāṇini. His treatment of compounds is to pair them with canonical phrasal paraphrases with which they share a common derivational ancestor; in addition to their semantic relation, they bear the syntactic relations of having the same heads and of having the same constituency. Hence, the constituency of compounds mirrors that of their canonical phrasal paraphrases. A condition on compound formation is that two elements cannot undergo compounding, the deletion of morphology from the subordinate element, unless the two elements form a constituent (A 2.1.4). A consequence of this is that inflected lexical items exterior to a compound are not construable with subordinate constituents within it. The applicability of this rule is illustrated both by Patañjali, in his *Mahābhāṣya*, or *Great Commentary*, on Pāṇini's *Aṣṭādhyāyī* (at A 2.1.1), and by Bharṭṛhari, in his work on the semantics of Sanskrit (VP 3.14.46), with the following example:

- (4.1) [<sub>NP<sub>1</sub></sub> ((ṛddha <rāja ) <puruṣaḥ ) ]  
           rich     king         man  
           servant of a rich king  
 (4.2) \* [<sub>NP<sub>1</sub></sub> [<sub>AP<sub>6</sub></sub> ṛddhasya ] (rāja <puruṣaḥ ) ]  
                   of rich             king-man  
           servant of a rich king

Thus, the expression in (4.1) is acceptable, whereas the one in (4.2) is not, as signalled by the asterisk.

Though the generalization holds of English compounds, it does not of Sanskrit compounds. Counter-examples are furnished both by Patañjali (MBh on A 2.1.1) and by Bharṭṛhari (VP 3.14.47):

- (5.1) [ $NP_6$  [ $NP_6$  Devadattasya ] guroḥ ] kulam  
of Devadatta of teacher family
- (5.2) (Devadatta- $\leftarrow$ guru)- $\leftarrow$ kulam  
Devadatta-teacher-family
- (5.3) [ $NP_6$  Devadattasya ] (guru- $\leftarrow$ kulam)  
of Devadatta teacher-family  
Devadatta's teacher's family

Indeed, compounds appearing in configurations such as that in (5.3) are given a special name by Sanskrit grammarians: they call them *a-samartha* compounds (i.e., non-constituent compounds). Moreover, these compounds are well attested in the classical literature. A study of over three-hundred sentences, chosen essentially at random from the Sanskrit corpus, reveals thirteen clear cases of non-constituent (*a-samartha*) compounds. (See Appendix I in Gillon 1993.) And a study of the first approximately five-hundred sentences of a single text reveals forty-three clear cases. (See Appendix II in Gillon 1993.)<sup>3</sup>

Thus, for example, in the best known play by the finest dramatist of Sanskrit literature, Kalidāsa's *Śakuntalā*, one finds precisely these configurations.

- (6) Ś 3.9.16 (= SG 3.1.6)  
[ $NP_1$  [ $NP_3$  [ $NP_7$  tasyām ] (snigdha- $\leftarrow$ dr̥ṣṭyā) ]  
on her fixed-gaze  
(sūcita- $\leftarrow$ abhilāṣaḥ)-B ]  
indicated-affection-ed  
.. whose affection was indicated by his gaze being fixed on her

Here, the past passive participle, *sūcita* (*indicated*), which is a subordinate constituent within the exocentric (*bahuvrīhi*) compound *sūcita- $\leftarrow$ abhilāṣaḥ* (*\*indicated-affectioned: whose affection was indicated*), is construed with the third, or instrumental, case noun phrase *tasyām snigdha- $\leftarrow$ dr̥ṣṭyā* (*by his gaze being fixed on her*). Moreover, this noun phrase itself exhibits a non-constituent (*a-samartha*) compound, for the past passive participle, *snigdha*

---

<sup>3</sup>To put these frequencies in perspective, I should point out that non-constituent (*a-samartha*) compounds occurred more frequently in each corpus taken separately or jointly than either indirect questions or relative clauses.

(*fixed*) is found as a subordinate constituent in the compound *snigdha-dr̥ṣṭyā* (*by fixed-gaze: by his gaze being fixed*), yet it is construed with *tasyām* (*on her*), a seventh, or locative, case noun phrase, for which the verb *snih* (*to fix*) subcategorizes.

One person to attempt to meet the challenge presented by these compounds to Pāṇini's grammar of Sanskrit was Bhartṛhari, who suggested that non-constituent (*a-samartha*) compounds are limited to cases where the subordinate constituent in the compound expresses a relation. Bhartṛhari's insight is a deep one. The remainder of this paper is devoted to showing how this insight might be captured within constituency grammars and then applied to not only non-constituent compounds but also to exocentric ones. The main concept is that of an argument frame, or an enriched subcategorization frame.

An argument frame has its classical quantificational logic. In elementary model theory, the set of predicates is partitioned into cells of the same degree, or adicity. The set of predicates are partitioned into the family of sets, one of which comprises all the one-place predicates, another all the two-place predicates, etc. This partitioning of the predicates has two effects. On the one hand, it helps to determine which string of symbols is a well-formed formula and which is not; on the other hand, it determines what kinds of values can be assigned to it. Thus, if one is told that  $P$  is a two place predicate and that  $a$ ,  $b$  and  $c$  are terms, then one knows that  $Pab$  is a formula and that neither  $Pa$  nor  $Pcba$  is. At the same time, one knows that  $P$  is to be assigned a set of ordered pairs in the model.

A similar effect can be achieved with a slight enrichment of subcategorization frames. Subcategorization frames were introduced into constituency grammar by Chomsky (1965 ch. 2.3.4; 2.4). They greatly simplified the constituency rules by having individual lexical items specify their complements. Subcategorization frames effectively formalized and generalized the lexicographical practice of distinguishing between transitive and intransitive verbs. Because the subcategorization frame of a word is silent about those constituents which are not its complements, the subcategorization frame of a verb does not specify that it has an argument corresponding to the subject of the clause in which it might occur. Argument frames will have this specification. As a result, argument frames will fulfill the same functions in constituency grammar which predicate adicity fulfills in classical quan-



tificational logic. It specifies the arguments associated with the word and it constrains the value assigned to it in a model to those relations with a corresponding arity, or degree of the relation. For example, a verb such as *to die*, which is intransitive, and thereby corresponding to a monadic predicate of classical quantification logic, takes only one argument and is assigned a unary relation (a subset of the model's domain), while a verb such as *to admire*, which is transitive, and thereby corresponding to a dyadic predicate, takes two arguments and is assigned a binary relation (a set of ordered pairs of members of the model's domain). Such a specification accounts for the contrasts in acceptability of the sentences given below.

- (7.1) Bill died.
- (7.2) \*Bill died Fred.
- (8.1) \*Mary admires.
- (8.2) Mary admires Bill.

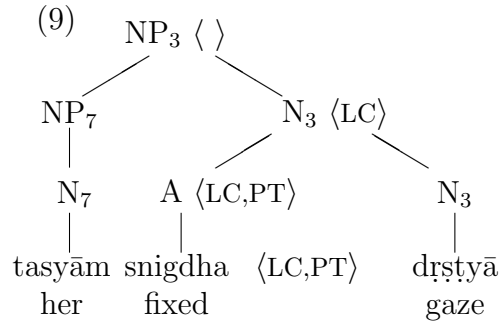
Moreover, just as each predicate of a given adicity is interpreted by a relation of a corresponding arity, so each relational word is interpreted by a relation of a corresponding arity. It is crucial that this correspondance be properly established. To see why, consider this example from model theory. Let  $R$  be a binary predicate and let  $a$  and  $b$  be individual constants. Let  $M$  be a model whose domain is  $\{1, 2, 3\}$  and whose interpretation function  $i$  assigns 1 to  $a$ , 2 to  $b$  and the set of ordered pairs  $\{\langle 1, 2 \rangle, \langle 2, 3 \rangle, \langle 3, 1 \rangle\}$  to  $R$ . The clause of the truth definition of an atomic formula guarantees the following:  $Rab$  is true if and only if  $\langle i(a), i(b) \rangle \in i(R)$ . It is crucial that the order of appearance of the individual constants  $a$  and  $b$  in the formula  $Rab$  be correlated with the ordered pair  $\langle i(a), i(b) \rangle$ , not with the ordered pair  $\langle i(b), i(a) \rangle$ . As the reader can easily verify, the ordered pair  $\langle i(a), i(b) \rangle$  is a member of  $i(R)$ , but not the ordered pair  $\langle i(b), i(a) \rangle$ .

The situation in natural language is, of course, much more complex than the situation in logical notation. To begin with, a syntactic structure of even a simple clause is much more complex than that of an atomic formula, the formation of the latter being simply a matter of concatenation. In addition, many argument positions have associated with them values, called *valences* or *thematic roles*. Indeed, these valences are at the heart of Pāṇini's grammar, the *Aṣṭādhyāyī*, where they are known as *kāraka*, or *factors*. Typical valences include those recognized by Pāṇini: *agent*, *patient*, *beneficiary*, *source* and

*location*. Valences are confined to verbs and to nouns and adjectives derived from verbs. They are not associated with the arguments of underived relational nouns (for example, *friend*, *cousin*, *neighbor*, *colleague*) or with underived relational adjectives (for example, *equivalent*, *opposite*, *proud*, *domestic* and *local*). For the sake of simplifying the discussion, I shall avail myself of the terminology of valences to index the argument positions relevant to the exposition of the examples in the treatment below.

Following Bhartrhari’s conjecture, I shall assume that non-constituent (*a-samartha*) compounds appear when the subordinate constituent in the compound has an argument which is construed with an inflected lexical item external to the compound. A survey of the cases mentioned above, as culled from the classical literature, shows that such is the case. Indeed, for the most part, the subordinate constituent is a deverbal noun or adjective, requiring an NP complement and often associating with it a particular valence.

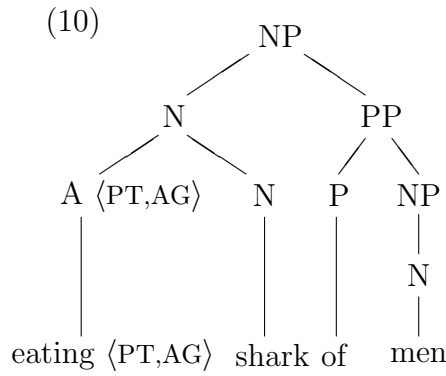
Let us consider the case of a non-constituent compound *snigdha-dṛṣṭyā*, cited above. Recall that it is preceded by the pronoun *tasyām*, which is construed with the word *snigdha*, itself subordinate to *dṛṣṭyā*. The past passive participle *snigdha* has two arguments, one of which must appear in the seventh case.



(where PT denotes *patient* and LC denotes *location*). The idea is that the argument frames are passed, as it were, up the tree. It is the *location* argument which is passed up to the top node of the tree for the compound.<sup>4</sup>

<sup>4</sup>This compound is also an exocentric compound. This aspect of the compound is not being addressed here.

This contrasts with the situation in English. Sanskrit, as we just saw, permits unsaturated arguments associated with a non-head to be transmitted to the mother node, while English prohibits non-heads from having unsaturated arguments. Thus, for example, an expression such as (3.2) is prohibited in English. The reason is that, although one of the arguments associated with *eating*, namely the one whose valence is *agent* is saturated by the noun *shark*, the other argument associated with *eating*, namely the one whose valence is *patient* is not.<sup>5</sup>



This treatment of non-constituent (*a-samartha*) compounds extends to exocentric (*bahuvrīhi*) compounds. Let us consider the following exocentric (*bahuvrīhi*) compounds in Sanskrit:

- (11.1) SK 830  
 Compound: prāptātithiḥ grāmaḥ  
 Analysis: (prāpta-←atithiḥ)-B grāmaḥ  
           reached-guest-ed village  
 Paraphrase [ <sub>RC</sub> atithayaḥ prāptāḥ yam ]  
               guest reached which  
           saḥ prāptātithiḥ grāmaḥ  
           that village  
           the village which guests have reached

<sup>5</sup>Evidence that the argument with the valence of PATIENT is relevant comes from the acceptability of (3.1), namely *man-eating shark*.

(11.2) SK 830  
 Compound: ūḍharathaḥ anaḍvān  
 Analysis: (ūḍha-ṛathaḥ)-B anaḍvān  
           drawn-cart-ed      bull  
 Paraphrase: [<sub>RC</sub> rathaḥ ūḍhaḥ yena ]  
               cart    drawn    by which  
 saḥ ūḍharathaḥ anaḍvān  
 that                              bull  
 the bull by which a cart is drawn

(11.3) SK 830  
 Compound: upahr̥tapaśuḥ puruśaḥ  
 Analysis: (upahr̥ta-ṛpaśuḥ)-B puruśaḥ  
           offered-cattle-ed      man  
 Paraphrase: [<sub>RC</sub> paśuḥ upahr̥taḥ yasmai ]  
               cattle    offered    to whom  
 saḥ upahr̥ta-paśuḥ puruśaḥ  
 that                              man  
 the man to whom cattle is offered

(11.4) SK 830  
 Compound: uddhr̥taudanā sthālī  
 Analysis: (uddhr̥ta-ṛodanā)-B sthālī  
           removed-rice      vessel  
 Paraphrase: [<sub>RC</sub> odanaḥ uddhr̥taḥ yasyāḥ ]  
               rice    removed    from which  
 sā uddhr̥taudanā sthālī  
 that                              vessel  
 the vessel from which rice has been removed

(11.5) SK 830  
Compound: pītāmbaraḥ puruṣaḥ  
Analysis: (pīta-<ambaraḥ)-B puruṣaḥ  
            yellow-garment-ed man  
Paraphrase: [<sub>RC</sub> pītam ambaram yasya ]  
            yellow garment whose  
            saḥ pītāmbaraḥ puruṣaḥ  
            that man  
            the man whose garments are yellow

In the paraphrase and translation of Sanskrit exocentric (*bahuvrīhi*) compounds, the relative pronoun of the paraphrasing relative clause may be construed with either the second constituent (11.5) or the first constituent (all other examples). Both of these constituents are subordinate within the structure of the exocentric *bahuvrīhi* compound. This is reflected in the canonical paraphrase, where the relative pronoun is construed with the subject of the relative clause (11.5) or with its predicate (all other examples). Indeed, as noted by Coulson (1976, p. 121), Sanskrit exocentric (*bahuvrīhi*) compounds are ambiguous between two readings: on one, the denotation of the lexical item modified by the exocentric compound is interpreted as the possessor of what is denoted by the final constituent of the compound; and on the other, it is interpreted as bearing a valence of any unsaturated argument associated with the initial constituent of the compound.

(12) Coulson 1976, p. 121  
Compound: dr̥ṣṭakaṣṭā strī  
Analysis: (dr̥ṣṭa-kaṣṭā)-B strī  
            witnessed-misfortune-ed woman  
Reading 1: a woman whose misfortune has been witnessed  
            (i.e., a woman whose misfortune people have witnessed)  
Reading 2: a woman by whom misfortune has been witnessed  
            (i.e., a woman who has witnessed misfortune)

Moreover, an exocentric compound has available a reading corresponding to each of the unsaturated arguments associated with its initial constituent.

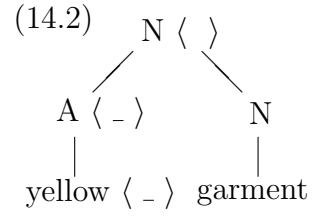
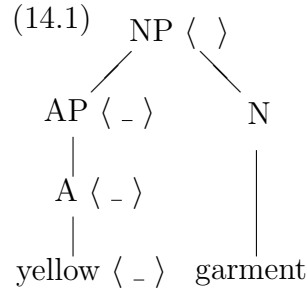
- (13) Coulson 1976, p. 121  
 Compound: dattādarā rajñī  
 Analysis: (datta-ādarā)-B rajñī  
           given-respect-ed queen  
 Reading 1: a queen by whom respect is given  
           (i.e, a respectful queen)  
 Reading 2: a queen to whom respect is given  
           (i.e., a respected queen)

Here emerges the difference between English and Sanskrit exocentric compounds alluded to above. Notice that, of the six examples, only the fifth allows an acceptable English calque: *\*reached-guested*, *\*drawn-carted*, *\*offered-cattled* and *\*removed-vesseled* but *yellow-garmented*. At the same time, while an English exocentric compound is paraphrasable with a relative clause, yet the relative pronoun of the paraphrase, “whose”, is construed only with the subject of the relative clause, which corresponds to the final constituent of the compound paraphrased. Thus, *mean-spirited* is paraphrasable as *one whose spirit is mean*, *level-headed* as *one whose head is level*, and *long-legged* as *one whose legs are long*.

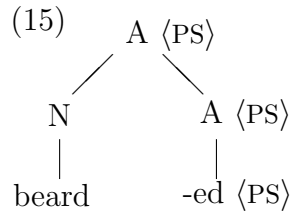
English and Sanskrit exocentric (*bahuvrīhi*) compounds differ as follows: the English adjectival suffix *-ed* does not permit the transmission of unsaturated arguments of an exocentric compound’s initial constituent; whereas the Sanskrit adjectival suffix B does permit the transmission of such arguments.

The foregoing differences between compounds in English and Sanskrit suggests the following hypothesis: the argument frame of initial constituents in lexical structure, in particular, in compound, percolate in Sanskrit but does not in English. This hypothesis accounts for two facts: first, that, in Sanskrit, unsaturated arguments associated with the initial constituent of an exocentric compound can be assigned to the lexical item the compound modifies, whereas in English they cannot be; second, that Sanskrit productively forms non-constituent (*a-samartha*) compounds whereas English does not. Let us see how this account works.

Each adjective has at least one argument which is saturated either by the noun it modifies or by the subject noun phrase of which it is predicated. This is illustrated below, for modification both within phrasal structure and within compound structure.

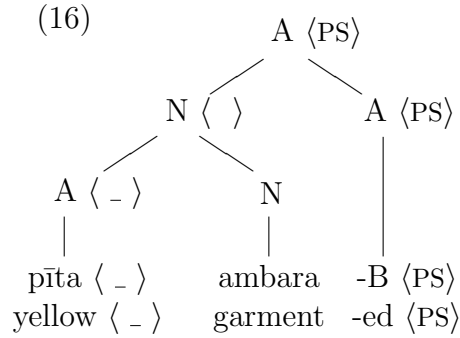


Now, both the *-ed* suffix in English and the *-B* suffix in Sanskrit create adjectives from nouns. This means that they create an argument. Associated with the resulting argument is the valence *possessor* (annotated PS). When the English suffix is applied to a simple noun like *beard*, one obtains the following:



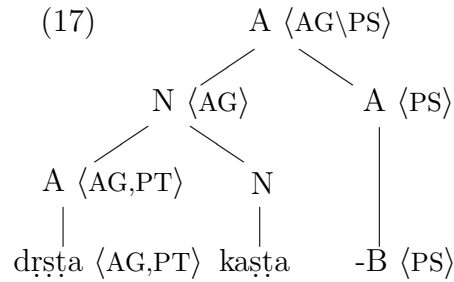
And when the resulting form modifies a word such as *man*, the resulting interpretation is *man who possesses a beard*. Combining what has been said so far, one obtains an analysis for both the Sanskrit compound in (11.5) and its English claque translation.

Moreover, the foregoing analysis shows precisely where Sanskrit and English differ. A morphologically complex English word accepts unsaturated arguments associated only with its head. Whereas, a morphologically complex Sanskrit word accepts the unsaturated arguments either of its head or of its head's sister. When an exocentric compound has no unsaturated argument other than the one associated with its possessive suffix, then its English and Sanskrit versions are equally acceptable.



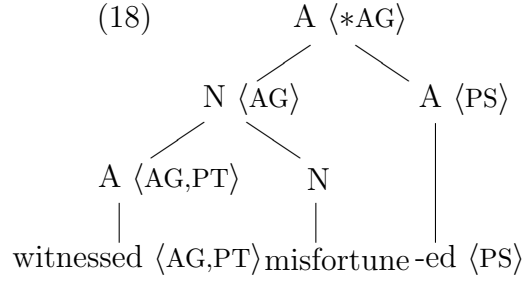
In this example, the argument associated with *yellow* (*pīta*) is saturated by *garment* (*ambara*), and so the complex word *yellow garment* (*pītāmbara*) has no unsaturated argument. The suffixation of *-ed* (*-B*) to *yellow garment* (*pītāmbara*) creates an unsaturated argument with an associated valence, namely that of possessor (PS).

The situation is otherwise when the left-hand constituent of an exocentric compound has an unsaturated argument. Sanskrit permits unsaturated arguments associated with either a head and a non-head to be transmitted to the mother node; and, depending on which unsaturated argument is transmitted, the compound receives one or another interpretation. Thus, in the compound in (12) the unsaturated argument associated with the entire compound may have associated with it either the value AG or the value PS (annotated below as  $\langle AG \setminus PS \rangle$ ).



In contrast, English prohibits any unsaturated arguments from being associated with a non-head, with the consequence that the English counterparts to (9) are ungrammatical (annotated below as  $\langle * \rangle$ ).





### 3 Conclusion

Above, we examined two kinds of compounds in Classical Sanskrit, non-constituent (*a-samartha*) compounds and exocentric (*bahuvrīhi*) compounds. The former compounds were considered problematic by the Indian grammatical tradition for Pāṇini's grammar, the *Aṣṭādhyāyī*. An insight due to Bharṭṛhari shows how they can be satisfactorily analyzed. This insight was recast using the notion of an argument frame, a generalization of subcategorization frame. A bonus of this solution is that it provides insight into well-known properties of the exocentric compounds of Classical Sanskrit, properties which exocentric compounds in English do not have.

#### ACKNOWLEDGEMENTS:

I thank Jan Houben and Peter Scharf for bringing to my attention errors in an earlier version of this paper.

## WORKS CITED OR CONSULTED

- Aklujkar, Ashok 1992 *Sanskrit: An Easy Introduction to an Enchanting Language*. Richmond, British Columbia (Canada): Svādhyāya Publications.
- Apte, Vāman Shivarām 1885 *The Student's Guide to Sanskrit Composition. A Treatise on Sanskrit Syntax for Use of Schools and Colleges*. Poona, India: Lokasamgraha Press, 24th edition (1960).
- Apte, Vāman Shivarām 1890 *The Practical Sanskrit-English Dictionary*. Poona, India: Prasad Prakashan, revised and enlarged edition (1957).
- Bhaṭṭoji Dīkṣita *Siddhānta-Kaumudī*. Sanskrit Edition and English Translation: Vasu, Śrīśa Chandra (ed) (tr) 1906.
- Bloomfield, Leonard 1933 *Language*. New York, New York: Holt.
- Boose, Emery and Gary Tubb 1981 *Rough Draft of Portions of a Handbook Designed to Aid Students in the Use of Sanskrit Commentaries*. Unpublished mss., Harvard University.
- Cardona, George 1988 *Pāṇini: His Work and Its Traditions. Background and Introduction*. New Delhi: Motilal Banarsidass.
- Chomsky, Noam 1957 *Syntactic Structures*. The Hague, The Netherlands: Mouton and Company (Janua Linguarum: Series Minor n. 4).
- Chomsky, Noam 1965 *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The MIT Press.
- Coulson, Michael 1976 *Sanskrit: An Introduction to the Classical Language*. London, England: Hodder and Stoughton (Teach Yourself Books).
- Deshpande, Madhav M 1985 *Ellipsis and Syntactic Overlapping: Current Issues in Pāṇinian Syntactic Theory*. Poona, India: Bhandarkar Oriental Research Institute (Post-graduate and Research Department Series no. 24).
- Di Sciullo, Anna Maria and Edwin Williams 1987 *On the Definition of Word*. Cambridge, Massachusetts: The MIT Press.
- Gillon, Brendan S. 1993 Bharṭṛhari's solution to the problem of *asamartha* compounds. *Asiatische Studien/Études Asiatiques*: v. 47, n. 1, pp. 117–133.
- Gillon, Brendan S. 1995 Autonomy of word formation: evidence from Classical Sanskrit. *Indian Linguistics*: v. 56, n. 1-4, pp. 15–52.

- Godabole, N. B. (ed) 1933 *Śakuntalā*. Bombay: Nirṇaya Sāgara Press (revised by W. L. Paṇṣīkar)
- Harris, Zellig 1946 'From Morpheme to Utterance'. *Language*: v. 22, pp. 161-183.
- Iyer, K. A. Subramania (ed) 1973 *Vākyapadīya of Bhartṛhari with the Prakīrṇakaparakāśa of Helarāja*. Poona: Deccan College.
- Iyyar, Seshādri (ed) 1896 *Mālavikāgnimitra*. Poona, India: publisher unknown.
- Kielhorn, F. (ed) 1880 *The Vyākaraṇa Mahābhāṣya of Patañjali*. Poona, India: Bhandarkar Oriental Research Institute (4<sup>th</sup> edition revised by R. N. Dandekar 1985).
- Kiparsky, Paul 1983 *Word Formation and the Lexicon*. In: Ingemann, Frances (ed) 1983 ??-??.
- Malvania, Dalsukhabhai (ed) 1955 *Dūrveka Miśra's Dharmottarapradīpa*. Patna, India: Kashiprasad Jayaswal Research Institute (Tibetan Sanskrit Works Series: v. 2). 2<sup>nd</sup> edition revised, 1971.
- Manaster-Ramer, Alexis and Kac, Michael B. 1990 *Linguistics and Philosophy*: v. 13, pp. 325-362.
- Marchand, Hans 1969 *The Categories and Types of Present-Day English Word-Formation. A Synchronic-Diachronic Approach*. Munich: C. H. Beck'sche Verlagsbuchhandlung. (reprint)
- Monier-Williams, Monier 1899 *A Sanskrit-English Dictionary*. Oxford, England: Oxford University Press.
- Selkirk, Elizabeth O. 1982. *The Syntax of Words*. Cambridge, Massachusetts: The MIT Press.
- Speijer, J. S. 1886 *Sanskrit Syntax*. Leiden, The Netherlands: E. J. Brill.
- Vasu, Śrīśa Chandra (ed) (tr) 1906 *The Siddhānta Kaumudī of Bhaṭṭoji Dikṣita*. Allahabad, India: The Pāṇini Office. Reprint: Delhi, India: Motilal Banarsidass, 1962.
- Wells, Rulon S. 1947 'Immediate Constituents' *Language*: v. 23, pp. 81-117.
- Whitney, William Dwight 1881 *Sanskrit Grammar: Including both the Classical language, and the older Dialects, of Veda and Brahmana*. Cambridge, Massachusetts: Harvard University Press, 2nd edition (1889), 11th reprint

(1967).

Williams, Edwin 1981 On the Notions “Lexically Related” and “Head of a Word”. *Linguistic Inquiry* 12, 245-274.