Journal of South Asian Linguistics

Contents

Editorial statement

The editorial team is thrilled to present volume 12 of the *Journal of South Asian Linguistics*, bringing together three new articles published in 2022, as well as an erratum from a previously published article. This volume is particularly notable in that it presents accounts of syntactic and phonetic phenomena in two languages of South Asia that despite their ubiquity may sometimes escape analytical attention: Sanskrit and Indian English. The diversity of frameworks and approaches explored in this volume signal the rich nature of research in languages of the region and the robust ongoing contribution of linguistics of South Asian languages to the discipline at large.

The first two articles in this volume were initially conceived as linked efforts in a wider project assessing the effectiveness of the use of ellipsis in the analysis of Sanskrit *bahuvrīhis* (an exocentric compound construction). Portions of the research were presented at discussed at the South Asian Language Association 35 Conference (INALCO, Paris) in October 2019. The first paper, entitled "Dispensing with zero in the analysis of Sanskrit bahuvrīhi: Resurfacing, testing and assessing Pāṇini's model", by Maria Piera Candotti and Tiziana Pontillo, opens with an intriguing puzzle. Recent approaches to Sanskrit *bahuvrīhis* have assumed that the derivation of the compound includes a zero (covert) suffix (Kiparsky 1982; Gillon 2008). Despite Pāṇini's own extensive use of empty categories and other "zero devices" to describe and explain syntactic phenomena in Sanskrit, he does not propose such an account of *bahuvrīhi*. Condotti and Pontillo examine shortcomings of existing zero-suffix accounts and explore the advantages of adopting Pāṇini's original rule for deriving these compounds. In so doing, they contribute to a wider understanding of compounding generally, which they point out sits at an important nexus of syntax and morphology.

The second article in this pair focusing on *bahuvrīhi* is entitled "The possessive relation in Sanskrit bahuvrīhi compounds" by Davide Mocci. The paper focuses on the syntax of a particular type of *bahuvrīhi* which encodes a possessive relation in which one of member of the compound is the possessum and an expression not mentioned within the *bahuvrīhi* is the corresponding possessor. Mocci considers two approaches to explain the absent possessor: the first in which the possessor is elided, and the second in which the possessor undergoes movement outside of the constituent containing the compound. Ultimately, he advances a formal derivation of *bahuvrīhis* based on the movement strategy, and shows that restrictions on *bahuvrīhis'* internal order are concordant with locality conditions on movement.

The third paper systematically examines the long-held auditory impression that Indians apparently use unaspirated voiceless stops rather than their aspirated counterparts when speaking English, in the same word positions where English speakers in the US, the UK, and elsewhere have predictable aspiration. The authors Jahnavi Narkar and Peter Staroverov question the initial assumption, by comparing the voice onset time (VOT) of the allophonically-aspirated stops of British English to the phonemically-aspirated stops of various Indic languages, finding that the VOT of British English aspiration is in fact not a perfect match with either that of the aspirated or the unaspirated stops of Indic languages. Furthermore, by looking at the production of a wide range of speakers in an online corpus of short recorded passages, the authors find that Indians indeed produce their English voiceless stops with a VOT comparable to that of the unaspirated stops of their L1s, which ranged across several Indo-European languages of India.

In addition to their theoretical contributions, these three papers in volume 12 also demonstrate distinct approaches to pre-existing data, a methodology that is in no way new but is presumably experiencing a surge in popularity as researchers have looked at ways to investigate important linguistic questions while respecting the needs of social distancing and decreased travel.

In closing, we would like to thank our editorial team colleague, Mythili Menon (Wichita State University), our senior advisors Miriam Butt (University of Konstanz) and Rajesh Bhatt (University of Massachusetts at Amherst), Sebastian Sulger (University of Konstanz) for setting up our website, and Sebastian Danisch (KIM at the University of Konstanz) for maintaining it. Special thanks also are due to the anonymous reviewers, without whose careful and constructive feedback our work would be impossible.

Emily Manetta (University of Vermont) & Sameer ud Dowla Khan (Reed College)

Dispensing with zero in the analysis of Sanskrit bahuvrīhi: Resurfacing, testing and assessing Pāṇini's model[*]

MARIA PIERA CANDOTTI, *UNIVERSITY OF PISA*
TIZIANA PONTILLO, *UNIVERSITY OF CAGLIARI*

ABSTRACT

   In modern linguistics it is common to analyse bahuvrīhis as derived from an endocentric compound to which a zero suffix applies (Whitney 1889:501-502; Kiparsky 1982a:139; Gillon 2008:2-3). All these descriptions owe something to the systematization handed down to us by the later pāṇinian tradition. By contrast, the present contribution highlights how, despite his extensive use of zero devices, Pāṇini himself does not adopt any of them to explain bahuvrīhis. This study attempts to recover Pāṇini's original handling of compound analysis, namely the fact that he does not focus on the head, but rather on the so-called *upasarjana* constituent, characterized in the source-phrase by a frozen case ending expressing the syntactic relation with another constituent of the compound. A frozen syntactic relation is furthermore established between one of the constituents and the denotatum of the whole compound, and is reflected in the case ending of the pronoun used (in the traditional analysis) to signify this relation. It is exclusively the syntactic meaning conveyed by this case ending that is assumed to explain the final meaning of the bahuvrīhi. Such an analysis scales back the importance of the endo- vs. exocentric polarity in the classification of typologies in compounding, in line with some quite recent achievement of contemporary linguistics (Scalise-Bisetto 2009: 45).

1.   Pāṇini's bahuvrīhi: what was the explanatory model?
The present inquiry stems from a simple question: given that Pāṇini makes extensive use of zero postulation in derivation and in phono-syntax, why does he not use it in the case of the so-called exocentric or bahuvrihi compounds as in (7a) and (7b)? Nevertheless, before illustrating Pāṇini's model to account for exocentric compounds, we should first make a preliminary reflection on what data we can assume that Pāṇini knew (and, more cogently, what data he thought he should account for) and also on the aim that made such an account necessary.

1.1   Aims and linguistic target of Pāṇini's grammar
In the present work, we quote Vedic examples selected from the *Ṛgveda*-Saṃhitā, a choice that, if not exhaustive, nevertheless grants us relevant data. The range of Vedic literature which Pāṇini might have been acquainted with and above all might have taken into account for his description, has been the subject of lengthy debate.[1] However, no one has ever questioned the fact that Pāṇini was acquainted with all the Saṃhitās and with the Black Yajurveda Brāhmaṇa tradition. There are over two hundred *Aṣṭādhyāyī* rules dealing with Vedic special features, explicitly referring to *chandas* (and *nigama*), i.e. plausibly to the four Vedic Saṃhitās, *yajus*, i.e. the sacrificial formulas in prose, *brāhmaṇa*, i.e. the homonymous class of works, *mantra*, i.e. Vedic stanzas and Vedic formulas as a whole. Some other rules are explicitly restricted to the common language (*bhāṣā*), while the bulk of the non-restricted rules, *de facto* apply to both, as convincingly shown by Deshpande (1985).[2] A crucial feature of Pāṇini's Vedic rules is that, despite their restricted domain of

---

[1] See, among the others, Goldstücker (1861: 130-149), Liebich (1891: 17–37) Thieme (1935: 74-76) and Kiparsky (2012).

[2] Several specific inquiries on single aspects of this grammar were focused on verifying its reliability. For instance, Cardona (2002: 236) states that "the tense system which Pāṇini describes for the speech of his time and area is actually attested in Vedic literature", despite what some other scholars have maintained. See also Cardona (1999:216).

application, they are, on the other hand, often 'generalizations rather than restrictions of Classical rules' and 'typically involve listing additional environments where a rule applies in Vedic' (Kiparsky 1979:57). We believe that a convincing explanation of this feature was supplied by Bronkhorst in 1981 and by Kiparsky 2012, when they maintain that Pāṇini's grammar might have assisted the editorial activity targeted on consolidating and normalizing the *Ṛgveda*. In fact, such a role played in the editorial activity on the Vedic Canon, can properly explain the need to make 'judgments of relative grammatical acceptability', like those emerging in the variationist pattern discovered in Pāṇini's architecture by Kiparsky in 1979.

This has also convinced us that Pāṇini's model, rather than being defined as generative, is better described as distributional and variational, following, again, Kiparsky's ground-breaking work. Pāṇini's architecture of the word-formation-rules is based on a variationist pattern, one of the most evident applications being that denominal derivational nominal stems (*taddhita*s) and compounds are taught as alternants of meaning equivalent phrases (see below § 4). Furthermore, a number of regional variations of Sanskrit usage (see Deshpande 2019:11;16) and several other variants (which could today be labelled as diatopic, diastratic, diaphasic and diamesic variants) of several linguistic phenomena are also recorded and explained within a distributional morphology framework. We are therefore well informed by Pāṇini himself about the boundaries of these variants. Instead, despite several intriguing hypotheses, according to which for instance the main object of Pāṇini's grammar could be a "frontier" language (Deshpande 1983; 2019; Hock 2012), a language spoken in a context of diglossia (Houben 2018), or a 'semi-colloquial language of scholarly community' (Kulikov 2013), we cannot indeed guess which kind of language is described by the non-restricted rules. This is why it could be quite risky to try and compare Pāṇini's rules with Classical Sanskrit. On the contrary, Pāṇini's rules must intentionally be equal at least to the task of accounting for the Ṛgvedic linguistic forms we know.

1.2.   Ellipsis and zero in Pāṇini

Endorsing Deshpande's (1989:103) working definition, we are using "ellipsis" as a cover term under which – in the present framework at least – all the linguistic phenomena which involve 'some sort of "incompleteness" of the surface structure or expression' should quite rightly be encompassed. As is well known, both "natural" and "regulated" ellipsis play a great role in Pāṇinian descriptions of language. By natural and regulated ellipsis we refer on the one hand to ellipsis phenomena that are assumed as a fact but never obtain any specific provision in Pāṇini's grammar and, on the other, to phenomena explicitly governed by rules. The two types of ellipsis on the whole match a distinction between syntactic ellipsis (assumed but not regulated by Pāṇini, as clearly pointed out by Kiparsky 1982c: 37) and morphological and phonic ellipsis.

A telling example of what is meant by Kiparsky may be found in rule A 1.4.105 teaching the occurrence of first and second verbal endings in co-occurrence with the pronouns *yuṣmad* 'you' and *asmad* 'we' with the specification *sthāniny api* 'even as mere place-holder [as co-occurring word]'.[3]

(1)   *pacasi/tvam pacasi* '(You) cook'

No rule in Pāṇini accounts for these two alternants. Nevertheless, the terminology used, in particular the mention of the technical term *sthānin* 'place-holder', puts this phenomenon against the background of the broader and technically determined concept of zero as a 'substitute' for something that is elsewhere overtly realized. But also in morphological derivation there is room for non-regulated or loosely regulated phenomena. An interesting example comes from Pāṇini's device of using the meaning (*artha*) of a given constituent as a relevant morphological context, thus avoiding a biunivocal link between a given meaning and a specific and overt form. In some rules, the reference to the meaning is a means of widening the set of morphological constituents quoted.[4] But the same

---

[3] A 1.4.105: *yuṣmady upapade samānādhikaraṇe sthāniny api madhyamaḥ* 'The second person verbal triplets are co-referential with a co-occurring second person (*yuṣmad*) personal pronoun even if merely understood (*sthāniny api* lit. "even when it is a place-holder")', i.e. even when it is zero-replaced. Other interesting examples of syntactic ellipsis may be found in the above-mentioned contribution by Kiparsky.

[4] See, e.g. A 1.4.25: 'The actant called *apadāna* is used to signify the cause in verbal bases meaning "to fear" or "to protect".': e.g. *caurebhyo bibheti/caurebhya udvijate* 'he is scared of robbers'.

device may on the contrary hint at the absence of the specific morphological constituent conveying that meaning:

> A 2.3.15: The dative is also used after a nominal base expressing a condition (*bhāva-va-canāt*) and having the meaning as the suffix *-tum* (*tum-arthāt*).

The above-mentioned suffix *-tum* is taught by A 3.3.10 and accounts for the infinitive competing forms such as (2b).

(2) a. *pākāya*   *vrajati*   → *pākāya vrajati*
     act of cooking-DAT. SG   to go-PRS. 3 SG   'he goes to cook'
     b. *paktum*   *vrajati*   → *paktum vrajati*
     to cook-INF.   to go-PRS. 3 SG   'he goes to cook'

In the nominal derivate *pāka-* 'act of cooking' there is no morphological constituent conveying the meaning of the infinitive: the pure meaning of the constituent elsewhere used may in these cases act as a left-hand constituent constraint: the dative is used after an action noun provided this action noun has the meaning of an infinite, i.e. co-occurs with a finite verb as its object.[5] The latter device is of particular relevance for us, as the most important rule regulating bahuvrīhis is formulated precisely with the constraint *anyapadārthe* 'in the meaning of another word'. This will be discussed below, §§ 5-6. Alongside these examples of assumed yet non-regulated ellipsis, on the other hand, several morphological and phonic phenomena are analysed by Pāṇini by means of the notion of "zero" (*lopa*, defined as *adarśana* 'non-perception' in A 1.1.60), which indeed – and this is a crucial point, even though it is far from universally accepted in indological and in linguistic studies in general – is exclusively a specific kind of "substitution" (*ādeśa*, lit. 'special teaching').[6] Such a "zero" is essentially targeted (as we have extensively discussed and hopefully proven elsewhere)[7] on identifying the place where a given word-form is expected, because it is overtly perceived in analogous linguistic contexts. By the way, this is not exclusively proper to zero, rather it is a feature of the whole mechanism of substitution as a device to account for context-specific alternants.

It is well-known, in fact, that the descriptive method adopted by Pāṇini is consistently synchronic, with almost no concession to the diachronic dimension of language nor consequently to deletion. Bloomfield (1933: 209) himself (who was one of the first modern linguists who shed light upon 'the apparently artificial but in practice eminently serviceable device of speaking of a zero element' on the part of Indian grammarians) explained his English examples just in terms of "replacement": 'in *sheep* : *sheep* the plural-suffix is replaced by zero – that is by nothing at all'. Indeed structuralists generally recognize their debt to Pāṇini whose method they properly connect to the requirements of generalization in morphology.[8] Yet modern linguistic studies rarely take into account the overall framework, despite the fact that they often refer to some specific aspects or, in any case, make use of the relevant terminology.[9]

Referring to our previous contributions for a comprehensive discussion of the data on Pāṇinian zero, we will focus here on what is strictly relevant both for interpreting Pāṇini's approach to these phenomena and for highlighting its possible differences with other theories. Firstly, it is important to describe the breadth of linguistic phenomena dealt with by Pāṇini by means of this device that

---

[5] This example allows us to highlight a feature that will be important for the next steps of our inquiry, namely the fact that the status of primary derivative stems is in-between verbal and nominal categories. This will prove fruitful in assessing the problems of unsaturated arguments in compounds (see above, fn. 20 and below § 6).

[6] See e.g. the translations ot the relevant Pāṇini *lopa* rules by Böhtlingk 1887 ('Schwund') Vasu 1891 ('elision'), Renou 1947-1954 ('amuissement'). Sharma (1990-2003) does not translate the term sometimes uses the expression 'zeroing' to describe the operation. See also Renou 1942, s.v. 'amuissement, chute'); Varma 1978 ('elision'); Dvivedi 1978:77 ('elision'); Das 1985 ('elision'); Sen 1999:102 ('elision'); Roodbergen 2008, s.v. ('deletion'), but Katre (1987:24) and Cardona (1997:46) interpret *lopa* as 'zero-replacement'.

[7] See Candotti-Pontillo (2013: 122-125); Candotti-Pontillo (2014: 64-68).

[8] See e.g. Allen 1955:113; Pinault 1989:323; Zakharyin 2007:338. As regards some modern linguists who might have been aware of (and influenced by) such a device used by Pāṇini, see also Pontillo (2002:559-570).

[9] e.g. in Allen 1955:109; Collinder 1962-68:15 only discuss examples of type (3) c; Meier (1961:140-143) on the other hand merely refers to the zero of single phonemes, here exemplified in (3a).

encompasses substitutions taught in phonic or almost phonic terms, such as (3a) and (3b) as well as substitutions with zero in place of affixes, whether declensional (3c) or derivational (3d):

(3)   a. **anañj-ti-* → *anakti* 'he anoints' (with Ø in the place of the nasal phoneme due to the specific present verbal class A 6.4.23)
       b. *sarit-s NOM.SG → *sarit* 'river' (with Ø in the place of the NOM.SG monoconsonant ending –*s* A 6.1.68)
       c. *vyoman-i LOC.SG → Vedic *vyoman* 'in heaven' (with Ø in the place of LOC.SG ending –*i* A 7.1.39),
       d. *bhās-* 'to shine' (vb. base) → *bhās* 'light' lit. 'the shining one' (with Ø in the place of the agentive abstract morpheme *KVIP* A 3.2.177).

With one single exception, namely rule A 5.3.82, Pāṇini does not teach zero-substitution of full words, even though early commentators already use this device and apply it in particular in the case of compounds.[10] Pāṇini further organizes the vast domain of zero in place of affixes through a broad division between
    a) zero of the *lopa*-type, characterized by the fact that the zero affix can trigger a specific ablaut on the base, as if it were the alternant overt affix (4)
    b) *LU*-type zero (with three subtypes: *LUK, ŚLU* and *LUP*), whose affixes do not trigger ablaut modifications of the base, but might show syntactic effects. Particularly interesting for us is the opposition between *LUK* and *LUP* exemplified in (6). The starting point is the comparison between two derivatives, a more general one (5a) with the affix –*aÑ*, meaning the modification/part of (a plant) and the second, -*aṆ* (5b), explicitly restricted to naming the fruit of the plant. The *LUK*-zero-derivation *de facto* neutralizes the feminine affix of the base by assuring a generic nominal stem, suited to the application of the derivational affix, while the *LUP*-one maintains it, crucially freezing the nominal stem proper of the etymon. In both cases, the zero forms alternate (with or without syntactic and semantic restrictions) with overt forms, i.e. they precisely work as "zero-alternants" according to Bloomfield's (1933: 215-216) terminology, as we can see in examples (5) and (6).

(4)   *a-* + *bi-* + *bhṛ-* 'bring' + Ø in the place of -*t* or –*s* → *abibhar* 'you/he brought'. The alternants -*s* and -*t* as taught by A 7.3.84 trigger a specific ablaut (the so-called *guṇa*-ablaut of the verbal base).

(5)   a. *jambū́-*   +   *aÑ*   → *jā́mbava* A 4.3.139[11]
          black plum tree F   modification of/part of   'produced from/part of the black plum tree'
       b. *jambū́-*   + optional *aṆ*   → *jāmbavá* A. 4.3.165
          black plum tree F   modification of/part of   'black plum fruit'

(6)   a. *jambū́-*   *LUK* of *aÑ* → ø   *jambú* A 4.3.163
          black plum tree F   modification of/part of   'black plum fruit'
       b. *jambū́-*   optional *LUP* of *aṆ* → ø   *jambū́* A 4.3.166
          black plum tree F   modification of/part of   'black plum fruit'

Both zero phenomena are thus to be interpreted in a wider substitution frame that links the zero affixes to their overt forms, seen as realized elsewhere. In the background of postulating a zero-affix in (6a) *jambú* and (6b) *jambū́*, we have the matching overt forms *pálāśa-*, *jā́mbava-* (with *aÑ*) – in the sense of a part of/modification of the relevant tree – and *pālāśá- jāmbavá-* (with *aṆ*) already meaning the fruit.[12] In (6a) the zero is justified by syntax, semantics and prosody while (6b) might seem to be based only on semantics. Nevertheless, the derivation through zero in the place of *jambū* 'black plum fruit' from *jambū* 'black plum tree' may be equated only superficially to the grammaticalisation of a metaphoric/metonymic transfer, since it enters a network of realisation with different levels of grammatical soundness: (6a) *jambú* is accepted, but preferably (5b) *jāmbavá* and (6b)

---

[10] Vt. 12 as A 2.2.24: *saptamy-upamāna-pūrvapadasyottarapadalopaś ca* 'And there is zero-substitution of the second constituent in [a bahuvrīhi] whose first constituent is either inflected in the locative case or a standard of comparison.' – ex. *kaṇṭhekāla-* = *kaṇṭhe-* [+ Ø in the place of *sthaḥ*] *kālo 'sya* = 'black colour persists on his throat'; *e.g., uṣṭramukha-* = *uṣṭrasya* + [Ø in the place of *mukham*] *iva mukham asya* 'his face is similar to a camel-face'.

[11] The final form is reconstructed but not attested.

[12] The zero being of the *LU*-type, there is no ablaut of the base.

jambū́ are used. It seems quite clear that, through a compelling Pāṇinian version of an 'elsewhere principle', not all metonymic transfers may be accounted for in the Pāṇinian system by postulating the zero in the place of some sort of affix.

3. Zero and exocentricity in modern linguistics

Exocentric compounds of the kind (7) are in fact a prototypical example of formations that, to use Kastovsky's (1982:152) perspective, deviate from the natural morphology of motivated words founded on the identification-specification scheme relation, which may be a variant of Deshpande's above-mentioned perception of 'some sort of "incompleteness" of the surface structure or expression'.

(7) a. *archád-* + *dhūmá-* → *archáddhūma-* (ṚV 10.46.7)
     gleam-PRS.PTCP   smoke-M.SG   'having gleaming smoke'-ADJ;
     b. *á-mita-* + *krátu-* → *ámitakratu-* (ṚV 1.102.6)
     non-NEG-measure-PST.PTCP strength-M.SG 'having immense strength'-ADJ

This perception goes back as far as the first modern descriptions of Sanskrit, beginning with Bopp's (1827 [1861: 455]) assumption that in bahuvrīhi compounds a specific notion is to be integrated in compound analysis.[13] Almost a hundred years later, Wackernagel ([1905] 1957:288) still mentions the presence of a third notion as even a pre-condition for the formation of the most ancient compounds themselves. Besides this purely semantic level there are, albeit less studied, morphological traces of the specificities of exocentric compounding. A major landmark is Kiparsky's 1982b contribution focused on the concept of (3)-level-ordered morphology. In this framework Kiparsky identifies two different morphological layers for endocentric and exocentric compounds, the latter showing the traces of a lower morphological level:

> Consider e.g., *milk teeth* (endocentric) vs. *sabertooths* 'sabertooth tigers' (exocentric). [...] endocentric noun compounds are formed at level 2 by combining words, including words derived at level 1 such as *teeth*. Exocentric compounds however, must on our assumptions be assigned zero derivational suffixes since they would otherwise share the properties of their heads, i.e. be endocentric. But [...] derivational affixes cannot be added to derived plurals. Therefore exocentric compounds come out of level two with exclusively singular morphology and can receive plural endings only at level 3 where they are adjoined to the whole compound. (Kiparsky 1982b: 10-11)

Kiparsky's model thus creates a very close link between exocentricity, zero-affixation, a lower morphological layer and accent. In particular the first two are mutually dependent: to be exocentric means to have a zeroed head. Nevertheless, as endocentric and exocentric compounds follow a completely different pattern of formation, zero-affixation, as proposed by Kiparsky, is not to be interpreted in terms of a conversion from a pre-existing endocentric compound.

3.1 Zero derivation patterns in exocentric compounds

Significantly, zero either of a suffix or of a word has often been posited both by traditional and by modern scholarship to account for this discrepancy between form and meaning. In fact, one may wonder whether Pāṇini's descriptive strategy has some interesting clue to give on how to deal with these specific compounds and with the device of ellipsis itself, its power and its limits. Concerning the postulation of zero in the analysis of bahuvrīhi compounds in modern scholarship, two basic

---

[13] The "incompleteness" of constituent analysis in bahuvrīhi compounds was pointed out by the first modern grammars and interpreted, besides Bopp's 'possesive Komposita', either as 'relative Komposita/relative compounds' (see Monier Williams 1857:329-330): 'for the obvious reason of their being *relatively* and not absolutely employed [...] translated into English by the aid of a relative pronoun') – or as attributive compounds (see F. Kielhorn 1870:250): '[A bahuvrīhi] denotes something else than what is expressed by its members [and] has the nature of an adjective.'). For ancient criticism of these analyses, see § 6.

procedures may be identified, resorting to either the postulation of a zero-suffix[14] or the zeroing of a lexeme.[15]

The first model is the most widespread and seems to account elegantly for the productivity of possessive bahuvrīhi compounds and for their attributive function. It is also widely used to account for Sanskrit bahuvrīhis, in particular of Classical Sanskrit, where e.g. Gillon (2008: 2) considers as a *natural hypothesis* that an exocentric compound is derived from a descriptive compound suffixed with a phonetically null suffix which converts it into an adjective.

A major argument often used to reinforce this interpretation in the specific case of Sanskrit bahuvrīhis is the fact that a parallel set of exocentric compounds with overt suffix is found in Sanskrit and, at least as far as the suffix *-ka* is concerned, is freely productive in Classical Sanskrit (see e.g., Gillon 2008:4). We will discuss this point in greater detail below but it suffices to point out, for the moment, that suffixation is originally very rare, almost excluded in Ṛgvedic compounds,[16] and most of the time lexically restricted. Moreover, suffixation makes it difficult to analyse compounding in syntactic terms, thus relegating to the margins some well-attested compounds such as the tradition-ally-called *a-samartha* compounds where an internal constituent may have a syntactic argument outside the compound itself. Gillon (2008:13) gives a tentative answer to this difficulty by postulat-ing a special (zeroed) B suffix that does permit the transmission of unsaturated arguments of the initial constituent of a bahuvrīhi. Such a suffix is language-specific, as it is contrasted by Gillon himself with the matching *-ed* English suffix that allows the transmission of unsaturated arguments associated with *-ed* itself (*yellow garment-ed*) but not with the initial constituent of the compound (**drawn-cart-ed*). This accounts for occurrences such as the classic example *ūḍharatha-* "by which a cart is drawn". See below (22).

While we maintain that the original Pāṇinian model did not follow this zero-model, we are aware that it is the principal model against which it competes, not only in its modern variants but also in post-Pāṇinian traditional scholarship. Moreover, it may also be that the zero-model better accounts for later Sanskrit compounding (from which both Gillon and Lowe seem to draw their data) but, on the other hand, we contend that Pāṇini's description is better suited for Vedic data.

The second model, on the other hand, which freezes the reference of the exocentric compound to a specific zeroed denotatum, fails to account for some crucial features, such as its being a qualifier that can refer to more than one qualificand – e.g. Skt. *bahuvrīhi-* 'much-rice-ed' may qualify a man 'wealthy in rice' but also e.g., a field 'yielding much rice'.

3.2 Alternative paths to account for exocentricity

Only recently other lexical-oriented approaches have tried alternative paths to zero-suffixation. For instance, Lowe (2015) concentrates on a specific syntax characterising compounding in Sanskrit, a syntax sharing only some of the features of ordinary syntax, but a syntax – at any rate – dealing with full words.[17] Lowe's two crucial points concerning the analysis of Classical Sanskrit bahuvrīhis are, on the one hand, independent evidence for the existence of words that do not project phrases. This amounts to saying that some words, albeit words, cannot act as heads and cannot enter in syntactic relation with other words in the sentence. Of course, the distinctions between word and morpheme, and between word and phrase is consequently less sharp (Lowe 2015:91). On the other hand, in order to account for the passage from a projecting to a non-projecting category of the last

---

[14] Marchand (1967 [1974]:335; 1969 [1960]:14) was 'tempted to argue that the determinatum of a bahuvrīhi compound is zero'. The analysis (followed by many other authors) is of the type [[bird+brain]Ø-ed] : [[pig+tail]ed]

[15] Štekauer (1998:148-149) resorts to the 'formation of an auxiliary-complete syntagm'. The example of analysis he proposes is the abusive term [red+skin+man], in which he singles out an operation of "shortening" of word forms, consisting in cancelling the head of compounds. Nevertheless, the term is known to be also used to denote a variety of potatoes, similarly the prototypical.

[16] In the ṚV, only one bahuvrīhi has a *-ka* suffix, *tryàmbaka-* 'with three mothers' (ṚV 7.59.2); see also Melazzo (2010:41).

[17] Lowe (2015:71) stresses the role of the so-called *asamartha* compounds in calling for a syntactic analysis of compounds: '[...] it should be impossible for syntactic relations to hold between subparts of morphologically formed compounds and words external to the compound. The fact that in Classical Sanskrit such relations are relatively common provides strong evidence for the syntactic status of Sanskrit compounding processes [...].'

constituent[18] of the compound, he uses the theory of *lexical sharing* which assumes that certain lexical items can instantiate two nodes in the phrase structure. Lowe (2015:104) defines these forms as *portemanteau words*, displaying a number of properties typical of single lexical elements and other features characteristic of two-word sequences, for example, the fact that they alternate with unambiguous two-word sentences. From Lowe's (2015:3) perspective, this is meant to account for the widely accepted fact that a noun form used at the end of a bahuvrīhi must be considered at least partially adjectival.

If we accept independent evidence for the existence of non-projecting words, bahuvrīhis are no longer disturbing exceptions but show a pattern that, though marginal, is recognized elsewhere. Other approaches however resort to the non-grammatical notion of metaphoric/metonymic usage: *red cap* would thus be a *pars pro toto* metonymy. This is explicitly stated for instance by Pennanen (1982:245-246), who indeed interpreted Marchand's proposal in this way, and later*, e.g.,* by Booij 2005, Barcelona (2008:210) and Bauer (2008:59), as an alternative to the zero-reading. Nevertheless, since bahuvrīhis actually trigger a specific accent, it is not possible it get rid of any grammatical role of morphology or syntax, and to postulate a mere figurative relation between the whole compound and the constituent in the head position.

3.3 Open issues in modern models accounting for bahuvrīhi compounds

In short, there still seem to be room for discussion in the modern treatment of exocentric compounds, in particular we would like to recall some points here, which are relevant to a discussion on the morphological role of ellipsis.

First of all there is a still-unresolved issue at the basis of the constituent analysis of bahuvrīhi, which is whether their source must be identified as a matching endocentric compound (to be, so to say, "converted" into a bahuvrīhi)[19] or as an independent string of constituents. A number of Ṛgvedic examples seem to favour the second line of reasoning: compounds such as (8) have no matching endocentric compound that could be used as a source.

(8)   a. *aja-aśva-*   → *ajā́śva-* (ṚV 1.138.4; 6.55.3, epithet of Pūṣan)
       goat-STEM-horse-STEM 'having goats for horses'-ADJ
       b. *go-mātṛ-*   → *gomātṛ-* (ṚV 1.85.3, said of the Maruts, sons of Pṛśni)
       cow-STEM-mother-STEM 'having a cow for mother'-ADJ
       c. *iheha-mātṛ*   → *ihehamātṛ* (ṚV 6.59.2)
       here and there-ADV-mother-STEM 'whose mothers are here and there'-ADJ

Endocentric formations of the same type are lexically restricted and non-productive. This is crucial in understanding up to what point compounding may be properly accounted for by syntax rather than by morphology. While the pattern starting from a matching endocentric compound seems to favour the postulation of an ellipsis of some morphological element, this is much less so in the competing pattern in which the two types of compounds are derived independently. In this sense, particularly interesting are those models, in particular Kiparsky's, which identify some specifically morphological traits characterising exocentric compounds (e.g., *milkteeth* vs *sabertooths*): these need to be accounted for by any model starting from an independent string of constituents. Some similar phenomena are in fact also attested in Sanskrit and are dealt with by Pāṇini through affixation in the context of exceptions to the more general pattern (see below § 3).

A crucial feature of modern debate on the bahuvrīhi in general is the contrast between exocentric and endocentric compounds, which is indeed rooted in any analysis based on the notion of head. In such a head-oriented framework, exocentric compounds, whose surface-constituents typically do not include a head, are interpreted as marginal forms of compounding and recourse to the postulation of a possessive zero-suffix is required by the presence elsewhere of overt forms:

---

[18] Like most modern scholars, Lowe does not apply compounding rules to the last element of the compound itself.

[19] Gillon limits its statement to the case of homodenotative compounds (such as *dīrgha-kaṇṭha* 'long neck' vs *dīrgha-kaṇṭha-* 'long neck-ed'): however he fails to see that the 'conversion' is less easy to postulate in the case of Noun + Noun homodenotative compounds, such as *candra-mukha* 'moon-face(d)' or *ayo-muṣṭi* 'iron-fist', where 'the predication relation may be metaphorical, instead of literal' (Gillon 2008:4) because the matching endocentric forms are not freely productive in Sanskrit and are lexically restricted by Pāṇini himself.

- [pig+tail] : [[pig+tail] + -ed] = [bird+brain] : [[bird+brain] + Ø -ed]
  pigtail : pigtailed ('endowed with a pigtail') = *birdbrain : birdbrain/birdbrained ('endowed with a birdbrain')
- long leg-s : [[long+leg+Ø-s] + -(g)ed] : [[long+leg+-s] + Ø-ed]) → longlegged X/daddy long-legs ('characterized by long legs');
- blue eyes : [[blue+eye+Ø-s] -ed] /[[blue+eye+-s] +Ø -ed] → blue-eyed Y/Y (= proper name) blue-eyes ('characterized by blue eyes')

Now, since the appropriateness of identifying a specific class of exocentric compounds was called into question in recent times, the link between exocentricity and the postulation of zero might become weaker.

3.4 The role of affixation in accounting for bahuvrīhis

A strong element in favour of an interpretation through the the postulation of a zero-suffix lies in the fact that some bahuvrīhis are sometimes endowed with affixes (above all –*ka*: freely productive in Classical Sanskrit). Moreover, Pāṇini himself devotes the final section of his chapters on secondary affixation to a number of phenomena concerning compound nominal bases in the so-called *samāsānta*-section (A 5.4.68-160) where secondary affixation to bahuvrīhis plays an important role (5.4.113-160). Gillon (2008:3-4) considers this datum as independent confirmation of the postulated null suffix, at least with specific reference to the *homodenotative* bahuvrīhis, even though he admits that the distribution of two alternants (with and without -*ka*) is *somewhat different* (A 5.4.151 ff.).

Nevertheless, at least in considering the most ancient data, we see that, as already pointed out by Wackernagel ([1905] 1957:102), such suffixes are extremely rare in the *Ṛgveda-Samhitā* (only two occurrences of -*ka*- for instance) and in the *Atharvaveda*, while they are increasingly frequent in the *Yajurveda-Samhitas, Brāhmaṇas*, etc. Pāṇini himself labels the forms with -*ka* as marginal (see below under A 5.4.154). They can hardly be used to account for the need to postulate a zero suffix to form a bahuvrīhi. Moreover, the whole *samāsānta*-section is complex[20] and deserves further scrutiny. Yet we consider that some important basic facts can be safely stated and will help in understanding the phenomenon at stake. In the above-mentioned section:

— the suffixes themselves are said to be 'final constituents of compounds' (A 5.4.68),[21]
— their affixation is taught *prātipadikāt* 'after the last sound of a nominal base', like any secondary suffix of §§ 4-5,
— moreover, the nominal base is further specified as a compound nominal base and may consist of any compound or there may be rules that explicitly restrict affixation to specific compounds e.g., after a tatpuruṣa (A 5.4.86), a dvigu (A 5.4.99), a dvandva (A 5.4.106),
— compound nominal bases are further described in terms of listed final constituents or strictly phonic conditions; in other words, the affixation described is, for the most part, not productive.

These suffixes are added to compound verbal bases and in turn create compound verbal bases: in no way do they act as transcategorizers.[22] Their functions are very rarely connected with semantic aspects; mostly, they are useful either in providing a shift towards vocalic declension for compound bases with a last constituent showing a rare declensional class and/or in dealing with some accentual issue. In Pāṇinian terms, end-of-compound suffixes are radical *svārtha*-suffixes inasmuch as, as pointed out by Cardona (1983:68), a *samāsānta*-suffix 'neither signifies nor cosignifies any particular meaning' (some exceptions may be found, e.g., A 5.4.114; 127). Moreover, as pointed out by Cardona (1983:53) they 'generally follow the gender and number of the bases to which they are added. The affix aC (10a), for example, taught by

> A 5.4.78: The affix *aC* occurs after the nominal stem ending in –*varcas*- co-occurring with *brahman*- or *hastin*-

---

[20] Although affixation is predominant in this section, some specific substitution rules specifically restricted to compound nominal bases also find their place here.

[21] They occur after the last part of a compound and are themselves the final part of that same compound; see K *ad* A 5.4.68 What seems crucial for commentators is to state explicitly that the condition of being a bahuvrīhi must concern both the constituent acting as left-conditioning and the end-constituent. These suffixes are added to compounds and have compounds as their output.

[22] See also Wielińska-Soltwede (2019:10-13).

accounts for the final pitch of the whole compound (A 6.1.163-164), while affixes with marker P, such as *-kaP* in (9b), taught by A 5.4.154 (see below), are enclitic (A 3.1.4).

(9)   a.   *brahma-* +   *varcas-* +   *-aC* (*-á*)   → *brahmavarcasá-*
              sacred knowledge-   eminence-   *samāsānta*   'eminence   in   sacred
              STEM   STEM   SUFFIX   knowledge' N.SG
       b.   *bahu-* +   *khaṭva-* +   *-kaP*   → *bahukhaṭváka-*
              many-STEM   bed-STEM   *samāsānta*   'in   which   there   are   many
              SUFFIX   beds' ADJ.[23]

The most consistent group of rules of this section is specifically restricted to bahuvrīhis. The first rule of the section runs as follows:

> A 5.4.113: The affix ṢaC occurs after a base ending in *sakthi-* 'thigh' or *akṣi-* 'eye' – meaning one's own limb – in the context of a bahuvrīhi (*bahuvrīhau*). [24]

(10)   *dīrgha-* +   *sakthi-*   +   *-ṢaC*   → *dīrghasaktha-*
       long-STEM   thighs-STEM   *samāsānta* SUFFIX   'having long thighs' ADJ.

It is true that in most of these cases a contrastive value of the suffix is recognisable and has already been pointed out by commentators, in particular the KV *ad* A 5.4.113: Why [the mention] 'in case of a bahuvrihi'? [cf.] *paramasakthiḥ* 'supreme, excellent thigh'/*paramākṣi* 'excellent eye'. Nevertheless, such suffixes are all lexically restricted and show no hint of productivity: they testify to lexicalisation phenomena much more than to conversion. The only significant exception is the already-mentioned rule:

> A 5.4.154: After all the other [bahuvrīhi stems], marginally, the affix *-kaP* may occur.

In this case, besides some occurrences of lexicalized uses of *-ka* listed in the immediately preceding rules (A 5.4.151-153), both forms with and without *-ka* alternate freely, even though the form with *-ka* is dispreferred. Nevertheless, even more important is the restriction, seldom quoted in literature, taught in the following rule, a restriction which in fact limits the usage of *-ka* to non-lexicalized, merely attributive, contexts.

> A 5.4.155: Not in the context of a conventional denotation.

(11)   *viśvadeva*
       Viśvadeva- PROPER NAME (lit. 'having all the gods')

Rather than disambiguating between an exocentric and a (problematic) matching homodenotative endocentric, this productive rule seems to account for the distinction between the original usage of bahuvrīhis as idionyms and a – probably later – expansion of the purely attributive function. This is not the right place to probe further, but this matches quite well with what we know of Indo-European naming practice and with the already-mentioned posteriority of suffixed bahuvrīhis.[25]

4.   The role of zero in Pāṇini to account for denominal and compound stems

As we have just seen, no special zero in the place of a derivational suffix is taught by Pāṇini for the bahuvrīhi compounds. Indeed, only a wide-ranging zero-substitution is adopted in bahuvrīhi analysis, as in any other class of compounds: it is a zero-substitution of case endings, in order to explain how each compound-internal constituent is used in its mere stem status, although each stem

---

[23] Alternating with *bahukhaṭvá-* in the same meaning.

[24] As already mentioned by commentators, the usage of the locative case is unexpected here: since the term qualifies the two bases ending in *sakthi* and *akṣi*, an ablative or a genitive is expected, as elsewhere in the section. See Sharma (1990-2003 vol. IV:734).

[25] Indeed, a similar disambiguation may, in ancient times, have been attributed to accent, a fact well known when it comes to nominal bases (see, e.g. Lazzeroni 1995) but that may also have had a role to play in the domain of bahuvrīhi compounds: see the almost homophonous pairs such as *ádeva/adevá*, discussed by Melazzo (2010:121-137) albeit with a different interpretation.

represents precisely an inflected noun with its relevant meaning and syntactic behaviour in the final compound. In fact, as we already recalled above, Pāṇini's architecture of the word-formation rules is based on a variationist pattern teaching compounds (but also secondary derivatives, as we shall see) which alternate with a meaning equivalent phrase, with respect to which compounds are declared to be preferable (*vā*).[26] Each nominal constituent, i.e. each inflected nominal lexeme involved in the source-phrase taught in the formation of a secondary derivative or compound nominal stem is consequently obtained by means of a *LUK* zero-substitution of endings, according to the following rule:

> A 2.4.71: [*LUK* zero-substitution] of a nominal case ending which occurs as a part of a verbal base (*dhātu*) or of a nominal stem (*prātipadika*).

And consistently with Pāṇini's whole system, where no derivation or compounding rule aims at forming a noun endowed with an ending, the final secondary derivative or compound form thus obtained is a pure nominal stem and will in turn obtain appropriate endings. For instance, the following rule teaches how to form endocentric compounds, such as *rājaputrá-* 'king's son', occurring as a NOM.M.DU form in ṚV 10.40.3 referring to the divine Aśvin twins (Ved. *rājaputrá̄* = Skt. *rāja-putrau*):

> A 2.2.8: a noun inflected in the genitive case preferably (*vā*) combines with another inflected noun to obtain an endocentric tatpuruṣa compound.

Thus, the source-phrase taught by this rule, i.e. (12a) – which is made up of inflected nouns – has to be converted into the matching meaning-equivalent compound (12b) – which instead consists of pure nominal stems. To obtain (12b) from its source-phrase, the genitive ending (*-as*) of *rā́jnaḥ* and the nominative ending (*-as*) of *putraḥ* in (12a) are *LUK*-zero-replaced according to A 2.4.71, but the meaning conveyed by them in (12b) is maintained by their relevant zero-substitutes.

(12) a.   *rā́jnaḥ*   *putraḥ*
             (*rā́jnas*)   (*putras*)
             king-GEN.M.SG   son-NOM.M.SG.
             'king's son (i.e. prince)'
      b.   **rā́jṇ-(Ø-as) → rāja-   putrá-(Ø-s)*
             rāja-putrá-
             king-STEM-son-STEM
             'king's son (i.e. prince)'

The first constituent shows a specific (and, we will see, frozen) syntactic relationship with the second constituent; the second constituent on the other hand shows a generic syntactic valence represented by the nominative that will be variously implemented in the specific linguistic contexts. In fact it is the compound stem that will receive the final case ending (in our case NOM.M.DU) required by the broader context of the sentence. Since the compound is endocentric, it will very often receive the same number, gender and case of the corresponding word in the meaning-equivalent phrase. Analogously, when we form a mandatory endocentric compound, such as *iṣu-kārá-* 'arrow-maker' – which is also a bound primary derivative stem – the left-hand constituent, i.e. the nominal lexeme bound with the primary derivative stem (*kṛt*) *-kārá-* is subject to A 2.4.71.

---

[26] The present analysis of compounding relies on a couple of specific exegetic choices. We are here following Kiparsky's (1979:3) and Radicchi's (1988:56-58) overall interpretation of Pāṇini's compounding rules, in considering that the preferable option taught by *vā* in A 2.1.18 has to be continued in the following compounding rules. *vibhāṣā* 'marginally' is instead exclusively valid from A 2.1.11 to 2.1.17, where it is precisely dismissed by *vā*, and it is not to be continued in all the following compounding rules, as suggested by Patañjali, and commonly accepted – see e.g., Cardona (1997:219). As a consequence, the majority of compounding rules preferably apply because they are grouped together under the scope of *vā* (A 2.1.18-2.2.9), instead of *vibhāṣā*. Moreover, as for the relation between the compound and the source-phrase, we resort to A 2.1.1, interpreted according to Pontillo (2013: 111-118).

A 3.2.1: the primary derivative suffix *-á*[27] applies to a verbal base, provided that an inflected noun playing the role of direct object co-occurs.

(13)   *íṣum* + (**kṛ-* →) *kār-* + *-á*
         íṣu-(Ø-m)-kār-á-
         arrow-ACC.SG.-make-STEM+ SUFFIX
         'arrow-maker'

In *iṣu-kār-á-* the accusative case ending of the bound lexeme is zero-replaced, but its meaning is maintained, so that the nominal stem *íṣu-* acts as the direct object of the action denoted by the verbal base of the primary derivative nominal stem, i.e. the bound morpheme *-kārá-*.

The secondary derivative nominal stem *vaiyākaraṇá-* (Nir. 1.12; 9.5 and 13.8) denotes a 'grammarian', i.e. one 'who studies grammar, one who knows grammar'. It is formed by applying the secondary (*taddhita*) suffix *-áṆ* (taught by A 4.2.59) to its etymon. In rule (14), which is also the source-phrase used to form our example, the demonstrative pronoun (i.e. *tád* 'this', in the present rule) plays the role of a sort of variable X in the rule itself, which can be read as in (14b). The variable has to be replaced by the due lexeme inflected according to the syntax taught in the rule itself, i.e. in the accusative case, in order to form the desired derivative lexeme (14c).

(14) a. *tád*   *ádhīte*   *tád*   *véda*
        this-ACC.N.SG   study-PRS.3SG   this-ACC. N.SG   see-PF.3SG(= he knows)
        'he studies this, he knows this'
     b. he studies X, he knows X
     c. *vyākáraṇam*   *ádhīte/véda*
        grammar-ACC.SG   study-PRS.3SG/see-PF.3SG (=he knows)
        'he studies grammar, he knows grammar'

 Therefore, provided that the final derivative lexeme denotes one who performs the action conveyed by the verbal forms *ádhīte* or *véda* and the object of this action is denoted by the etymon of the derivative lexeme, the noun inflected in the accusative case is the starting point of the derivation. It is the accusative *vyākáraṇam*, envisioned as the object of these two above-mentioned verbal forms, that assures the right meaning of the derivative lexeme *vaiyākaraṇá-* and the right syntactic relation of the derivative lexeme with its etymon. Nonetheless, from a morphological point of view, a nominal stem (*vyākáraṇa-*) is required to which the suffix *áṆ* has to apply (as taught by 4.1.1 and 4.1.82), so that a zero-substitution of the accusative case ending of *vyākáraṇam* is mandatory, even though its accusative meaning has to be kept in the final meaning of the relevant derivative lexeme.

The *LUK*-zero substitute that Pāṇini postulates in all the above-surveyed derivations or compounds is an effective device he adopts to assure the desired meaning of inflected words for the pure stems involved in the relevant formations. If Pāṇini merely dealt with all these formations only by paying attention to their morphological features, he would not be able to account for the specific meanings conveyed by such formations.

In his analysis of bahuvrīhi-compounds, Pāṇini also involves *LUK*-zero substitutes for all the constituents. We shall return below to the other crucial details in Pāṇini's analysis of this type of compound, but let us first look at the constituents of a bahuvrīhi, such as the one occurring in (15a), as a qualifier of the goddess Aditi. Once again, all the case endings of the constituents explained in (15b) are *LUK*-zero-substituted according to A 2.4.71 in order to obtain a compound stem (15c).

(15)   a. *áditī*   *rā́ja-putrā-* (ṚV 2.27.3)
         Aditi-F.SGBAHUVRĪHI F.SG
         'Aditi whose sons are kings (i.e. "Aditi having kings as her sons")'
         b. *rā́jānaḥ*   *putrā́ḥ*
         **rā́jṇ-(Ø-as)>-rāja-*   *putra-(Ø-as)*
         king-NOM.PL   son-NOM.PL
         c. *rā́ja-putra-*
         king-STEM-son-STEM
         'sons who are kings'

---

[27] The marker *Ṇ* entails a *vṛddhi*-ablaut substitution of the penultimate short *a* vowel of the verbal base to which the primary derivational suffix applies (A 7.2.116).

Again, the compound stem will receive the final case endings required by the broader context of the sentence, more specifically, as we will see, required by another word in the sentence with which the compound is syntactically linked.

5. Pāṇini's analysis of bahuvrīhi compounds

The theorical frame within which one must evaluate Pāṇini's description of bahuvrīhis is the general frame sketched above. Within this frame, his simple and almost laconic treatment of bahuvrihis – about which Laurie Bauer (2010:175) complains – becomes fully significant. Indeed, after explaining a lot of details about endocentric compounding in dozens of rules of his grammar, Pāṇini only devotes two general rules to bahuvrīhis:

> A 2.2.23: *śeso*   *bahuvrīhih*   [*samāsaḥ* 2.1.3]
>   remainder-M.SG.   bahuvrīhi (technical term)-M.SG   compound-M.SG
> 'The remainder is the bahuvrīhi [compound]'.
> A 2.2.24: *an-ekam*   *anyapadārthe*
>   not-NEG=one-N.SG   other-inflected.word-meaning-LOC.M.SG
> 'The combination of two or more inflected words when the denotatum (*artha-*) is that of
> another (*anya-*) inflected word (*pada-*) (i.e. of an inflected word other than the compound-
> internal constituents) [is a bahuvrīhi compound A 2.2.23]'

Nonetheless, read against the background of the previous compounding rules, what Pāṇini seems to advance by means of the first *sūtra* is to designate the bahuvrīhi as a sort of default case of compounding typology and this is not a trivial principle. In fact, in order to stick with the *modus operandi* that the *śeṣa* device typically triggers in his grammar, *śeṣa* denotes, within the set of combinations of inflected words forming a compound, those units 'that are other than' the combinations listed in the previous rules.[28] However broad this rule may seem, a restriction is thus provided by A 2.2.23: this is operatively based on the exclusion of all the specific combinations listed in the previous compounding section. In other words, with respect to the set containing all possible combinations of inflected words forming a compound, these previous combinations are considered as a subset, and *śeṣa* denotes the residual subset, containing all and only the members not included in the first subset. And since the relations conveyed by all the case endings – with the exception of the nominative case ending – are included in the previous rules, one of Patañjali's explanations for *śesaḥ* assumes that it might just refer to the relation conveyed by the nominative ending.[29] We shall try to understand what new scenario this exclusion plausibly opens, and consequently what this residual of compounding rules actually consists of.

Such a "default status" assumedly attributed to the bahuvrīhi type is indeed well-tuned to Wackernagel's already-mentioned opinion about the origin of these compounds. Wackernagel ([1905] 1957:288) indeed maintains that primigenial Indo-European compounds might have substantially been formed only when a third concept was to be designated by such a combination. Therefore, even though the analysis of exocentric compounds as derived from their endocentric counterparts is a common practice,[30] tatpuruṣas would have to be relegated to a secondary and especially diachronically later role, since bahuvrīhis are far more frequent than tatpurusas in the earliest stages of ancient Indian language, exactly as in ancient Greek. It is noteworthy that e.g., bahuvrīhi (15a), i.e. the feminine *rā́japutrā* referring to the goddess Aditi, occurs in the so-called "Family Books", i.e. the most ancient sections of the ṚV, while tatpuruṣa (12b), i.e. *rāja-putrá-* is in the latest book and there is no further occurrence in the ṚV of the same combination of constituents as a compound.

A 2.2.24 *an-ekam anya-padārthe* is Pāṇini's most renowned rule on the structure of bahuvrīhis. The Western label "exocentric" might even have been inspired by its segment *anyapadārthe* (Sadovski 2002:352), although the term *artha* hardly matches the modern syntactic and semantic notion

---

[28] The present definition is tentatively modelled on a formula used for A 2.3.50 (*ṣaṣṭhī śeṣe*) by Cardona (2013:104). Cf. Sharma (2010:1).

[29] See M 1.419 ll.7-8 *ad* Vt. 2 *ad* A 2.2.23, about which recently Kobayashi (2016:180).

[30] See, e.g., Gillon (2008:2) §1 above, but cf. § 6 below: in particular cf. (21a) with (21b), (21a), (22a), with (21b), and (23a), (23b), (23c), with (23d).

of head. Indeed, Candotti and Pontillo (2019: 21-26) explained the drawbacks of using the notion of head, especially in analysing bahuvrīhis and, on the contrary, the advantages of making such an analysis contingent on the non-head of compounds (i.e. the *upasarjana*), as Pāṇini taught, with marked discontinuity with the previous grammatical tradition and his posterity, including his commentators.[31]

In fact, we maintain that precisely the concept of *upasarjana* (which only provisionally and for the sake of the discussion can we identify as the non-head) is the basis of a model that may successfully compete with the model of zero to account for exocentric compounds. Analysing the wording of A 2.2.24, we see that it is constructed precisely upon the concept of *upasarjana*, because *anekam* is inflected in the nominative,[32] which is the metalinguistic case reserved for this constituent in Pāṇini's system according to the following rule:

> A 1.2.43: What is stated in the nominative in a compound-[rule] is called *upasarjana.*

Due to this metalinguistic rule, for example, we are able to identify the constituent that plays the role of *upasarjana* in (16c), which is an endocentric compound formed according to pattern (16a) = A 2.1.32 on the basis of source-phrase (16b):

(16) a. <u>kartrkarane</u>   *kṛtā*   *bahulam*
       agent-STEM-instrument-NOM.N.DU   primary derivative noun-INS.SG   variously-ADV
       [An inflected word denoting] an agent or an instrument variously combines with a primary derivative noun.
       b. *nakhair nirbhinnaḥ*
       nail-INS.PL-tear- PAST.PTCP.NOM.M.SG
       'torn by means of nails'
       c. <u>nakha</u>-*nirbhinna*-
       nail-UPASARJANA-STEM-tear-PAST.PTCP.STEM
       'torn by means of nails'

The *upasarjana nakha*- occupies the first slot in (16c) and – as is well known – such a location is quite a common feature of the *upasarjana* constituent in endocentric Indo-European compounds, a feature that Pāṇini also fixed in the following general rule:

> A 2.2.30: A constituent termed *upasarjana* is placed first in a compound.

Nevertheless, there are some exceptions, such as (17b), generated on the basis of the source-phrase (17a) taught by the following rule:

> A 2.2.18: The indeclinable *pada ku-*, the units termed *gati* (A 1.4.60) or included in the list beginning with *pra-* (A 1.4.58) – such as *pra-, ati-, nis-* – mandatorily combine with an inflected word to form a tatpuruṣa compound.

(17)   a. *niḥ*   *kauśāmbyāḥ*
         out of-ADV   *kauśāmbī* (toponym)–ABL.F.SG
         'out of [the town named] Kauśāmbī'
       b. *niṣkauśāmbi*-
         out of-ADV- *kauśāmbī* (toponym)–STEM
         '[somebody] who is out of [the town named] Kauśāmbī'[33]

The indeclinable word *nis* – a prefix[34] treated in the same way as any inflected word in Pāṇini's grammar, as revealed by the fact that he teaches the zero-substitution of its case ending – is mentioned in the wording of rule A 2.2.18 by means of a metalinguistic nominative and, moreover, it is placed first in compound (17b), so that it is expected to play the role of an *upasarjana*. By contrast, *nis* is not the *upasarjana* and in the traditional constituent-analysis of compound *niṣkauśāmbi* it is interpreted as 'gone out of, departed from', i.e., as if it were the past participle *niṣkrānta-* including *nis-* as a preverb, which merely makes it easier for grammarians to show the relevant case ending (zero-substituted in the indeclinable word *nis*).

c. *niṣkrāntaḥ*   *kauśāmbyāḥ*
go out of-PAST.PTCP.NOM.M.SG   *kauśāmbī* (toponym)–ABL.F.SG
'gone out of [the town named] Kauśāmbī'

This is in fact the commentarial example used to explain the second rule defining the *upasarjana*, which states that it is recognisable even when, in a compound, it does not occupy the first slot (and is thus not mentioned in the nominative case):

A 1.2.44: And what has one single ending (*ekavibhakti*) is the *upasarjana*, even when, in a compound(-rule), it is not in the first place.

In the diagram below, following the above-mentioned traditional analysis, we can see how in the source-phrase of compound *niṣkauśāmbi-*, whatever the case ending applying to the resultant compound when used in a sentence, *niṣkrānta-* is inflected in several cases while *kauśāmbyāḥ* remains unchanged, *i.e.*, has always 'one single ending':

d. *niṣkrāntaḥ* go out-NOM.M.SG	*kauśāmbyāḥ* Kauśāmbī-ABL.F.SG from Kauśāmbī	→ *niṣkauśāmbiḥ* gone out of-Kauśāmbī-NOM.M.SG
niṣkrāntam go out -ACC.M.SG		→ *niṣkauśāmbyam* gone out of-Kauśāmbī-ACC.M.SG
niṣkrāntena go out-INS.M.SG		→ *niṣkauśāmbinā* gone out of-Kauśāmbī-INS.M.SG. [...][35]

Therefore, the second definition rule exclusively allows recognition of the *upasarjana* – not mentioned as a nominative in the formation rule, and not occupying the first slot in the compound – because the constituent *kauśāmbyāḥ* is linked to the other constituent by means of a frozen syntactic relationship. Thus, the *upasarjana* becomes a purely lexical subordinate constituent, just as the inflected noun *vyākáraṇam* 'grammar', used as a direct object in the source-phrase of *vaiyākaraṇá-* 'who studies grammar, who knows grammar', i.e. 'grammarian' – see above (14c), § 4 becomes a pure lexeme in its stem status (i.e. *vyākáraṇa-*, obtained by means of a *LUK*-zero-substitution according to A 2.4.71), to which the derivational affix *-á* and a specific ablaut apply. This subordinate status also involves a different treatment of the lexeme (*prātipadika*) itself by determining, e.g.*,* a shortening mechanism, such as the one explained by the following productive rule.

A 1.2.48: The final vowel of *go* or of a feminine-affixed noun applying to a nominal stem that is] termed *upasarjana* is replaced by a short vowel.

We have already seen the effect of this rule in (17), where the final long vowel of the feminine noun *kauśāmbī*, which is a final *prātipadika* and an *upasarjana*, is replaced by a short vowel. Now, since we are sure – as already emphasized at the beginning of the present chapter – that the combination of two or more inflected words of a bahuvrīhi have to be labelled *upasarjana-*, because *anekam* in A 2.2.24 is inflected in the nominative (see above, A 1.2.43), we must ponder the consequences of this analysis.

---

[34] *nis* is generally used as a verbal or nominal prefix, but sometimes it occurs as a preposition. See AVŚ 6.18.3: *ta īrṣyāṃ muñcāmi nír* 'I set your thought free from your jealousy'; AVŚ 7.115.3: *nír itáḥ* 'out from here'.
[35] Examples from KV *ad* A 1.2.44.

First of all, the *upasarjana*-status of the constituents affects some crucial morphological features in Pāṇini's formation of nominal compound or derivative stems. For instance, according to the just-mentioned rule A 1.2.48, the shortening of the final vowel of *go*-, e.g., in (18) *ariṣṭagu*- 'whose cows (*go*-) are unhurt (*ariṣṭa*-)' occurring in AVŚ 10.3.10, depends on this *upasarjana* label extended to the bahuvrīhi's *prātipadika*, i.e. to the combination of its surface-constituents.[36] Conversely, the shortening of the final vowel of *go*- does not apply either out of compound, when used as a fully syntactic word, or in several endocentric compounds where *go*- is not the *upasarjana*. See, e.g., the forms attested by the traditional lexicons, such as *bahusūtigo*- 'a cow that calves often'. Moreover, in the frequent tatpuruṣa type e.g., *gopati*- 'the lord of cows/of cowherds' (occurring several times in the ṚV), *go*- is also termed *upasarjana*, but the nominal stem (*prātipadika*) to which it belongs does not end with *go*- and its shortening dpes not apply.

6. In between morphology and syntax: the frozen relation of the constituents with the denotatum

Within such a framework of compounding, in our opinion, A 2.2.24 emphasizes the fact that, in the absence of a non-*upasarjana* constituent in the surface linguistic form of the bahuvrīhi, all its surface-constituents play an *upasarjana*-role and thus comply with A 1.2.44, characterizing them as having a single, frozen, case ending. As we can see in (18), in fact, each of these constituents is declined according to one single case ending, whatever the syntactic role of the compound in the sentence. Indeed, such a frozen syntactic combination of inflected nouns within the bahuvrīhi works as an etymon, contributing to its final meaning, which has to be *anyapadārtha*, i.e. it has to convey the denotatum of an inflected noun other than the compound-internal constituents. This final meaning crucially depends on the relationship of one single surface-constituent and the denotatum, because any other relation is realized through that constituent in a subordinate manner. Coming back to the already quoted example *ariṣṭagu*-, we propose the following analysis:

(18)

ariṣṭā (*ariṣṭāḥ*) unhurt-NOM.M.PL	*gāvo* (*gāvaḥ*) cows-NOM.M.PL	'sya (*asya*) that-GEN.M.SG	→ *ariṣṭagu-s* one whose cows are unhurt-NOM.M.SG
			→ *ariṣṭagu-m* one whose cows are unhurt-ACC.M.SG
			→ *ariṣṭagu-nā* by one whose cows are unhurt-INS.M.SG. [...]

The possessive relation expressed by the genitive *asya* in the traditional constituent analysis (*vigraha*) especially links *gāvaḥ* 'cows' to his possessor, who can thus be called *ariṣṭagu*- 'one whose cows are unhurt'. On the other hand, the relation between the denotatum (i.e. the possessor) and *ariṣṭāḥ* 'unhurt' is only indirectly assured by the relation between *gāvaḥ* and *ariṣṭāḥ*. In the source-phrase, in fact, both *upasarjana*s, i.e. *gāvaḥ* and *ariṣṭāḥ*, remain unchanged in their inflected form regardless of the case ending applying to the resultant compound when used in a sentence. The two frozen nominative cases play the role of subject and predicate in the nominal sentence constituting the source-phrase of the bahuvrīhi. They have the same function but a different scope. The nominative in *ariṣṭāḥ* signifies its synctactic–relation to the other constituent *gāvaḥ*, 'cows'; the nominative in *gāvaḥ* signifies its synctactic–relation with *ariṣṭāḥ* but also with an inflected–word other than the compound-internal constituents.

The specific shape of this relation between the etymon and the denotatum is expressed by the case of the demonstrative pronoun, a point we will tackle below. It is first necessary however to confirm that the internal structure of the *upasarjana* constituents matters, as can also be shown in the easiest examples of bahuvrīhis, e.g., the possessive ones, when the surface constituents do not

---

[36] Conversely, another set of provisions headed by A 4.1.14 concerns constituents that are not *upasarjana*. For instance, A 4.1.54 teaches an option between the derivational feminine affixes *-ā* and *-ī* for a specific set of nominal stems acting as *upasarjana*s. Accordingly, this option applies to the *upasarjana* compound nominal stem *candramukhī-/candramukhā*- 'moon-faced', but not to the negative tatpuruṣa stem *aśikhā*- 'without crest', where neither *śikhā* nor *aśikhā* is an *upasarjana*.

share the same case ending. In some bahuvrīhis, such as the Vedic *parjánya-retas-* (19), the constituent-analysis might be as follows:

(19)   a. *parjánya-retas-*
         thunder-STEM-seed-STEM
         'whose seed is from the thunder' (bahuvrīhi qualifying the feminine noun *iṣuḥ* 'an arrow/a reed' in
             ṚV 6.75.15)
         b. *parjanyād   reto   'syāḥ (= retaḥ + asyāḥ)*
         thunder-ABL.M.SG.   seed-NOM.N.SG   that-GEN.F.SG.
         'her seed is from the thunder'

   No direct syntactic relation holds between the denotatum of the bahuvrīhi (i.e. 'an arrow' typically made of reed growing in muddy soil) and the left-hand constituent of the compound (i.e. 'the thunder'), but only between the denotatum and the right-hand constituent (i.e. 'the seed'). The left hand constituent is fully encapsulated in the source phrase, as shown by the ablative case ending; the nominative of the right-hand constituent however makes it available to the relation with its denotatum. The specific nature of this relationship is further specified by the case of the pronoun in the constituent-analysis.
   The new method we here propose to explain Pāṇini's approach to bahuvrīhi is indeed inspired by the original descriptive method Pāṇini himself often used in explaining the source-phrase of secondary derivative nouns (*taddhita*s). In several *taddhita*-rules, indeed, besides using the pronoun *tad*, as a sort of pure place-holder, i.e. as something to be replaced by the lexeme actually used in the derivation – as we have just recalled above (§ 4) – Pāṇini employs a second pronoun, namely *idam*. He uses this second pronoun as a sort of anonymous "coat stand", on which a specific case ending may be hung, in order to convey the syntactic relation between the *upasarjana* of the secondary derivative stem (which is also its etymon)[37] and the denotatum (expressed or understood) of the secondary derivative stem. In fact, in the final surface-form of the derivative stem, no lexeme replaces or represents *idam*, but the meaning of the case-relation it conveys is indeed the distinctive trait of the final meaning of the *taddhita*, i.e., its specific (= non-generic) relationship with its etymon. We shall continue to label the variable expressed by *tad* (to be replaced by a specific lexeme) as X – as we did above (14c) – and instead, we shall call the second assumed variable Y, expressed in Pāṇini's rules by *idam.* Let us now focus on an easy example in rule A 5.2.94, which teaches the application of the affix -*mat* or -*vat* in the sense explained in the following source-phrase, in order to form two ṚV *taddhita* examples, *gómat* 'rich in cows' (20b¹) and *púṣpavat*- 'full of flowers' (20b²):

(20)   *tad   asyāsty   asmin*
         a¹. *tad   asya   asti*
         this-NOM.N.SG   that-GEN.M./N.SG   be-PRS.3SG
         'X belongs to Y'
         a². *tad   asmin   asti*
         this-NOM.N.SG   that-LOC.M./N.SG   be-PRS.3SG
         'X in Y'
         b¹. *gavo 'sya santi*
         (*gavaḥ*)   (*asya*)   *santi*
         cow-NOM.M.PL   that-GEN.M./N.SG   be-PRS.3PL
         'cows (= X = *upasarjana*) belong to Y,'
         i.e. cows belong to one who (or to an object that) can be called *gómat*- 'rich in cows' by using the
         matching secondary derivative nominal stem)
         b². *puṣpāṇy asmin santi*
         (*puṣpāṇi*)   (*asmin*)   *santi*
         flower-NOM.N.PL   that-LOC.N.SG   be-PRS.3PL
         'there are flowers (=X= *upasarjana*) within Y,

---

[37] From some of Pāṇini's rules, we learn that *upasarjana* is a technical term which also indicates the non-head constituent in secondary derivation. According to A 6.2.104, for example, *pāṇini-* is termed "*upasarjana*" of the *taddhita* derivative stem *pāṇinīya-* 'disciple of Pāṇini' in the compound *pūrvápāṇinīyāḥ* 'earlier disciples of Pāṇini'.

i.e. there are flowers within a place[38] which can be called *púṣpavat-* 'full of flowers' by using the matching secondary derivative nominal stem.

Above, in (14a) no second variable Y is involved, simply because *vaiyākaraṇá-* 'grammarian' is one 'who studies grammar' (*tád ádhīte*), i.e. the relation between the denotatum and the etymon *vyākáraṇam* (see 14c) is conveyed by the verbal ending of *ádhī-te* 'he studies', signifying the agent of the relevant action.

The present comparison between taddhita and bahuvrīhi patterns is not a ungrounded proposal, since the wording of two specific bahuvrīhi rules, i.e. A 2.2.27-28,[39] involve the mentioned demonstrative pronouns in the same way as taddhita rules to restrict the generic relation of the etymon with the denotatum. The earliest indigenous commentators, as we have seen, used to adopt either the pronoun *idam* in their constituent-analysis of the bahuvrīhi compound or the relative pronoun, which finally prevailed in the traditional commentaries.

Moreover, the relation between the etymon (more precisely: a specific constituent of the etymon) and the denotatum cannot be mechanically identified in a possessive relation: even though compound rules (with the notable exception of the already mentioned A 2.2.27-28) do not specify the exact nature of the postulated relationship, it is undeniable that different syntactic relations, besides the most common possessive genitive, may account for the final meaning of a bahuvrīhi. In particular, the earliest commentaries (Vt. 17-19 ad A 2.2.24) already identify two classes of bahuvrīhis: on the one hand, those characterized by a genitive or even by a locative relation with the denotatum (21) and, on the other, generally with a verbal noun as first constituent, characterized by a relation consisting of any verbal actant with the exception of the agent (22). The syntactic relation between the right-hand constituent (*mātaṅgās*) and the denotatum (here expressed as *vana* 'forest') is conveyed by means of the locative case. In the example (22), the frozen syntactic relation – expressed by means of the instrumental case – intervenes between the left-hand constituent (*ūḍhas*, which is in its turn combined with *rathas*) and the denotatum (here expressed as *anaḍvān* 'ox'). In the bahuvrīhi-analysis, Pāṇini does not actually teach the variable Y by means of a pronoun, but merely puts a meaning constraint on its constituent-combination, which has to convey the denotatum of another inflected word. Nevertheless, we maintain that this is indeed the 'residual' of the compounding rules (see above § 5: A 2.2.23 *śeso bahuvrīhih*). It is operatively based on the adoption of the whole range of specific combinations listed in the previous compounding sections, provided with the new constraint –unexplored in the previous rules[40] – namely *anyapadārthe.* Moreover, we consider it noteworthy that Pāṇini, just as we have seen in example (2a), involves the mere *artha* of the external constituent in the scheme of the bahuvrīhi source-phrase. In fact, such 'another inflected noun' can be expressed or understood: the identification of a specific external constituent is not required.[41]

In this context, the fact that Pāṇini's compounding rule pattern does not aim at identifying the *pradhāna* (lit. 'principal', i.e. head constituent), as on the other hand Patañjali and later tradition suggest, becomes fully significant. Indeed, the concept of *upasarjana* from Pāṇini's point of view is sufficient to account for all types of compounds, and even though the *upasarjana* is not independent at the morphological and syntactic level, it is effective in bringing about the construction of the final meaning of the resultant (compound or derivative) nominal stem.

If this interpretation of compounds is accepted, both the derivation through the zero-substitution of the affix and the explanation through the conversion of bahuvrīhis are proven to be completely extraneous to Pāṇini's original model. In fact, the present interpretation of bahuvrīhi compounds does not entail a recursive application of the previous rules by adding something and eventually replacing it with zero at the surface morphology level. It is not the output of the previous rules, to be merely embedded in the new exocentric compound, as if the endocentric compounds were to be used as a "substrate" for the matching exocentric ones. In fact, we have to bear in mind that it is

---

[38] In the ṚV there is a single feminine plural occurrence, referring to plants 'where there are plenty of flowers' (*óṣadhīḥ* […] *púṣpavatīḥ* ṚV 10.97.3).

[39] See, e.g., A 2.2.27: *tatra tenedam iti sarūpe* 'Two inflected nouns having the same linguistic form combine with one another to form a bahuvrīhi denoting "it (Y) is in this (X)" or "it (Y) is by means of this (X)"' to account for forms such as *keśā-keśi* (*yuddham*) 'fight where there is mutual pulling of the hair'.

[40] With the exception of rule A 2.1.21, teaching the same constraint for indeclinable compounds such as *lohita-gaṅgám* 'where Ganges river (*gáṅgā*) is red (*lóhitā*)'.

[41] The expression *vibhaktyarthe* involved in Vt. 7 *ad* A 2.2.24 seems to hint at this interpretation.

only one of the surface-constituents of the bahuvrīhi, in a privileged syntactic relation with the bahuvrīhi's denotatum.

To illustrate this point, let us focus again on *ūḍha-ratha-*. The denotatum of the whole compound *ūḍha-ratha-* is the agent (*kartṛ*) of the action expressed by the past participle *ūḍha-*, which occupies the left slot in the compound; by contrast, no direct relation links *ratha-* 'cart' (*i.e.,* the right-hand constituent) with such a denotatum. This is reflected in the traditional source-phrase (22a) of *ūḍha-ratha-*, where *anena* (which is the representation of the denotatum of the whole bahuvrīhi) combines with (or is in a constituency relation with) the past participle *ūḍha-* and not with *ratha-*.

On the other hand, since no possessive relation holds between the denotatum of *ūḍharatha-* and (one of) its surface-constituents, this bahuvrīhi cannot be explained as a mere transformation of (22c) by postulating that a zero affix conveying the sense of possessing applies to such a karmadhāraya.

(21)   a. *mattā bahavo mātāṅgā yasmin*

(*mattās*)	(*bahavas*)	(*mātāṅgās*)	(*yasmin*)
excited-ADJ.	many-NOM.M.PL	elephant-NOM.M.PL	which-LOC.N.SG

      tad mattabahumātaṅgaṃ [vanam]

(*tat*)	(*mattabahu-mātaṅgam*)	(*vanam*)
that-NOM.N.SG	BAHUVRĪHI	forest- NOM.N.SG

      b. *mattās + bahavas + <u>mātaṅgā</u>* = X; *yasmin* = Y

(22)   a. *ūḍho ratho'nenoḍharatho 'naḍvān*

(*ūḍhas*)	(*rathas*)	(*anena*)	(*ūḍharathas*)	(*anaḍvān*)
drive-PAST.PTCP.NOM.M.SG	cart-NOM.M.SG	that-INS.M.SG	BAHUVRĪHI	ox- NOM.M.SG

'A cart is drawn by that,' i.e. [it is] an ox [which can be called] *ūḍharatha-* 'by which a cart is drawn' (by using the matching bahuvrīhi compound).
      b. *<u>ūḍhas</u> +rathas* = X; *anena* = Y
      c. **ūḍha-*   *-ratha-*

drive-PAST.PTCP.STEM	cart-STEM

'drawn cart'

Brugmann (1905-6:75) had already underlined that such a transformation is not correct, and it does not suffice to consider such a typology of compounds as a sort of exception, as Whitney (1889: 510) suggested. There are many ancient examples of bahuvrīhi compounds comparable to *ūḍha-ratha*, i.e. compounds whose denotatum is in a privileged syntactic relation with the left-hand constituent. For instance, *kṛtábrahman-* occurs three times in the ṚV, always as NOM.M.SG (*kṛtábrahmā*), on each occasion with a different syntactic role played by the denotatum of the whole compound with respect to the *upasarjana*-constituent *kṛtá-*:
- in ṚV 2.25.1 (21a) the denotatum of masculine *kṛtábrahmā* is a man who won the favour of the god Bṛhaspati, since he was the agent (*kartṛ*) of the action expressed by the past participle *kṛtá*;
- in ṚV 6.20.3 (21b), it is the recipient (*sampradāna*) of the action, i.e. the god Indra, mentioned in the same hemistich;
- in ṚV 7.70.6 (21c), it is the substratum (*adhikaraṇa*) of the action, namely an (explicitly mentioned) sacrifice.

The relevant constituent analysis of these three occurrences might be as follows. None of these three occurrences of the same bahuvrīhi can be obtained 'by adding *having* or *possessing* to the meaning of the determinative' (Whitney 1889:501), which could be assumed as "substrate" of bahuvrīhis (23a), (23b), (23c), made up of the two surface-constituents *kṛtám* and *bráhman* – merely because no possessive relation holds between the denotatum of the assumed karmadhāraya, and the denotatum of the bahuvrīhi itself.

(23) <u>BAHUVRĪHIS</u>

    a. *kṛtám*   *bráhman*   *anena*   *kṛtábrahmā*

create-PAST.PTCP.NOM.N.SG	sacred formulation- NOM.N.SG	that-INS.N.SG	BAHUVRĪHI

      'One <u>by whom</u> the sacred formulation is created is a *kṛtábrahmā*'
    b. *kṛtám*   *bráhman*   *asmai*   *kṛtábrahmā + indras*

create-PAST.PTCP.NOM.N.SG	sacred formulation- NOM.N.SG	that-DAT.N.SG	BAHUVRĪHI	Indra-NOM.M.S

'Indra <u>for whom</u> the sacred formulation is created is *kṛtábrahmā*'

c. *kṛtám*   *bráhman*   *asmin*   *yajñas*   *kṛtábrahmā*

create-PAST.PTCP.NOM.N.SG   sacred formulation-NOM.N.SG   that-LOC.N.SG sacrifice-NOM.M.S BAHUVRĪHI

'That <u>in which</u> the sacred formulation is created is a *kṛtábrahmā* sacrifice'.

<u>KARMADHĀRAYA</u>

d. **kṛta-bráhman-*

create-PAST.PTCP.STEM-sacred formulation-STEM

'created sacred formulation'

More significantly, the inconsistency of such a transformation paradigm might also be shown by a pair of the above-quoted synchronically attested compounds, i.e. the two sole ṚV occurrences of compounds combining *rā́jan* and *putrá*, namely (15a), which is actually a possessive bahuvrīhi and (12b), i.e. a mere genitive tatpuruṣa. Aditi is indisputably [a goddess] 'having sons' according to the phrase *áditī rā́ja-putrā-* in ṚV 2.27.3, but not 'having king's sons', as we might be compelled to understand, if we persisted in adopting the mechanism of transformation of a matching endocentric (12b: *rāja-putrá-* 'king's son') into an exocentric one (15c: *rā́ja-putra-*), to explain the meaning of the latter. Indeed, Aditi's sons are kings (15b: *rā́jānaḥ putrā́ḥ*), i.e. she is 'having kings as her sons'. The source-phrase of (15a) cannot be the synchronically attested compound (12b).

Furthermore, the schema of (15a) is confirmed by other fairly ancient bahuvrīhi occurrences such as *vṛtráputrā* (ṚV 1.32.9) who is unequivocally Vṛtra's mother, i.e. 'a being whose son is Vṛtra', and a Vṛtra's son is nowhere mentioned.

To conclude, Pāṇini does not need to postulate a (zero) head, merely because he concentrates on the non-head constituent of compounds in order to achieve a quite different explanation of the compounding phenomenon in general. Modern linguists who rely on Pāṇini's model usually do not pay attention to this relevant specificity and consequently underexploit it. Therefore, there is no need of special ellipsis in the bahuvrīhi, but the generic *LUK* zero-substitution of case endings that is taught to obtain a stem status from each inflected word involved in the source-phrase to form compounds (as well as secondary derivative nominal stems, etc.), be it a head or non-head constituent. The non-head constituents (*upasarjana*) are characterized by a frozen syntactic relation expressed by a specific case ending. In the case of bahuvrīhi compounds, the specific syntactic relation linking the denotatum with *one* of the surface-constituents of the compound is represented by the case ending of the pronoun involved in the traditional constituent analysis of the compound itself. The existence of such privileged surface constituents indeed accounts for cases in which the syntactic valence of the verbal noun is exhausted outside the surface-constituents, which is a handy way of dealing with an aspect of the blurred status of compounding, in-between morphology and syntax that has puzzled and still concerns modern linguistics. There is thus no need to postulate another constituent besides the surface constituents (the etymon) to account for the final meaning of the bahuvrīhi: it is sufficient that such an etymon expresses the relation with another inflected word (overt or understood).

References
Primary Literature

A   [*Aṣṭādhyāyī*] *The Aṣṭādhyāyī of Pāṇini*, ed. and tr. by R. N. Sharma, 5 vols. Delhi: Munshiram Manoharlal, 1990-2003.

AVŚ   [*Atharvaveda Śaunakīyaa*] *Atharvaveda* (*Śaunaka*) *with the Padapāṭha and Sāyaṇācārya's Commentary*, ed. by V. Bandhu, Hoshiarpur: Vishveshvaranand Vedic Research Institute, 1960-62.

KV   [*Kāśikā-Vṛtti*] *Kāśikā. A Commentary on Panini's Grammar* by Vāmana and Jayāditya, ed. by Ā. Sarmā, Kh. Deshpande, and D.G. Padhye, Hydearabad: Sanskrit Academy Osmania University, 1969–1970.

M   [*Mahābhāṣya*]*The Vyākaraṇa-Mahābhāṣya of Patañjali*, ed. by F. Kielhorn, Bombay: Government central book depot 1880-1885. Third edition, revised and furnished with additional readings, references, and select critical notes by K. V. Abhyankar. Poona: Bhandakar Oriental Research Institute 1962.

Nir.   [*Nirukta*] The *Nighaṇṭu and the* Nirukta, ed. and tr. by Lakshman Sarup. Delhi: Motilal Banarsidass, 1920–1927.

ṚV   [*Ṛgveda*] *Ṛgveda-Saṁhitā* with a Commentary of Sāyaṇācārya, ed. by N. S. Sontakke – C.G. Kashikar, 4 Vols., Poona: Vaidika Saṃṣodhana Maṇḍala, 1983[2] ⌈1933-1951⌉.

VSM   [*Vājasaneyi Saṁhitā Mādhyandina* Recension] *Vājasaneyi Saṁhitā in the Mādhyandina and the Kāṇvaśākhā with the commentary of Mahidhara*, ed. by A. Weber, Berlin-London: Stenzler, 1849.

Secondary Literature

Allen, W. Sidney. 1955. Zero and Pāṇini, *Indian Linguistics* 16 (Chatterji Jubilee Volume,): 106–13.

Barcelona, Antonio. 2008. The interaction of metonymy and metaphor in the meaning and form of 'bahuvrihi' compounds. *Annual Review of Cognitive Linguistics* 6:208-281.

Bauer, Laurie. 2008. Exocentric Compounds, *Morphology* 18:51-74.

Bauer, Laurie. 2010. The typology of exocentric compounding. In S. Scalise, and I. Vogel, eds., Cross-disciplinary issues in compounding, pages 167-176. Amsterdam: Benjamins.

Böhtlingk, Otto. 1887. *Pāṇini's Grammatik,* hrsg., übers., erl. und mit versch. Indices, Leipzig: Haessel.

Booij, Geert. 2005. Compounding and derivation: evidence for construction morphology. In W.U. Dressler, F. Rainer, D. Kastovsky and O. Pfeiffer, eds., *Morphology and its Demarcation*s, pages 109-132. Amsterdam: John Benjamins.

Bopp, Franz. 1827. *Ausführliches Lehrgebäude der Sanskrita Sprache*. Berlin: Dümmler.

Bronkhorst, Johannes. 1981. The Orthoepic Diaskeuasis of the Ṛgveda and the Date of Pāṇini. *Indo-Iranian Journal* 23: 83-95.

Brugmann, Karl. 1905-1906. *Grundriß der Indogermanischen Sprachen*. Berlin: De Gruyter.

Candotti, Maria Piera and Tiziana Pontillo. 2013. The Earlier Paninian Tradition on the Imperceptible Sign. In T. Pontillo and M.P. Candotti, eds., *Signless Signification in Ancient India and Beyond* (Cultural, Historical and Textual Studies of Religions), pages 99-153. London: Anthem Press.

Candotti, Maria Piera and Tiziana Pontillo. 2014. Pāṇini's zero morphs as allomorphs in the complexity of linguistic context. In É. Aussant et J.-L. Chevillard, eds., *Proceedings of the XI International Conference on the History of Language Sciences, Workshop Les études sur les langues indiennes. Leur contribution à l'histoire des idées linguistiques et à la linguistique contemporaine*, pages 55-85. *Bulletin d'Etudes Indiennes* (ed. by N. Balbir), 32.

Candotti, Maria Piera and Tiziana Pontillo. 2019. Lexical subordination in derivation and compounding. *Studi e Saggi Linguistici.* LVIII (1):155-175.

Cardona, George. 1983. *Linguistic Analysis and Some Indian Traditions.* Pandit Shripad Shastri Deodhar Memorial Lectures. First Series. Post-graduate and Research Department Series 20. Poona: Bhandarkar Oriental Research Institute.

Cardona, George. 1997. *Pāṇini: His Work and Its Tradition*, Vol. 1: Background and Introduction**.** Delhi: Motilal Banarsidass.

Cardona, George. 1999. *Recent Research in Pāṇinian Studies*. Delhi: Motilal Banarsidass.

Cardona, George. 2002. The Old Indo-Aryan Tense System. *Journal of the American Oriental Society*, 122, 2 (Indic and Iranian Studies in Honor of Stanley Insler on His Sixty-Fifth Birthday): 235-243.

Cardona, George. 2013. Indological research: different standpoints (ed. by P.C. Muraleemadhavan), pages 99-146. Delhi: New Bharatiya.

Das Karunasindhu. 1985. *On Concept of Elision as a Phenomenon in Pāṇini's Grammar*, in *Proceedings of the Fifth World Sanskrit Conference, Varanasi October 21-26, 1981*, New Delhi, Rashtriya Sanskrit Sansthan: 131-137.

Deshpande, Madhav M. 1985. Historical change and the theology of eternal Sanskrit. *Zeitschrift für vergleichende Sprachforschung* (KZ) 99: 122-49

Deshpande, Madhav M. 1989. Ellipsis in modern Linguistics and Pāṇini. *Annals of Bhandarkar Oriental Institute* 70:103-124.

Collinder, Björn. 1962-1968. Les origines du stucturalisme. *Acta Societatis Linguisticae Upsaliensis*, n.s., 1: 1-15.

Das, K. 1985. On Concept of Elision as a Phenomenon in Pāṇini's Grammar. In R.N. Dandekar, P.D. Navathe, eds., *Proceedings of the Fifth World Sanskrit Conference,* Varanasi, October 21-26, 1981, New Delhi: Rashtriya Sanskrit Sansthan, pages, 131-37.

Deshpande, Madhav M. 2019. Scope of Early Sanskrit Usage: A Wider Approach. In: M. Kulkarni and P. Scharf, eds., P*roceedings of the 17th World Sanskrit Conference, Vancouver, Canada, July 9-13, 2018, Section 3: Vyākaraṇa.* <https://open.library.ubc.ca/cIRcle/collections/70440/items/1.0379841>

Dvivedi, H.P. 1978. *Studies* in Pāṇini. Technical Terms of the *Aṣṭādhyāyī*, Delhi: inter-India Publications.

Gillon, Brendan. 2008. Exocentric Compounds in Classical Sanskrit. In G. Huet and A. Kulkarni, eds., *First International Sanskrit Computational Linguistics Symposium*, Oct. 2007, Rocquencourt, France. http://hal.inria.fr/SANSKRIT/fr/. <inria-00202860>

Goldstücker, Theodor. 1861. *Panini: His Place in Sanskrit Literature*. London: Trübner [Reprint Osnabrück: Zeller 1966].

Hock, Hans Henrich. 2012. Sanskrit and Pāṇini: Core and periphery. *Sanskrita Vimarsha*, vol. 6: 85-102.

Houben, Jan E. M. 2018. Linguistic Paradox and Diglossia: The Emergence of Sanskrit and Sanskritic language in Ancient India. *Open Linguistics* 4: 1–18.

Kastovsky, Dieter. 1982. *Wortbildung und Semantik*, Schwann-Bagel: Pädagogischer Verlag.

Katre, Sumitra M., ed. 1988. *Aṣṭādhyāyī of Pāṇini*. Austin: University of Texas Press.

Kielhorn, Franz. 1870. *A Grammar of the Sanskrit Language*. Bombay: (Presidency). Educational Department.

Kiparsky, Paul. 1979. *Pāṇini as a Variationist*. Ed. S. D. Joshi. Cambridge, MA: MIT Press/Pune: Publications of the Centre of Advanced Study of Sanskrit.

Kiparsky, Paul. 1982a. Cyclic Phonology to Lexical phonology. In H. Hulst and N. Smith, eds., *The Structure of Phonological Representation*, Vol. 1, pages 131-175. Dordrectht: Foris.

Kiparsky, Paul. 1982b. Lexical morphology and phonology. In The Linguistic Society of Korea, ed., *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co.

Kiparsky, Paul. 1982c. Some Theoretical Problems in Pāṇini's Grammar. Poona: Bhandarkar Oriental Research Institute.

Kiparsky, Paul. 2012. Pāṇini, Variation, and Orthoepic Diaskeuasis *Asiatische Studien* 66: 327-335.

Kobayashi, Masato. 2016. Patañjali on Borderline Cases of Compound Classification. In G. Cardona and H. Ogawa, eds., *Vyākaraṇaparipṛcchā. Proceedings of the Vyākaraṇa Section of the 16th World Sanskrit Conference*, pages 175-194. New Delhi: DK Publishers/Bangkok: Silpakorn University.

Kulikov, Leonid, 2013. Language vs. grammatical tradition in Ancient India: How real was Pāṇinian Sanskrit?, *Folia Linguistica Historica* 34.

Lazzeroni, Romano. 1995. La baritonesi come segno dell'individuazione. Il caso del vocativo indoeuropeo. *Studi e Saggi Linguistic* 35:33- 44.

Liebich, Bruno 1891. *Panini. Ein Beitrag zur Kenntnis der Indischen Literatur und Grammatik*. Leipzig: Haessel.

Lowe, John J. 2015. The syntax of Sanskrit compounds. *Language* 91 (3):71-114.

Marchand, Hans. 1967. Expansion, transposition, and derivation. In D. Kastovsky, ed., *Studies in Syntax and Word-Formation*. Selected articles by H. Marchand, pages 322-337. Münchener: Fink.

Marchand, Hans. 1969. The Categories and Types of present-day English word-formation. München, Verlag C. H. Beck.

Meier, Georg Friedrich. 1961. *Das Zéro-Problem in der Linguistik* (Schriften zur Phonetik, Sprachwissenschaf und Kommunikationsforschung nr. 2). Berlin: Akademie Verlag.

Melazzo, Roberta. 2010. *I Bahuvrīhi del* ṚgVeda. Institut für Sprachen und Literaturen der Universität Innsbruck: Innsbruck.

Monier Williams, Monier. 1857. *A Practical Grammar of the Sanskrit Language*. Oxford: Oxford University Press.

Pennanen, Esko V. 1982. Remarks in Syntagma and Word-Formation. *Folia Linguistica* 16:241-261.

Pinault, George-Jean. 1989. Parole articulée et vérité; Travaux à partir du corpus védique; Pāṇini et l'enseignement grammatical; Procédés pāṇinéens; Le système de Pāṇini. In S. Auroux, ed., *Histoire des idées Linguistiques*, 1. Liege-Bruxelles: Mardaga, 293-401.

Pontillo, Tiziana. 2002. La prima ricezione del modello morfologico di Pāṇini nella linguistica occidentale: il caso dello zero. In V. Orioles, ed., *Idee e parole. Universi concettuali e metalinguistici* ("Lingue, linguaggi, metalinguaggio" 3, collana dir. da C. Vallini e V. Orioles), Roma; Il Calamo, 535-587.

Pontillo, Tiziana. 2013. Where the sense is intended although the corresponding speech unit is not employed: the *ekaśeṣa* case. In G. Cardona, ed., *Vyākaraṇa Across the Ages. Proceedings of the 15th World Sanskrit Conference*, Delhi 5-10 January 2012 (Vol. II: Vyākaraṇa Session). Pages 107-43. New Delhi: D.K. Printworld.

Radicchi, Anna. 1985; 1988. *La teoria pāṇiniana dei samāsa secondo l'interpretazione delle scuole grammaticali indiane dal quinto all'ottavo secolo d.C.* 2 Volumes. Firenze: Elite.

Renou, Louis. 1942. *Terminologie Grammaticale du Sanskrit*, Paris: Champion, 3 voll.

Renou, Louis. 1947-1954. *La Grammaire de Pāṇini*, traduite du Sanskrit avec des Extraits des commentaires indigènes. 3 Volumes. Paris: Klincksieck.

Roodbergen, Jouthe A. F. 2008. *Dictionary of Pāṇinian Grammatical Terminology*. Pune: Bhandarkar Oriental Research Institute.

Roodbergen, Jouthe A. F. 1974. *Patañjali's Vyākaraṇa-Mahābhāṣya. Bahuvrīhi/Dvandvāhnika* (P 2.2.23-2.2.38): Text, Translation and Notes. Pune: Publications of the Centre of Advanced Study of Sanskrit C n° 9. Class).

Rousseau, Nathalie. 2016. *Du syntagme au lexique. Sur la composition en grec ancien.* Les Belles Lettres: Paris.

Sadovski, Velizar. 2002. Dvandva Tatpurusa and Bahuvrīhi on the Vedic Sources for the Names of the Compound Types in Pāṇini's Grammar. *Transactions of the Philological Society* 100 (3):351-402.

Sen, Subhadra Kumar. 1999. Word Ordering in the Aṣṭādhyāyī. *Journal of Indoe/European Studies* 27, 1-2, 101-103.

Sharma, Rama Nath, ed. and tr. 1990-2003. *The Aṣṭādhyāyī of Pāṇini*. 6 Vols. New Delhi: Motilal Banarsidass.

Sharma, Rama Nath. 2010. Rule Interaction, Blocking and Derivation in Pāṇini. In *Sanskrit Computationl Linguistics, 4th International Symposium*, New Delhi, India, December 10-12, 2010 (Lecture Notes in Artificial Intelligence 6455), pages 1-20. Berlin-Heidelberg: Springer.

Štekauer, Pavol. 1998. *An Onomasiological Theory of English Word-Formation*. Amsterdam – Philadelphia: John Benjamins.

Thieme, Paul. 1935. *Pāṇini and the Veda (Studies in the Early history of Linguistic Science in India)*. Allahabad: Globe Press.

Vasu, S.C., ed., tr. 1891. *The Aṣṭādhyāyī of Pāṇini,* Allahabad: Pāṇini Office. [Reprint Delhi: Motilal Banarsidass 1988].

Varma, Siddheshwar. 1978. *Pāṇini and Elision, being An Analytical Study of Pāṇini's Sūtras on Lopa (Elision) Sanskrit* (Panjabi University Indological Series 12). Hoshiarpur: Vishveshvaranand Vishva Bandhu Institute of Sanskrit and Indological Studies.

Wackernagel, Jakob. [1905] 1957. *Einleitung zur Wortlehre. Nominalkomposition* (Altindische Grammatik, Band II.1). Göttingen: Vandenhoeck & Ruprecht.

Wielińska-Soltwede, Małgorzata. 2019. Svārthika Suffixes vs. Endocentric Taddhita Derivatives. In: M. Kulkarni and P. Scharf, eds., P*roceedings of the 17th World Sanskrit Conference, Vancouver, Canada, July 9-13, 2018, Section 3: Vyākaraṇa.* <https://open.library.ubc.ca/cIRcle/collections/70440/items/1.0379841>

Whitney, William D. 1889[2]. *A Sanskrit Grammar, including both the Classical language, and the older Dialects, of Veda and Brahmana*. Harvard University Press, Cambridge, MA.

Zakharyin, Boris A. 2007. PaaNini's Lopa and Zeroing in Modern Standard Hindi. In: Masica, C.P., ed., Old and new perspectives on South Asian languages: Grammar and Semantics. *Papers growing out of the Fifth International Conference on South Asian Linguistics* (ICOSAL-5), held at Moscow, Russia in July 2003 (MLBD Series in Linguistics XVI). Delhi: Motilal Banarsidass, 337-348.

The possessive relation in Sanskrit bahuvrīhi compounds:
Ellipsis or movement?

DAVIDE MOCCI, *SCUOLA UNIVERSITARIA SUPERIORE IUSS PAVIA*

ABSTRACT

Many Sanskrit bahuvrīhis involve a possessive relation whereby one of the bahuvrīhi-members is the possessum and an expression not mentioned within the bahuvrīhi is the corresponding possessor: e.g., *ugra-putra-* (ṚV 8.67.11), not 'mighty son(s)' but 'Aditi having mighty sons' or 'Aditi whose sons are mighty'. This study addresses the following research question: how is this possessive relation established in Sanskrit bahuvrīhis? We consider two possible strategies. According to the first strategy, a linguistic unit which conveys the meaning 'having' and undergoes ellipsis combines with the bahuvrīhi stem: e.g., the combination of this elided unit with *ugra-putra-*, which per se would convey the meaning 'mighty son(s)', yields the meaning 'having mighty sons'. According to the second strategy, the possessor starts out within the phrase projected by one of the bahuvrīhi-members: e.g., *áditi-* (i.e., the Sanskrit term for 'Aditi') starts out as the specifier of the phrase projected by *putrá-* in the above example; in this configuration *áditi-* is read as the possessor of *putrá-*; only subsequently will *áditi-* exit the bahuvrīhi. We argue that the second strategy is superior because only it captures certain restrictions on the internal order of bahuvrīhis.

1 Introduction

Compounding is an extremely productive phenomenon in Sanskrit: in fact, "almost any meaning that can be expressed using two or more separate words can also be expressed using a single compound" (Lowe 2015:72) in this language. For this reason, special attention has been devoted to Sanskrit compounding in the Indian grammatical tradition (starting from Pāṇini's section A 2.1-2) as well as in modern linguistic treatises such as Wackernagel (1905). The development of formal theories of language in the second half of the Twentieth Century has offered a fresh take on this subject matter. Thus Gillon (2008) capitalises on theories of word-formation such as Selkirk (1982) and Di Sciullo and Williams (1987) to formulate explicit constraints on the interpretation of Sanskrit bahuvrīhi compounds. On the other hand, Lowe (2015) analyses Sanskrit major compound types within the framework of Lexical-Functional Grammar, showing that they are generated by syntactic rules (i.e., the rules that combine words into sentences), rather than by morphological rules (i.e., the rules that combine morphemes into words).

Pursuing this line of research, in this work we provide a formal analysis of Sanskrit bahuvrīhi compounds (or simply bahuvrīhis), an exemple of which is *ugra-putra-* in (1) (for an explicit definition of bahuvrīhis, see §2 below).[1] The bahuvrīhi *ugra-putra-* is formed by combining *ugrá-* 'mighty' with *putrá-* 'son(s)'. However, *ugra-putra-* does not denote a mighty son, but rather a possessor of mighty sons. The information as to who this possessor is (i.e., the identity of this possessor) is not contained within the bahuvrīhi *ugra-putra-*: this information is supplied by a

---

[1] We use the hyphen at the end of a Sanskrit form (e.g., *ugrá-*) to cite that form as a stem (for an explicit definition of stem, see §2.1 below). *ugra-putra-* is accented on its first syllable (i.e., *úgraputre*) in (1) because it is a pāda-initial vocative (Macdonell 1910:81, 97). Since the accent displayed by a form inflected in the vocative and occupying the pāda-initial position need not coincide with the normal accent of that form, and since (1) constitutes the only occurrence of *ugra-putra-* as a bahuvrīhi in accented sources, we cannot be certain as to which syllable *ugra-putra-* is normally accented on. We therefore omit the accent of *ugra-putra-* when we quote this form outside (1). More on the accent of bahuvrīhis in §2.2 below.

bahuvrīhi-external expression. Thus, *adite* in the passage reported in (2) reveals that the possessor of mighty sons is indeed the goddess Aditi.

(1)   úgra-putre   (adite)   (ṚV 8.67.11)
       mighty-son.VOC.SG.F   Aditi.VOC.SG.F
       'O Aditi, having mighty sons (or 'O Aditi, whose sons are mighty').

(2) *utá tvā́m <u>adite</u> mahi/ aháṃ devi úpa bruve/ sumṛḷīkā́m abhíṣṭaye// párṣi dīné gabhīrá ā́m̐/ <u>úgra-putre</u> jíghāṃsataḥ/ mā́kis tokásya no riṣat//*
    'And, great goddess Aditi, I entreat you, the very merciful, (for us) to prevail. Deliver us, whether we're in the shallows or the deep, from someone who wishes to smite us, o you who have powerful sons. Let none of our offspring be injured.' (ṚV 8.67.10-11; tr. Jamison and Brereton 2014).

The bahuvrīhi-external expression that reveals the identity of the possessor denoted by the bahuvrīhi is known as the external referent of the bahuvrīhi. Thus, *adite* is the external referent of *ugra-putre* in (1). The brackets enclosing *adite* in (1) signal that *adite* is not actually adjacent to *ugra-putre* in the line quoted. We shall follow the same practice when exemplifying bahuvrīhis throughout this work: we shall systematically pair bahuvrīhis with their external referent, and bracket this external referent in case it is not adjacent, in the textual passage quoted, to the bahuvrīhi it is associated with.

A possessive relation is implied in *ugra-putra-*: *putrá-* 'son(s)' is the possessum and the external referent *adite* is the corresponding possessor in (1).[2] However, no overt linguistic unit signals the presence of the possessive relation in this example: there is no overt verb or suffix which conveys the meaning 'having' in (1), and no relative pronoun corresponding to *whose* of the English translation shows up either. We then wonder: how is the possessive relation that is perceived in bahuvrīhis like (1) established? At least two solutions come to mind, namely (3a) and (3b) (a third solution is considered in §2.4 below).

(3) a. <u>The Elided Possessive Unit Strategy</u>
    A linguistic unit that conveys the meaning 'having' undergoes ellipsis: to wit, the meaning 'having' conveyed by this unit is perceived, although this unit is not phonologically realised. Let us call this elided unit B. The combination of the bahuvrīhi stem, which conveys the meaning 'X', with B causes the bahuvrīhi to acquire the meaning 'having X'.
    b. <u>The Movement Strategy</u>
    An expression E_1 may be read as the possessor of another expression E_2 (E_2 being the possessum of E_1) when E_1 occupies the specifier of the phrase projected by E_2. The possessor (E_1) subsequently moves out of the phrase projected by E_2, so that no possessor of E_2 results as being visible within the phrase projected by E_2 itself.

According to (3a), the bahuvrīhi stem *ugra-putra-* ((1)), which would convey the meaning 'mighty son(s)', combines with B, thereby yielding *ugra-putra*-B, where B is silent qua elided: in this way, *ugra-putra*-B comes to denote the property of having mighty sons. The external referent of the bahuvrīhi will then supply the bearer of the property denoted by the bahuvrīhi itself: in (1), the goddess Aditi denoted by *adite* bears the property, denoted by *ugra-putra*-B, of having mighty sons. Hence, *adite* is interpreted as the possessor and *putrá-* as the corresponding Possessum in (1). To be noted that we are using 'elided' in (3a) in the sense of 'which has undergone ellipsis'. Ellipsis as we use it here is a cover term for all the linguistic phenomena in which "there is more meaning comprehended than what is presumably directly conveyed by the components of an expression" (Deshpande

---

[2] The possessive relation is merely one of the several types of relations that may hold between the external referent of the bahuvrīhi and one of the bahuvrīhi-members in Sanskrit. The traditional examples illustrating this point are reported in Cardona (1997:220). See also Pontillo (2021).

1989:103; see also Candotti and Pontillo 2022:2-5): in the context of (3a), the meaning 'having' is perceived but is not directly conveyed by any of the overt components of (1).[3]

On the other hand, according to (3b), the external referent *adite* starts out within the bahuvrīhi *ugra-putra-*, as the specifier of the phrase projected by *putrá-*: in this configuration *adite* is read as the possessor and *putrá-* as the corresponding possessum. *adite* subsequently moves to a bahuvrīhi-external position, so that the possessor of *putrá-* is not visible within *ugra-putra-*.

In this work we argue that the possessive relation perceived in Sanskrit bahuvrīhis like (1) is established by means of The Movement Strategy (3b), and that The Elided Possessive Unit Strategy (3a) is in fact unavailable. The argument that we build for achieving this conclusion capitalises on the internal order of bahuvrīhis (the internal order of a compound is the order in which that compound's members are arranged): only the Movement Strategy predicts the internal orders in (4), (5), (6), and (7) (the syntactic functions specified in the glosses are commented on in §2.3).

(4) úgra-putre   (adite).   (ṚV 8.67.11)
     [mighty]-[son].VOC.SG.F   [Aditi].VOC.SG.F
     [Predicate]-[**Subject**]   [Possessor of **Subject**]
   'O Aditi, having mighty sons' (or 'O Aditi, whose sons are mighty').

(5) vṛṣaṇ-aśvéna   (ráthena).   (ṚV 8.20.10)
     [bull]-[horse].INS.SG.M   [chariot].INS.SG.M
     [Predicate]-[**Subject**]   [Possessor of **Subject**]
   'With the chariot having bull-like horses' (or 'with the chariot whose horses are like bulls').

(6) vṛtrá-putrā   (dā́nuḥ).   (ṚV 1.32.9)
     [Vṛtra]-[son].NOM.SG.F   [Dānu].NOM.SG.F
     [Subject]-[**Predicate**]   [Possessor of **Predicate**]
   'Dānu, having Vṛtra as her son (or 'Dānu, of whom Vṛtra is the son').

(7) índro   vájra-hastaḥ   (ṚV 1.173.10)
     [Indra].NOM.SG.M   [mace]-[hand].NOM.SG.M
     [Possessor of **Predicate**]   [Subject]-[**Predicate**]
   'Indra, having a mace in his hand (or 'Indra, in whose hand is a mace').

Some preliminary remarks are in order at this point. First, the bahuvrīhis in (4)-(7) belong to Vedic Sanskrit, namely the earliest stage of Sanskrit (1500-600 BCE): we only consider bahuvrīhis from Vedic Sanskrit in this work. Second, by saying that we predict a bahuvrīhi's internal order, we mean that we make a claim to the effect that a certain bahuvrīhi, with a certain internal order, will be found in the Vedic texts hosted by the DCS (The Digital Corpus of Sanskrit—Hellwig 2010-2021) and by TITUS (Thesaurus Indogermanischer Text- und Sprachmaterialien—Gippert et al. 2016). Third, the uploading of Vedic texts to the DCS and TITUS has not yet been completed, so any of our statements of (non-)attestation can only be partial, and valid until proven otherwise. Last, the derivation which allows us to predict the internal orders in (4)-(7) will be cast in the framework of generative grammar.

The remainder of this paper is organised as follows. §2 provides an explicit definition of the relevant technical terms concerning Sanskrit compounding, and contains an argument for choosing between the Elided Possessive Unit Strategy and the Movement Strategy. §3 sets the stage for the Movement Strategy (3b), by reviewing some arguments in favour of a syntactic derivation of Sanskrit compounds. §4 advances a formal derivation of bahuvrīhis that is based on the Movement Strategy, and shows that the restrictions on bahuvrīhis' internal order intertwine with locality conditions on movement. §5 concludes the paper.

---

[3] Ellipsis in the sense used here is not to be confused with the notion 'copy' developed within the copy-theory of movement (Chomsky 1995:183-194; 2019:26). Simplifying somewhat, while copies result from an application of the movement operation, the ellipsis of suffix B described in (3a) does not involve movement.

2 A definition of bahuvrīhi

In this section we define bahuvrīhis. Then, we spell out the argument for choosing between The Elided Possessive Unit Strategy (3a) and The Movement Strategy (3b).
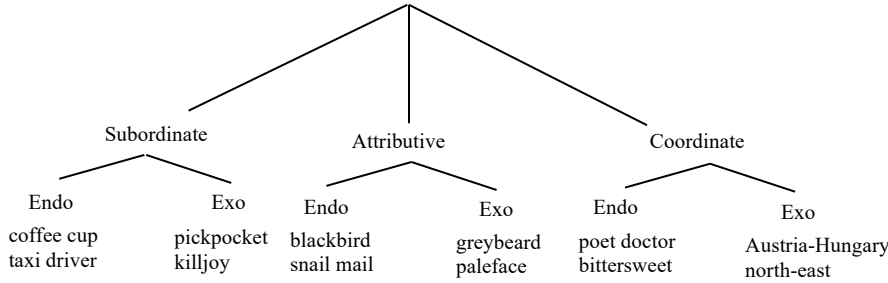
2.1 What is a bahuvrīhi?

A compound is a lexical category Z made up of two other lexical categories X and Y (Guevara and Scalise 2009:107, Scalise and Vogel 2010:6). More formally: $[X\ Y]_Z$, where X and Y are said to be the internal members of Z. Lexical categories include nouns, adjectives, verbs, and prepositions.[4] The morphological realisation of X and Y changes from language to language: thus, X and Y may be roots, stems, inflected words (Bauer 2017:4).

An important characteristic of Sanskrit (Latin, Ancient Greek, etc.) compounds is that X of "$[X\ Y]_Z$" is realised as a stem, while Y is a fully inflected word (Lowe 2015:72, 93): e.g., *aśvaśapháḥ* 'horse's hoof' (cf. *aśvaśaphéna* in ŚB 13.3.4)[5] is analysed as $[_Z\ [_X\ aśva]$-$[_Y\ śaphá$-$ḥ]]$, where *aśva-* 'horse' is a stem and *śapháḥ* 'hoof' a word inflected in the NOM.SG.[6] Given an independent word W, the stem of W is what is left when any inflectional ending of W is dropped (Lieber and Štekauer 2009:5; Kastovsky 2009:324): e.g., the stem of the independent word *aśvasya* is *aśva-*, obtained by dropping the GEN.SG nominal ending *-sya* of *aśvasya*.

Scalise and Bisetto (2009:§4) proposed an influential classification of compounds, which may be summarised by means of the following tree (from Scalise and Vogel 2010:7):[7]

(8)

However, there is no single node, in (8), into which the whole category of Sanskrit bahuvrīhis could fit: indeed, Sanskrit bahuvrīhis may be subordinate exocentric (e.g., *vájra-hasta-* of (7)), attributive exocentric (e.g., *ugra-putra-* of (4)), as well as coordinate exocentric (e.g., *somendrá-* TS 2.3.2.6 '[the oblation] belonging to Soma and Indra'). Defining Sanskrit bahuvrīhis as exocentric compounds would not work either; this is because while all bahuvrīhis are exocentric, not all exocentric compounds are bahuvrīhis (Bauer 2017:64-71): e.g., so-called *upapadasamāsa*s like *mayo-bhū́-* (AVŚ 1.5.1) '(the waters) by means of which delight obtains' would count as exocentric in accordance with the hyponymy test, but are not treated as bahuvrīhis by the Indian grammatical tradition. Bauer's (2017:65) definition of bahuvrīhis as "compounds which canonically label a part of the whole which the compound denotes" is problematic, too, because many bahuvrīhis in Sanskrit do

---

[4] See Corver (2013) on the distinction between lexical and functional categories.

[5] As the attentive reader may have noted, the accent on the initial *a-* of *áśva-* has disappeared in *aśva-śaphá-*. This illustrates a general tendency of compounds, namely that they have exactly one accent. For a general overview of compound accentuation, see Whitney (1899:484-485), Wackernagel (1905:40-43), and Macdonell (1910:91-97).

[6] Note that, in accordance with Pāṇini's rule A 2.4.71, *aśvaśaphénaḥ* is to be analysed as [[aśva]-[śapha]]-ḥ, where both *aśva-* and *śapha-* are fully inflected words (*pada*s) whose nominal endings have been zero-replaced: *-ḥ* qualifies as the ending of the whole compound-stem *aśvaśaphá-* rather than as the ending of *śapha-* (see Cardona 1997:21-23, 186, 207; Kiparsky 2009:67, 81-82; Candotti and Pontillo 2019:31).

[7] Endo and Exo in (8) stand for 'endocentric' and 'exocentric', respectively. We shall only consider semantic endo-/ exocentricity here: given the compound schema "$[X\ Y]_Z$", a compound Z is endocentric if and only if Z is a hyponym of either X or Y; conversely, Z is exocentric if and only if Z is not a hyponym of either X or Y. See Scalise and Fábregas (2010:111) and Bauer (2017:37, 64-65).

not involve any part-whole relation between the external referent and a bahuvrīhi-member: e.g., *kṛtá-brahman-* (ṚV 2.25.1) '(he) by whom the sacred formulation was created' (see Pontillo 2021:507-509 for discussion). Accordingly, we shall propose a working definition of bahuvrīhis modelled on the relative clause analysis of the Indian grammatical tradition (Cardona 1997:220; Lowe 2015:74). Taking again "[X Y]z" as our starting point, we define bahuvrīhis as those compounds Z whose paraphrasis involves a relative clause such that:[8]

(9) i. the relative pronoun is canonically inflected in a Case other than nominative;
     ii. and the functions of subject and predicate are canonically fulfilled, in the relative clause, by the same lexical categories that serve as the internal members of Z.

For instance, *ugra-putra-* of (1) is paraphrased as in (10). Here the relative pronoun is inflected in the genitive, in compliance with (9i). Moreover, the internal members of *ugra-putra-*, namely *putrá-* and *ugrá-*, serve as the subject and the predicate of the relative clause in (10): this satisfies condition (9ii). Since both conditions in (9) are satisfied, *ugra-putra-* qualifies as a bahuvrīhi.

(10) ugrāḥ   putrā   yasyā   aditiḥ.
      mighty.NOM.PL.M   son.NOM.PL.M   REL.GEN.SG.F   Aditi.NOM.SG.F
      'Aditi whose sons are mighty'.

The same remarks apply to *vṛṣaṇ-aśvá-* ((5)), *vṛtrá-putra-* ((6)), and *vájra-hasta-* ((7)). Let us briefly examine these compounds.

     vṛṣaṇ-aśvá- denotes a possessor of bull-like horses; this possessor is identified as the Maruts' chariot (*rátha-*) in ṚV 8.20.10.[9] Thus, *vṛṣaṇ-aśvá-* may be paraphrased as in (11). Since the relative pronoun is inflected in the genitive in (11), (9i) is satisfied. Since *áśva-* 'horse' and *vṛ́ṣan-* 'bull' are the internal members of *vṛṣaṇ-aśvá-* but also the subject and the predicate of the relative clause in (11), (9ii) is also satisfied. Accordingly, *vṛṣaṇ-aśvá-* qualifies as a bahuvrīhi.

(11) vṛṣaṇo   'śvā   yasya   rathaḥ.
      bull.NOM.PL.M   horse.NOM.PL.M   REL.GEN.SG.M   chariot.NOM.SG.M
      'The chariot whose horses are bulls'.

     Turning now to *vṛtrá-putra-*, this compound denotes the possessor—qua mother—of a son whose name is Vṛtra; this possessor is identified as Dānu in ṚV 1.32.9.[10] This allows us to paraphrase *vṛtrá-putra-* as in (12). Since the relative pronoun is inflected in the genitive in (12), (9i) is met. Furthermore, (9ii) is also met because the internal members of *vṛtrá-putra-*, namely *vṛtrá-* 'Vṛtra' and *putrá-* 'son', serve as the subject and the predicate of the relative clause in (12). Consequently, *vṛtrá-putra-* qualifies as a bahuvrīhi.

(12) putro   vṛtro   yasyā   dānuḥ.
      son.NOM.SG.M   Vṛtra.NOM.SG.M   REL.GEN.SG.F   Dānu.NOM.SG.F
      'Dānu of whom Vṛtra is the son'.

---

     [8] The 'paraphrasis' of a compound corresponds to the vigraha of the Indian grammatical tradition, i.e., the combination of inflected words into which the compound is solved or analysed. The first formulation of (9i) is due to Kātyāyana's *Vārttika* 19 *ad* A 2.2.24 (see also Benfey 1852:273). Pāṇini's definition of bahuvrīhis, contained in A.2.2.23-24, differs from the definition adopted in the text. See Candotti and Pontillo (2019:33-36); (2022) for an illustration of Pāṇini's model of bahuvrīhis, and Pontillo (2021) for a critical assessment of how that model was received by modern scholarship.

     [9] *vṛ́ṣanaśvéna maruto vṛ́ṣapsunā/ ráthena vṛ́ṣanābhinā/ ā́ śyenáso ná pakṣíno vṛ́thā naro/ havyā́ no vītáye gata*// 'O Maruts, with your chariot with its bullish horses, bullish breath, bullish wheel-naves, come here at will, like winged falcons, to pursue our oblations, o men.' (tr. Jamison and Brereton 2014).

     [10] *nīcā́vayā abhavad vṛtráputra/ índro asyā áva vádhar jabhāra/ úttarā sū́r ádharaḥ putrá āsīd/ dā́nuḥ śaye sahávatsā ná dhenúḥ*// 'The strength of Vṛtra's mother ebbed; Indra bore his weapon down upon her. The mother was above; the son below: Dānu lies like a milk-cow with her calf.' (tr. Jamison and Brereton 2014).

Likewise, *vájra-hasta-* denotes the (inalienable) possessor of a mace-holding hand; such a possessor is identified as Indra in ṚV 1.173.10.[11] Thus, the paraphrasis in (13) properly captures the meaning of *vájra-hasta-*. The relative pronoun is inflected in the genitive in (13), in compliance with (9i). In addition, *vájra-* 'mace' and *hásta-* 'hand', namely the internal members of *vájra-hasta-*, serve as the subject and the predicate of *vájra-hasta-* in (13), so that (9ii) is also complied with. Since both conditions in (9) are satisfied, *vájra-hasta-* qualifies as a bahuvrīhi.

(13)   haste   vajro   yasyendraḥ.
       hand.LOC.SG.M   mace.NOM.SG.M   REL.GEN.SG.M_Indra.NOM.SG.M

   'Indra in whose hand is a mace'.

Besides having the defining characteristics in (9), Sanskrit bahuvrīhis also display adjectival agreement (i.e., agreement in Case, gender, and number) with their external referent; moreover, Sanskrit bahuvrīhis show a peculiar accentuation which distinguishes them from other compound types. We examine the accentuation of bahuvrīhis in the next subsection, while we address the agreement facts in §3.5 below.

2.2   The accent of bahuvrīhis

The fundamental rule on bahuvrīhi accentuation was formulated by Pāṇini in rule A 6.2.1: *bahuvrīhau prakṛtyā pūrvapadam* 'the left-hand member retains its original accent in bahuvrīhis' (see, e.g., Wackernagel 1905:291; Katre 1987; Cardona 1997:385-386). This rule is meant to capture the fact that, when Z of "[X Y]$_Z$" is a bahuvrīhi, then Z is typically accented on the left-hand member X, more precisely on the same syllable on which X is accented when X occurs outside Z (Macdonell 1910:92). For illustration, consider *rāja-putra-*, formed from *rā́jan-* 'king' and *putrá-* 'son(s)'. When this compound is endocentric, such as in ṚV 10.40.3, it shows up as *rāja-putrá-* 'the king's two sons' (referring to the divine Aśvin twins); conversely, when this compound is a bahuvrīhi, such as in ṚV 2.27.7, it shows up as *rā́ja-putra-* '[Aditi] whose sons are kings' (see Candotti and Pontillo 2022:10-11, 19). The reader may easily verify that the bahuvrīhis *vṛtrá-putra-* ((6)) and *vájra-hasta-* ((7)) comply with A 6.2.1. As for the accent of *vṛ́ṣaṇ-aśvá-* ((5)) and *ugra-putra-* ((4)), see below in this subsection and fn. 1 above.

Indeed, A 6.2.1 allows for several exceptions: on the one hand, many bahuvrīhis are accented on the right-hand member and, on the other hand, many endocentric (hence, non-bahuvrīhi) compounds are accented on the left-hand member. The list of rules formulated by Pāṇini to capture these exceptions is reported in Cardona (1997:385-391). A full discussion of these exceptions will not be attempted here; the interested reader is referred to the standard treatments of accent in Sanskrit compounds, notably Whitney (1899:480-515), Wackernagel (1905), and Macdonell (1910:91-97). Here we limit ourselves to mentioning a couple of exceptions that are relevant for the examples quoted in §4.3 below as well as for *vṛ́ṣaṇ-aśvá-* ((5)).

When the left-hand member of a bahuvrīhi is *víśva-* 'all', that bahuvrīhi is accented on the second syllable of *víśva-* (Wackernagel 1905:292; Macdonell 1910:92): e.g., *viśvá-bhānu-* (ṚV 4.1.3) 'whose brightness is on everything', 'all-radiant'. When the left-hand member of a bahuvrīhi is instead *dúṣ-* or *sú-* 'well', that bahuvrīhi is accented on the right-hand member, typically on the same syllable on which the right-hand member is accented outside compounds (Whitney 1899:508-509; Wackernagel 1905:293-295; Macdonell 1910:93): e.g., *su-bhága-* (ṚV 1.86.7) 'whose portion is good', 'well-portioned' (cf. *bhága-* 'portion'). When a compositional suffix (also known as *samāsānta* suffix) is attached to a bahuvrīhi, the bahuvrīhi's accent may be on the suffix, in which case the suffix is said to be dominant (Whitney 1899:505; Macdonell 1910:86-87; Kiparsky 2010:32): e.g., *śiti-kakṣ-ín-* (TS 5.5.20.1; from *śiti-* + *kákṣa-* + *-ín-*) 'whose belly is white' (cf. *śiti-*

---

[11] *víspardhaso narā́ṃ ná śáṃsair/ asmā́kāsad <u>indaro</u> <u>vájrahastaḥ</u>/ mitrāyúvo ná pū́rpatiṃ súśiṣṭau/ madhyāyúva úpa śikṣanti yajñaíḥ//* 'The contenders—(crying) "Indra with mace in hand will be ours!"— seeking (the one in) the middle [=Indra], try to win him over with their sacrifices, as those seeking an ally, (approaching) a lord of strongholds in regard to his good command, (try to win him over) with lauds of men.' (tr. Jamison and Brereton 2014).

kákṣa- 'id.' in AVŚ 5.23.5);[12] *vārdhrā-ṇas-á-* (TS 5.5.20.1; from *vārdhra-* + *nás-* + *-á-*) 'whose nose is leathern' (cf. *ṛjū-nas-* lit. 'whose nose is straight' in RV 8.52.2). Finally, the accent of *vṛṣan-aśvá-* does not coincide with either the accent of *vṛ́ṣan-* or that of *áśva-*: see Whitney (1899:506); Wackernagel (1905:298).[13] *vṛṣan-aśvá-* may perhaps be treated along the lines of *tri-bandhú-* (RV 7.37.7) 'whose relations are triple' in Kiparsky's (2010) model of compound accentuation (see especially ibid., p. 32): first, the bahuvrīhi stem (**trí-bándhu-*; **vṛ́ṣan-áśva-*) combines with a dominant zero compositional suffix, which deletes the original accent of the bahuvrīhi-members (**tri-bandhu-Ø*; **vṛṣan-aśva-Ø*); next the Oxytone Rule assigns default final accent to the bahuvrīhi stem (*tri-bandhú-Ø = tri-bandhú-*; *vṛṣan-aśvá-Ø = vṛṣan-aśvá-*).[14]

Earlier we considered the question as to whether the possessive relation perceived in some Sanskrit bahuvrīhis is established via The Elided Possessive Unit Strategy (3a) or The Movement Strategy (3b). We now construct an argument which makes it possible to settle this question on empirical grounds.

2.3 The Elided Possessive Unit Strategy vs. The Movement Strategy

In analyses of bahuvrīhis that resort to The Elided Possessive Unit Strategy (3a), the elided possessive unit B expressing the meaning 'having' combines with a pre-formed compound-stem CS: e.g., B would combine with *ugra-putra-* as a whole in (1). Consider the possibility in which CS is the stem of an endocentric compound. This possibility is advocated by Gillon (2008:3) for at least one class of bahuvrīhis: so-called samāna-adhikaraṇa-bahuvrīhis, in which an attributive relation holds between the bahuvrīhi-members. He assumes that a (samāna-adhikaraṇa-)bahuvrīhi is the combination of an endocentric compound with B (cf. also Bopp 1827:357; Whitney 1889:501-502; Marchand 1967:335; Kiparsky 1982:139). E.g., by combining *śiti-* 'white' with *kákṣa-* 'lurking place', 'belly', we obtain the endocentric compound *śiti-kakṣa-* which expresses the meaning 'white belly'.[15] Then, by attaching B to the stem of the endocentric compound *śiti-kakṣa-*, the omophonous bahuvrīhi *śiti-kákṣa-* (= *śiti-kakṣa*-B) is formed which expresses the meaning 'having a white belly' (AVŚ 5.23.5).[16]

The Elided Possessive Unit Strategy looks particularly plausible in the light of the existence of *śiti-kakṣín-* (TS 5.5.20.1). *śiti-kakṣín-* is synonymous with *śiti-kákṣa-* insofar as *śiti-kakṣín-* and *śiti-kákṣa-* both denote an entity having a white belly in the texts quoted here: this entity is identified as the worms in the case of *śiti-kákṣa-* but as a bird species in the case of *śiti-kakṣín-*.

Now that we have established the synonymity between *śiti-kákṣa-* and *śiti-kakṣín-*, consider that - *ín* - is a possessive suffix which attaches to a nominal X to form a derivative nominal with the meaning 'endowed with X': cf. *rátha-* 'chariot' + *-ín-* 'possessing' > *rathín-* 'possessing a chariot', 'charioteer'. One may then take the B occurring in such bahuvrīhis as *śiti-kákṣa-* as the elided counterpart to the *-ín-* of *śiti-kakṣín-*: this would straightforwardly account for the synonymity between *śiti-kákṣa-* and *śiti-kakṣín-*, and for the establishment of the possessive relation between the external referent of the bahuvrīhi and one of the bahuvrīhi-members. Thus, the postulation of the elided possessive unit B by the Elided Possessive Unit Strategy (3a) appears to be well-motivated.

This is not the end of the story, however. We wonder: can the Elided Possessive Unit Strategy (3a) predict the internal orders in (4)-(7)? To repeat, the Elided Possessive Unit Strategy takes it that

---

[12] See §2.3 below on the synonymity of *śiti-kakṣ-ín-* and *śiti-kákṣa-*. The accent of *śiti-kákṣa-* also deviates from the general rule A 6.2.1 for bahuvrīhi accentuation. Indeed, *śiti-kákṣa-* belongs to the set of bahuvrīhis which have a disyllabic adjective ending in *-i* or *-u* as their left-hand member: these bahuvrīhis are commonly accented on the same syllable on which the right-hand member is accented outside compounds (Wackernagel 1905:296-297; Macdonell 1910:92-93; Kiparsky 2010:31-32).

[13] The accent required by A 6.2.1, i.e., *vṛ́ṣan-aśva-*, is attested in MS 2.5.5 and ŚB 3.3.4.18.

[14] The accentuation of compounds may be used to diagnose their internal constituency (Kiparsky 2010:34; 2016:16-17). We shall not pursue this interesting line of analysis here.

[15] Although *śiti-kakṣa-* is not actually attested as endocentric compound, we assume, for the sake of argument, that this form was nonetheless possible in principle.

[16] Lowe (2015:103) similarly assumes that (samāna-adhikaraṇa-)bahuvrīhis are the combination of an endocentric compound with a special unit, but such a unit does not bear any possessive meaning in his account, and only serves the purpose of endowing bahuvrīhis with adjectival agreement properties (see §3.5 below). For an overview of recent analyses of bahuvrīhis, see Candotti and Pontillo (2022:5-9).

the elided possessive unit B attaches to the stem of an endocentric compound to form a bahuvrīhi. But B does not affect the internal order of the compound-stem to which B attaches. This allows us to make a clear empirical prediction on the internal order of bahuvrīhis:

(14) The internal order of a bahuvrīhi will be always the same as the internal order of the endocentric compound corresponding to that bahuvrīhi (the endocentric compound corresponding to a bahuvrīhi is the endocentric compound to whose stem B attaches to form that bahuvrīhi).

The prediction in (14) is highly problematic, however. For one thing, none of the compounds in (4), (6), and (7) (i.e., *ugra-putra-*, *vṛtra-putra-*, *vajra-hasta-*) is attested as endocentric; in point of fact, it is very rare that, if a form is attested as an endocentric compound, then that form is also attested as a bahuvrīhi (and vice versa)—see Renou (1961:114). Therefore, the accuracy of (14) cannot be tested against the internal orders of (4), (6), and (7). Furthermore, *vṛṣaṇ-aśvá-* ((5)) does have an endocentric counterpart, namely *aśva-vṛṣá-*. Indeed, direct evidence for the endocentricity of *aśva-vṛṣá-* comes from the context in which this form occurs, i.e., ŚB 14.4.2.7-8 (= BĀU 1.4.4):[17] *aśva-vṛṣá-* here denotes the masculine analogue of *váḍavā-* 'mare', hence conveys the meaning 'male horse'; the meaning 'male horse' is in turn a more specific version of the meaning 'bull-like horse', on the grounds that a bull-like horse is a horse bearing the properties typical of bulls, and that such properties include masculinity in the Vedic-speaking world. Thus, *aśva-vṛṣá-* properly conveys the meaning 'bull-like horse' in the passage at stake. Since *aśva-vṛṣá-* conveys the meaning 'bull-like horse', *aśva-vṛṣá-* is hyponym of *aśva-* 'horse', which means that *aśva-vṛṣá-* is endocentric. Recall now that the bahuvrīhi *vṛṣaṇ-aśvá-* conveys the meaning 'having bull-like horses': it is then reasonable to take *aśva-vṛṣá-* as the endocentric compound stem to which the elided possessive unit B (conveying the meaning 'having') attaches to form the bahuvrīhi *vṛṣaṇ-aśvá-*. But in accordance with prediction (14), the internal order of *vṛṣaṇ-aśvá-* should match the internal order of *aśva-vṛṣá-*, contrary to fact.

Therefore, (14) is either wrong or not testable when the internal orders in (4)-(7) are at stake. But (14) is all that the Elided Possessive Unit Strategy has to say about the internal order of bahuvrīhis. We conclude that the Elided Possessive Unit Strategy fails to predict the internal orders in (4)-(7). The question is then whether The Movement Strategy (3b) succeeds in predicting these internal orders. To tackle this question, let us go back to (4)-(7), now focusing on the specification of the syntactic functions fulfilled by the bahuvrīhi-members.[18]

Interestingly, the position of the bahuvrīhi-internal predicate in (4)-(7) depends on whether the bahuvrīhi's external referent is construed with that predicate: the predicate precedes the subject when the bahuvrīhi's external referent is construed with (i.e., is interpreted as the Possessor of) the subject (as in (4)-(5)) but not when this external referent is construed with the predicate (as in (5)-(6)).[19] We may thus express the restrictions on the internal orders in (4)-(7) by means of the following generalisation:

(15) i. In a bahuvrīhi in which the predicate is nominal (i.e., an adjective or a noun) and the external referent is interpreted as the Possessor of the <u>subject</u>, the predicate must **precede** the subject;

ii. In a bahuvrīhi in which the predicate is nominal (i.e., an adjective or a noun) and the external referent is interpreted as the Possessor of the <u>predicate</u>, the predicate must **follow** the subject.

---

[17] *sā gaurábhavat vṛṣabha ítarastāṃ sámevàbhavattáto gā́vo 'jāyanta// váḍavétarábhavat <u>aśvavṛṣa</u> ítaro gardabhī́tarā gardabha ítarastāṃ sámevàbhavattáta ékaśaphamajāyata.* 'She became a cow, the other became a bull and was united with her; from that cows were born. The one became a mare, the other a stallion; the one became a she-ass, the other be-came a he-ass and was united with her; from that one-hoofed animals were born.' (tr. Mādhavānanda 1950).

[18] By saying that an expression α is 'Possessor of Subject' or 'Possessor of Predicate' (see (4)-(7) above), we mean that α is interpreted as the possessor of the entity denoted by another expression β, where β fulfills the function of Subject or Predicate.

[19] Note that the restriction that compound-external material can only be construed with (or 'view') the right-hand member of a compound (see Di Sciullo and Williams's model 1987, especially their pp. 24, 30-31) does not hold in Sanskrit: see Gillon (2008); Molina-Muñoz (2013); Lowe (2015:76-77); Candotti and Pontillo (2022:18). Thus, one cannot appeal to this restriction to account for the different positioning of the bahuvrīhi-internal predicate in (4)-(5) vs. (6)-(7).

This way, the question as to whether the Movement Strategy predicts the internal orders in (4)-(7) reduces to the question: can the Movement Strategy account for generalisation (15)? §4 is devoted to providing an affirmative answer to the latter question.

Before concluding this section, we should like to explore a third means of establishing the possessive relation in bahuvrīhis. We shall see that, albeit theoretically plausible, this third possibility faces the same empirical problems as the Elided Possessive Unit Strategy.

2.4 Bahuvrīhis and figures of speech

The Elided Possessive Unit Strategy (3a) and the Movement Strategy (3b) are not in fact the only ways in which the possessive relation between the external referent of the bahuvrīhi and one of the bahuvrīhi-members may be established. Indeed, Bauer (2008:59) (see also Bauer 2016:462-463; 2017:65; cf. Pennanen 1982:245-246; Barcelona 2008:210; Lundquist and Yates 2018:2120 for similar views) proposes that English bahuvrīhis such as *red-cap* (which names people who are abitually associated with a red cap) are regular endocentric compounds interpreted by synecdoche (the 'pars pro toto' figure of speech). For instance, the simplex *crown* properly denotes a headdress possessed by, or being part of (⊆), a monarch; by resorting to the synecdoche, one may use *crown* (the part) as a name for *monarch* (the whole). Likewise, the compound *red-cap* properly denotes a kind of head covering possessed by, or being part of (⊆), specific groups of people, typically the military policemen in the UK or the railway porters in the US; as such, *red-cap* is an endocentric compound because it is hyponym of the compound-member *cap*. By synecdoche, however, *red-cap* (the part) names those very groups of people that possess the red head covering (the whole). The synecdoche thus turns the groups of people possessing the red head covering into the external referent of *red-cap*: this way, a possessive relation is established between the external referent of *red-cap* and the head covering denoted by *cap*.

Although this may well be the correct analysis of bahuvrīhis in English (and possibly other languages considered by Bauer 2008:57-60), it cannot be extended to Sanskrit. Indeed, Bauer's synecdoche-based analysis and the Elided Possessive Unit Strategy (3a) alike commit themselves to the prediction in (14). Consider why. In the synecdoche-based analysis, bahuvrīhis are obtained by applying the synecdoche to their corresponding endocentric counterparts; but the synecdoche (much like the elided possessive unit B) is unable to affect the internal order of the endocentric compound to which it applies; accordingly, the internal order of a bahuvrīhi must be the same as the internal order of the endocentric compound corresponding to that bahuvrīhi ((14)).

As already seen, this prediction is either wrong or not testable when it comes to the internal orders in (4)-(7) (see §2.3 above).[20] Given that Bauer's synecdoche-based analysis makes the same predictions on the internal orders in (4)-(7) as the Elided Possessive Unit Strategy (3a), we shall disregard Bauer's analysis in what follows. However, whenever we claim that the Elided Possessive Unit Strategy (3a) fails to predict the internal order of Sanskrit compounds, this criticism can be extended to Bauer's analysis.

3 For a syntactic derivation of bahuvrīhis

In this section we report some evidence that justifies a syntactic derivation for Sanskrit bahuvrīhis. Specifically, we argue that the internal members of Sanskrit bahuvrīhis are phrases, and as such can undergo ordinary syntactic operations such as movement.

3.1 Sanskrit compound-members as words

Lowe (2015) provides several arguments to the effect that both members of Sanskrit compounds are words. These arguments can be divided into two families.

---

[20] An additional reason against extending the synecdoche-based analysis to Sanskrit bahuvrīhis is that the accent of bahuvrīhis typically differs from the accent of endocentric compounds in Sanskrit (see §2.2 above). It is not clear to us how the application of the synecdoche to an endocentric compound may effect this difference in compound accentuation (see Candotti and Pontillo 2022:7).

The first family is targeted at showing the word status of the left-hand member: more precisely, the left-hand member is claimed to be a 'non-projecting word' (i.e., a word of a special sort which does not project either complement or specifier—ibid. p. 91), but nonetheless a word. This first family consists of several arguments, but we only mention two of them here, drawing our data from Vedic Sanskrit (see ibid., pp. 75-86, for the complete list of these arguments).[21]

3.2 Sanskrit compounds are not anaphoric islands

The first argument of the first family capitalises on the fact that the left-hand member of Sanskrit compounds may be a pronoun that is coreferential with a compound-external expression. (16) illustrates this fact.

(16) | tád$_i$-annāya | tád$_i$-apase/ | tám |
|---|---|---|
| this$_i$-food.DAT.SG.M | this$_i$-work.DAT.SG.M | this.ACC.SG.M |
| bhāgám | upasedúṣe/ | tritắya |
| portion.ACC.SG.M | approach.PRF.PTCP.DAT.SG.M | Trita.DAT.SG.M |
| ca | dvitắya | *ca*/ |
| and | Dvita.DAT.SG.M | and |
| *úṣ*o | duṣvápn*iyam*$_i$ | vah*a*/ |
| dawn.VOC.F.SG | bad_dream.ACC.SG.N$_I$ | carry.IMP.2SG |

'To the one who has it as his food and as his work, to the one reverently approaching it as his portion, to Trita and to Dvita, o Dawn, carry the bad dream.' (ṚV 8.47.16a-d; tr. Jamison and Brereton 2014).

The demonstrative pronoun *tád-* 'this' figures as the left-hand member of the two bahuvrīhis *tád-anna-* 'he of whom this is the food' and *tád-apas-* 'he of whom this is the work': *tád-* here is coreferential with *duṣvápnya-* 'bad dream', which shows up in this same verse, as well as in ṚV 8.47.14 (see Jamison and Brereton 2014:1126-1127).

Thus, Sanskrit compounds can contain pronouns, and these compound-internal pronouns are coreferential with an expression placed outside the compound itself. But this is tantamount to saying that Sanskrit compounds are not anaphoric islands (Lowe 2015:78; see also Kiparsky 2009:83; Molina-Muñoz 2013:191-192). English compounds appear to obey different constraints:

(17) a. Whenever he meets an [old woman]$_i$ Pete says he hates young ones$_i$.
     b. *Whenever he meets an [old woman]$_i$ Pete says he is a one$_i$-hater.

(Postal 1969:216)

In (17b) the pronominal element *one*, which is the left-member of the compound *one-hater*, cannot be coreferential with *old woman*, which occurs outside the compound itself (cf. the well-formedness of (17a), in which *ones* is not part of a compound). This shows that a pronominal element contained within a certain compound cannot be coreferential with an antecedent placed outside that compound in English: to wit, English compounds, unlike Sanskrit compounds, are anaphoric islands. Now, anaphoric islandhood is considered a defining property of 'word'. Indeed, it has been assumed since Postal (1969:227) that words are anaphoric islands crosslinguistically (see Lowe 2015:77 and the references cited therein).[22] This is illustrated by the following contrast:

(18) a. I got a divorce$_i$ from Sally and I'll get one$_i$ from Louise too.
     b. *I [divorce]$_i$-d] Sally and I'll get one$_i$ from Louise too.

(Postal 1969:210)

---

[21] Indeed, Lowe (2015:107-108) confines the claim that the members of Sanskrit compounds are words to Classical Sanskrit, a standardised form of Sanskrit codified on the basis of Pāṇini's Aṣṭādhyāyī (4th c. BCE). However, to the extent that his arguments can still apply to Vedic Sanskrit compounds, we consider it reasonable to extend that claim to Vedic Sanskrit compounds, too.

[22] The claim that words and compounds are anaphoric islands in English is indeed an oversimplification and needs to be remodulated in the light of the counterexamples provided by Bauer (2017:20-21).

In (18b) *one* cannot be coreferential with the antecedent noun *divorce* that is contained within (i.e., is part of) the verb *divorced* (a denominal derivative verb formed from the noun *divorce*), *divorced* being a word. However, coreference between *one* and *divorce* becomes possible when the noun *divorce* is itself a word not contained within any other word, as in (18a). Thus, the ill-formedness of (17b) follows on the assumption that English compounds are words, and hence that compound-members are subparts of words in English (see Postal 1969:225-227). But then, Lowe (2015:80) argues, the well-formedness of (16) can be traced back to the fact that the left-hand member of Sanskrit compounds is a word, rather than a subpart of word. Note that this argument is also consistent with the conclusion that the left-hand member of Sanskrit compounds (e.g., *tád-* in the compound *tád-apas-* of (16)) is a phrase that does not contain complements or specifiers.

3.3 Sanskrit compounds and clausal scope

The second argument of the first family is based on the following fact: the left-hand member of Sanskrit compounds can be an interrogative pronoun. Thus, *kád-* is the stem form of the interrogative pronoun (*káḥ, kā́, kád/kím*), and shows up as the left-hand member of the bahuvrīhi *kád-artha-* (lit. 'whose purpose is what') in the following example:

(19)
ádha	gmánt*ā*	*u*śánā	pr̥chate
PCL	come.AOR.PTCP.ACC.DU.M	Uśanā.NOM	ask.IND.PRS.3SG
vām/	kád-arthā	na	ā́
you.ACC.DU	what-purpose.NOM.DU.M	we.GEN	to
gr̥hám/	ā́	jagmathuḥ	parākā́d/
house.ACC	to	come.IND.PRF.2DU	afar.ABL
diváś	ca	gmáś	ca
heaven.ABL	and	earth.ABL	and
mártiyam//			
mortal.ACC			

'Then Uśanā asks you both [=Indra and Kutsa] on your coming, "With what purpose (have you come) to our house? You have come from afar, from heaven and earth, to a mortal."' (R̥V 10.22.6; tr. Jamison and Brereton 2014).

kád- turns the sentence whose predicate is *jagmathuḥ* into an interogative sentence ('With what purpose have you come to our house?') in (19); technically, *kád-* takes clausal scope (see Kiparsky 2009:83; Lowe 2015:80). Since words and phrases—but not subparts of words—are known to have the ability to take clausal scope in Sanskrit (much like in English), *kád-*, which displays this ability in (19), should also qualify as a word. Accordingly, the left-hand member (*kád-*) of the bahuvrīhi *kád-artha-* is a word, rather than a subpart of word. In this case, too, the argument would stand even if we were to take *kád-* as a phrase (which does not contain complements or specifiers).

3.4 Word or phrase?

We have reviewed two of Lowe's (2015) arguments to the effect that the left-hand member of Sanskrit compounds is a word (for other arguments, see Lowe 2015:80-86). We have also noted that these arguments are consistent with an alternative conclusion: that Sanskrit compounds' left-hand member is a phrase rather than a word. On this alternative conclusion, the stem *áśva-* occurring in the compound *aśva-śaphá-* 'horse's hoof' would be represented as a bare phrase (i.e., a phrase which does not contain complements or specifiers) in the phrase marker.

Let us then assume that a compound's left-hand member that is morphologically a stem qualifies as a bare phrase from the syntactic point of view. We identify the head of such a phrase with the stem itself: thus, the structure of the *áśva-* contained within *aśva-śaphá-* would be [NP [N° aśva-]], which amounts to a bare phrase. However, we still need to explain why the phrase corresponding to the compound's left-hand member does not contain complements or specifiers and lacks inflection. According to Lowe (2015:91-93), the lack of inflection of a certain compound-member and its inability to contain complements or specifiers both follow from this compound-member being a non-projecting word. An alternative explanation that does not appeal to the notion of non-projecting

word may be the following. It is standardly assumed that fully inflected words are endowed with some functional projection above the lexical level (e.g., above N[oun] P[hrase]/ A[djectival] P[hrase]/ V[erb] P[hrase]).[23] Thus, a natural way to capture the absence of inflection in Sanskrit compounds' left-hand members is to say that these members only consist of the projection headed by the lexical head (N°/ A°/ V°/ P°), lacking instead the functional layers stacked above such a head. We return to this point in §4.2, where the inability of a compound-member to take complements and specifiers is also traced back to the fact that it lacks functional layers.

3.5 The right-hand member of Sanskrit compounds

Let us now turn to the second family of arguments, i.e., the arguments targeted at showing the word-status of Sanskrit compounds' right-hand member.

Lowe (2015:72, 93, 97-99) analyses Sanskrit endocentric compounds such as *aśvaśapháḥ* 'horse's hoof' as [aśva-[śapha-ḥ]], where the nominal ending -*ḥ* belongs to the right-hand member *śaphá*-. But this means that the right-hand member of endocentric compounds (e.g., *śapháḥ* of *aśvaśapháḥ*) is formally identical to a full-fledged inflected word which occurs outside compounding (e.g., *śapháḥ* is also attested as an independent word in MS 2.5.6.6). Then, the formal identity of Sanskrit endocentric compounds' right-hand member with full-fledged inflected words constitutes crucial evidence, in Lowe's analysis, for considering this right-hand member as a word rather than as a subpart of word. Things get more complicated when it comes to bahuvrīhis. Indeed, the right-hand member of a bahuvrīhi is not formally identical to an independent word:

> nouns have inherent grammatical gender in Sanskrit, but at the end of a bahuvrīhi a noun can be inflected in any gender, since it must agree with the compound's external referent. So an inherently masculine noun, for example, which cannot otherwise appear in neuter or feminine forms, can appear in such forms at the end of a bahuvrīhi. Therefore a noun form used at the end of a bahuvrīhi can be considered a rather different type of word; it is not, in fact, a noun of the standard type. It is a noun with adjectival agreement properties or, to put it another way, a noun that is partly adjectival. (Lowe 2015:103)

In fact, Lowe (2015:103-105) analyses a bahuvrīhi like *ugraputraḥ* '(he/ she) whose sons are mighty' ((4)) as having the following structure:

(20) i. Adj° → N̂ Adj°
     ii. N̂ → Âd̂ʝ N̂

Adj° attaches to the whole compound structure, which is dominated by N̂. The stem form *ugrá*- instantiates Âd̂ʝ of (20ii), while *putraḥ* instantiates two nodes by the mechanism of Lexical Sharing: the N̂ daughter of (20ii), as well as the Adj° daughter of (20i) (see ibid., pp. 104-105, for the technical details of this analysis). Therefore, according to this analysis the right-hand member of bahuvrīhis is a complex unit—constituted of a non-projecting word (N̂) and a word (Adj°)—which bears a special kind of inflection (i.e., adjectival inflection). Since the presence of inflection is a defining characteristic of words, the right-hand member of bahuvrīhis qualifies as a word.

However, we cannot immediately adopt Lowe's conclusion in this case. Indeed, we analyse bahuvrīhis such as *ugraputraḥ* as [[ugra-putra-]-ḥ], where both *ugrá*- and *putrá*- are in stem form, and the nominal ending -*ḥ* is affixed to the whole compound stem [ugra-putra-], rather than to the right-hand member only. In this way, -*ḥ* does not qualify as the inflection of the right-hand member, which thus results as being morphologically identical to the left-hand member.[24] This means that we can no longer appeal to the presence of inflection on the right-hand member of bahuvrīhis to justify the word status of this right-hand member. Nonetheless, there is still reason to maintain that

---

[23] For instance, nominal and adjectival projections in noncompound syntax may be endowed with the Agr(eement) P(hrase) of Cinque (2005:317-318, 321 n. 24), with the KP ('Case Phrase') of Bittner and Hale (1996), or, more generally, with a D(eterminer) P(hrase) (decomposable as in Zamparelli 2000). Verbal projections are endowed with T(ense) P(hrase), and C(omplementiser) P(hrase) at least, TP and CP being further decomposable along the lines of Cartography (see Rizzi 2013).

[24] This analysis of *ugraputraḥ* as [[ugra-putra-]-ḥ] is reminiscent of Pāṇini's analysis of compounds. See the references of n. 6 above.

a bahuvrīhi's right-hand member is a word or a phrase, and not a subpart of word.

Perfect passive participles (PPP, also known as *-ta-* adjectives) typically occur as left-hand members in Sanskrit bahuvrīhis. Thus, we find: *jāta-danta-* (VaikhGS 7.2) '(the boy) whose teeth have grown', where *jātá-* 'grown' is a PPP; *ā́hitāgni-* (TS 2.2.2.2) '(he) by whom the sacred fire has been established', where *ā́hita-* 'established' is a PPP; *hata-putra-* (KB 4.6.7) '(Vasiṣṭha) whose sons were killed', where *hatá-* 'killed' is a PPP. These bahuvrīhis fall outside the scope of generalisation (15) because they involve a verbal predicate (i.e., the PPP itself); nonetheless, they are very interesting in that their internal order may be inverted without any apparent change in meaning. Thus, the following bahuvrīhis are also attested, with the PPP occurring as the right-hand member of the compound: *danta-jāta-* (VaikhGS 5.9) '(the boy) whose teeth have grown'; *agny-āhita-* (MBh. 12.281.21) 'he by whom the sacred fire has been established';[25] *putra-hata-* (PB 4.7.3) '(Vasiṣṭha) whose sons were killed'—more examples of bahuvrīhis having a PPP as their right-hand member are given in Wackernagel (1905:302-303).

Now, if the left-hand member of a Sanskrit compound is to be considered a word or a phrase for the reasons spelled out above (see §§3.1-4), then the PPP occurring as the left-hand member of the bahuvrīhis *jāta-danta-*, *ā́hitāgni-*, and *hata-putra-* also counts as a word or a phrase. We have seen that the PPP can optionally occupy the right-hand slot in these bahuvrīhis (thereby yielding *danta-jāta-*, *agny-āhita-*, and *putra-hata-*): this should be impossible if these bahuvrīhis' right-hand slot were smaller than a word or a phrase. Therefore, these bahuvrīhis' right-hand slot must be at least as big as a word or a phrase. In keeping with the preceding discussion, we conclude that both the left-hand member and the right-hand member of the bahuvrīhis *jāta-danta-*, *ā́hitāgni-*, and *hata-putra-* (as well as of *danta-jāta-*, *agny-āhita-*, and *putra-hata-*) is a phrase.

This argument cannot be easily reproduced for proving the phrasal status of the bahuvrīhis in (4)-(7): the linear order of *ugra-putra-*, *vṛṣaṇ-aśvá-*, *vṛtrá-putra-*, and *vájra-hasta-* cannot be inverted (*putrogra-*, *aśva-vṛṣan-*, *putra-vṛtra-*, and *hasta-vajra-* are all unattested as bahuvrīhis). Nevertheless, the non-invertibility of the linear order of these bahuvrīhis will be shown to follow from independent reasons in §4. We may then assume that the right-hand member of the bahuvrīhis in (4)-(7) also has phrasal status.

In §3.4 we proposed to explain the stem status of Sanskrit compounds' left-hand member as a result of the fact that this left-hand member is a bare phrase headed by a lexical category and lacking functional structure (e.g., an NP rather than a DP). We have now claimed that, in all Sanskrit bahuvrīhis, both the left-hand member and the right-hand member are stems from the morphological point of view, and phrases from the syntactic point of view. Accordingly, we suggest that the phrases corresponding to the left- and right-hand members of Sanskrit bahuvrīhis are always headed by lexical categories and always lack functional structure.

A brief remark is in order at this point. We have made two important assumptions in this subsection: i. that the right-hand member of the bahuvrīhis in (4)-(7) is a phrase; ii. that the phrases corresponding to the left- and right-hand members of Sanskrit bahuvrīhis are always bare phrases projected by lexical heads and lacking functional structure. As a matter of fact, these assumptions do not follow from any logical necessity: they are tenable only to the extent that they allow decomposing complex empirical facts into simpler abstract entities.[26] In this specific case, these assumptions can be maintained only to the extent that they make it possible to reduce the well-formedness or ill-formedness of Sanskrit bahuvrīhis' internal order to the interaction of independently motivated syntactic constraints. Other assumptions are advanced in §4: they should all be considered in this perspective.

3.6 Summary

Section §3 contained several arguments to the effect that both members of Sanskrit compounds are words or phrases. These arguments capitalise on the fact that a compound-member can enter

---

[25] Although we could not find any earlier attestation of *agny-āhita-*, it is still safe to claim that this form was common at the time when Pāṇini lived (4th c. BCE), i.e., during the late Vedic era, insofar as Pāṇini's rule A 2.2.37 (*vāhitāgnyādiṣu*) alludes to it. See Cardona (1997:223).

[26] As remarked by Perrin (1913) (quoted in Moro 2017:41), the task of the scientist is to "explain what is visible and complicated by means of what is simple and invisible".

typically syntactic relations in Sanskrit (e.g., coreferentiality, clausal scope): only words and phrases are known to be able to enter such syntactic relations in Sanskrit; the left-hand member of Sanskrit compounds enters these relations; hence, Sanskrit compounds' left-hand member is a word or phrase. When independent restrictions are factored out, the left-hand member can invert with the right-hand member in Sanskrit bahuvrīhis, as in *jāta-danta-* vs. *danta-jāta-*: to make sense of this kind of inversion, we have suggested that the right-hand member of Sanskrit bahuvrīhis also has word- or phrasal status. Concerning instead the right-hand member of Sanskrit endocentric (hence, non-bahuvrīhi) compounds, the fact that it is formally identical to a word makes it possible to assign word- or phrasal status to it. We are then left with the following picture: the left- and right-hand members of Sanskrit compounds are words or phrases; consequently, the syntactic operations applying to these compound-members should be comparable to the syntactic operations applying to non-compound words and phrases. In §4 we shall argue that movement is among the syntactic operations which apply to compound-members in Sanskrit. Furthermore, we have assumed that the left-hand member and the right-hand member of Sanskrit bahuvrīhis are not words, but rather bare phrases projected by lexical categories and lacking functional projections: this assumption will allow us to implement the Movement Strategy (3b) in terms of phrasal movement, thus dispensing with word- (i.e., head-)movement.[27]

    We are now ready to explore a syntactic derivation of bahuvrīhis.

4   Movement and bahuvrīhis: The Movement Strategy

In this section we lay out a formal derivation of Sanskrit bahuvrīhis couched within the framework of generative grammar. This derivation incorporates the Movement Strategy (3b), which involves the movement of the possessor from the phrase projected by a compound-member to a position outside the compound. The restrictions on the internal order of bahuvrīhis will be seen to follow from independently motivated constraints on this movement of the possessor.

4.1   The predicative link within Sanskrit bahuvrīhis

Dikken (2006:10-12) proposed that every ascription of a property by a phrase XP to another phrase YP is mediated in syntax by a predicative link. E.g., *the cause of the riot* ascribes the property of having caused the riot to *a picture of the wall* in (21) (Moro 1997:24; 2017:164); therefore, a predicative link must occur between *the cause of the riot* and *a picture of the wall*.

(21) John renders [$_{SC}$ [$_{DP}$ a picture of the wall] [$_{DP}$ the cause of the riot]].

The predicative link is the link between the predicate and the subject-of-predication (Moro 2019:1). The predicative link is codified, at the beginning of the derivation, by a symmetrical structure referred to as SC 'small clause' (Moro 2000:71; 2019:4; Chomsky 2013:42; cf. Bittner and Hale 1996:8-10). The small clause is made up of two components α and β (namely, the subject-of-predication and the predicate) such that: i. both α and β are maximal projections; ii. α c-commands β and vice versa (Moro 2000:27; 2004:§5).[28] But how can we understand which of the two components of the small clause is the predicate and which is the subject-of-predication?

    Indeed, the phrase XP that ascribes properties to another phrase YP within the small clause is the predicate, YP being the subject-of-predication (see Moro 2017:49-52, 164, building on Geach 1980:60, 80; 1987:9-10). Therefore, *the cause of the riot* is the predicate and *a picture of the wall* the subject-of-predication in (21). Now, an alternative way exists of telling subject-of-predication

---

[27] This is reminiscent of Cinque's (2005), (2010) work on adjectives. He shows that a series of generalisations on the ordering and interpretation of adjectives (including Greenberg's Universal 20) are captured by assuming that only phrasal movement is available crosslinguistically, even when the unit that moves up the tree is formally identical to a simple word. See especially Cinque (2005:317 n. 6); (2010:37-41). Cf. also Ntelitheos's (2022:11) proposal that the formation of synthetic compounds in Modern Greek involves phrasal movement, as opposed to head-movement, of the compound's left-hand member.

[28] A node α c-commands a node β iff: i. α does not dominate β; ii. the first node that dominates α also dominates β. Dominance is instead defined as follows: let T be a syntactic tree, let α and β be nodes of T; then α dominates β in T if it is possible to draw a continuous line towards the bottom from α to β.

and predicate apart: the predicate contains (or coincides with) a phrase which 'selects' the subject-of-predication. For an expression E_1 to select another expression E_2 essentially means that E_1 sets up an abstract scenario and specifies the semantic role played by the denotatum of E_2 in that scenario (see Chomsky 1986:86-87; Cinque 1990:41). Thus, *cause of the riot* (i.e., a phrase contained within the DP *the cause of the riot*) selects *a picture of the wall* in (21): *cause of the riot* sets up a scenario, describable as 'some entity bears the property of causing the riot', and specifies that the picture denoted by *a picture of the wall* serves as the property bearer in that scenario. We may then define the predicate as the component of a small clause that contains (or coincides with) the selector of that small clause's remaining component, known as subject-of-predication (see Moro 1997:115 for the original observation that predicates select subjects, not vice versa). We shall see that the notion 'selection' plays a crucial role in the following subsections.

Let us summarise the results of these paragraphs:

(22) If a phrase XP ascribes some property to another phrase YP, then: i. XP and YP form a small clause at the beginning of the derivation; ii. XP selects (or contains a selector of) YP; iii. XP is the predicate of YP.

Interestingly, Lowe (2015:101) analyses Sanskrit bahuvrīhis as expressing an embedded predication at f(unctional)-structure:[29] thus, the f-structure of a bahuvrīhi such as *ugra-putra-* ((4)) specifies that 'mighty', i.e., the basic meaning of *ugrá-*, is the value of the attribute PREDLINK (corresponding to our notion of predicate) and 'son' (i.e., the basic meaning of *putrá-*) the value of the attribute SUBJECT. This can be recast in the terms of the theory of predication adopted here, along the following lines.

The AP *ugrá-* ascribes the property of being mighty to the NP *putrá-* in *ugra-putra-* ((4)). Equivalently, *ugrá-* selects *putrá-* in this compound: *ugrá-* sets up a scenario describable as 'some entity bears the property of being mighty', and specifies that the sons denoted by *putrá-* serve as the property bearer in that scenario. This means that *ugrá-* is the predicate and *putrá-* the subject-of-predication. Then, in accordance with (22), the structure assigned to *ugra-putra-* at the beginning of the derivation is a small clause:[30]

(23) [SC [NP putra-] [AP ugra-]].

4.2 The place of the possessor within Sanskrit bahuvrīhis

In accordance with the Movement Strategy (3b), the possessor *áditi-* of such examples as (4) starts out as the specifier of the phrase projected by the possessum, hence within the phrase projected by the possessum: more explicitly, the position in which *áditi-* starts out in (4) and the position in which a genitival modifier such as *aditeḥ* starts out in *putrā́so áditeḥ* (ṚV 8.18.5) 'Aditi's sons' are one and the same position. In (4) the possessum is *putrá-*, which is generated within the small clause (23). Now, in keeping with the remarks of §3.4, *putrá-* is a bare NP in (4), i.e., an NP that does not contain complements or specifiers. How can we reconcile the proposal that *áditi-* starts out in the specifier of the phrase projected by *putrá-*, with the assumption that *putrá-* is a bare NP? We submit that complement and specifier positions are indeed in principle available in bare phrases such as [NP [N° putra-]], but that these positions cannot host <u>overt</u> linguistic material at the end of the derivation: this means that some expression may sit in the complement or specifier of, e.g., [NP [N° putra-]] at the initial stage of the derivation, provided this expression vacates that complement or specifier

---

[29] In Lexical-Functional Grammar, "f-structure represents syntactic predicate-argument structure in terms of grammatical functions such as subject and object" (Börjars, Nordlinger and Sadler 2019:13). F-structures are graphically represented as sets of attribute-value pairs.

[30] This analysis, which takes the initial stage of the derivation of a nominal compound to involve a small clause, resembles the analysis proposed by Kayne (1994:106) for *that idiot of a doctor*: *that idiot of a doctor* is a complex nominal, but nonetheless involves a predicative link between *doctor* (the subject-of-predication) and *idiot* (the predicate), and this predicative link is implemented by a small clause at the initial stage of the derivation (see also Dikken 1998; 2006:161-246; Moro 2000:49-61).

during the derivation.[31] Of course, in order for this proposal to be tenable, some independent motivation should be found that justifies the expression's vacating the complement or specifier position of the bare phrase. Let us thus assume that the possessor *áditi-* of (4) starts out as the specifier of [$_{NP}$ [$_{N°}$ putra-]], as depicted in (24) (a revised version of (23)).

(24) [$_{SC}$ [$_{NP}$ aditi- [$_{N'}$ putra-]] [$_{AP}$ ugra-]].

Here *áditi-*, by virtue of being a nominal expression that sits in the specifier of the phrase projected by *putrá-*, constitutes a nominal modifier of *putrá-*, and is interpreted as the possessor of *putrá-*. Qua possessor and nominal modifier of *putrá-*, *áditi-* bears an unvalued Case feature (typically genitive) to be valued in a specific syntactic configuration: more precisely, this feature is valued when *áditi-* is in the specifier of a functional head, call it F°, such that F° is part of the DP-layer with which the possessum is endowed, as in (25).[32]

(25) [$_{DP}$ … [$_{FP}$ aditi- [$_{F'}$ [$_{NP}$ [$_{N'}$ putra-]] F°]]]

In this case the possessum is *putrá-*. But to account for the fact that bahuvrīhi-members are morphologically stems, we concluded above (see §3.4-5) that all the functional structure gravitating around these members (e.g., around *putrá-* in (4)) is missing; accordingly, *áditi-* cannot be hosted in any FP placed above *putrá-*. This in turn means that *áditi-* cannot be Case-marked in (24) (more precisely, its Case feature remains unvalued). We have thus found an independent reason for the obligatory movement of *áditi-* from the specifier of the NP headed by *putrá-* in (24): *áditi-* must move to a position outside such an NP because otherwise (i.e., if it stays in situ) it cannot be Case-marked, hence it cannot be licensed.

To sum up: Sanskrit bahuvrīhis' members are phrases projected by a lexical head, contain complements and specifiers, but lack the functional structure—including FP—gravitating around the lexical head. FP is given over, among other things, to valuing the Case feature of the lexical head's nominal modifiers. Therefore, if a nominal modifier of a bahuvrīhi-member is generated in the complement or specifier position projected by this member, such a modifier has to move to the FP gravitating around this member in order to get its Case feature valued. If FP is missing, the nominal modifier in question has to vacate the bahuvrīhi-internal specifier/ complement position in which it was generated, and reach a higher position in which the modifier's Case feature gets valued. Hence, the fact that Sanskrit bahuvrīhis' members lack the functional structure that ordinarily gravitates around non-compound phrases implies the following: that the specifier and complement positions of the phrases corresponding to Sanskrit bahuvrīhis' members cannot contain any overt nominal modifier at the end of the derivation. A welcome result, as we saw.

The take-home point from this subsection is thus the following. From a single hypothesis, i.e., that the functional structure gravitating around both bahuvrīhi-members is missing, we derive two facts: first, the stem status of the bahuvrīhi-members (i.e., the fact that they do not display inflection); second, the inability of the bahuvrīhi-members to contain overt complements or specifiers.

4.3 The linking element in Sanskrit bahuvrīhis

We have posited the structure in (24) in order to account for the interpretation of (4) (notably, the fact that *áditi-* is interpreted as the possessor of *putrá-* and that *ugrá-* ascribes the property of being mighty to *putrá-*). However, we need more structure than is contained in (24) if the Movement Strategy is to be implemented. For one thing, we should provide a landing site for *áditi-*, which we know has to end up outside the compound at the end of the derivation (as shown by (4)). But we also need some additional empty slot for theory-internal reasons.

---

[31] The idea that an illicit structure can be generated provided that some operation intervenes to rescue it before the end of the derivation is not new in the generative literature: indeed, this idea is at the core of Moro's (2000) principle of Dynamic Antisymmetry and Chomsky's (2013; 2015) Labelling Algorithm.

[32] Longobardi (2001:567) proposes that a nominal N_1 marked with an inflectional genitive ending may occupy either a high or a low position (called GenS and GenO, respectively) in the DP-periphery of the nominal N_2 modified by N_1. Longobardi and Silvestri (2013:100-101) offer a minimalist implementation of this proposal in terms of feature-valuation.

Indeed, small clauses are intrinsically unstable structures: the overt linguistic material contained within the small clause cannot be linearised (Moro 2000), and the small clause itself cannot be labelled (Moro 2009; Chomsky 2013; 2015). But linear order and labels are properties that every linguistic output must possess to qualify as well-formed.[33] Consequently, a small clause like (24) cannot reach, as it stands, the end of the derivation: it must somehow be rescued (i.e., linearised and labelled) during the derivation. In brief, the small clause is rescued if either phrase that makes up the small clause (i.e., the subject-of-predication or the predicate) is moved outside the small clause itself.[34] This means that (24) should be supplemented with a second landing site (over and above the landing site for the movement of *áditi-*), devoted to hosting either the subject-of-predication or the predicate of the small clause.

We submit that this second empty slot or landing site is the specifier of a silent linking element L°. Linking elements are meaningless extensions that occur either between the two internal members of a compound (Lieber and Štekauer 2009:§1.3) or, seldom, at the end of a compound (Okubo 2014:230-231). We propose that, in Sanskrit bahuvrīhis, L° occurs at the end of a compound stem, as in (26). More precisely, L° takes the small clause as complement, and projects an LP: indeed, LP is the syntactic structure of the whole compound stem *ugra-putra-* of (4).[35]

(26) [P __] … [LP __ [L' [SC [NP aditi- [N' putra-]] [AP ugra-]] L°]]

The P which shows up in (26) is a cover term for any position outside LP (i.e., outside the structure of the compound stem *ugra-putra-*): we identify P as the empty slot in which the movement of *áditi-* (i.e., the external referent of the bahuvrīhi *ugra-putra-*) terminates. The exact location of P within the sentence changes depending on the Case ending taken on by *áditi-*: if *áditi-* surfaces in the nominative, P coincides with some Spec-TP position (i.e., the position where nominative Case is assigned); if *áditi-* surfaces in the accusative, P coincides with the complement of a transitive verb (i.e., the position in which accusative Case is assigned), and so on.

In our account, the main reason for positing the presence of L° in (26) is theory-internal: either the subject-of-predication (i.e., *[NP aditi- [N' putra-]]*) or the predicate (i.e., *[AP ugra-]*) has to move outside the small clause of (26) in order to rescue the small clause itself; the linking element L° provides the landing site for this movement by projecting an empty specifier (i.e., Spec-LP). However, some evidence exists for the presence of a functional head at the end of Sanskrit bahuvrīhis. Thus, various changes take place at the end of bahuvrīhis (see Macdonell 1916:278-279, from which most of the following examples are taken).

First, if the stem of the right-hand member terminates in *-an-*, the latter may be replaced by *-a-*, as in (27). Second, the semantically empty suffixes *-a-*, *-ya-*, and *-ka-* may be added to the right-hand edge of the bahuvrīhi stem, as in (28). Third, the possessive suffix *-in-* (for which see §2.3 above) may also be added to the right-hand edge of the bahuvrīhi stem, as in (29). Last, the final *-i-* of some stems is replaced by *-a-* when such stems occur in the right-hand slot (i.e., at the end) of the bahuvrīhi; conversely, the final *-a-* of some other stems is replaced by *-i-* when these stems occur in the right-hand slot of the bahuvrīhi: see (30).[36]

---

[33] See more recently Moro and Roberts (2020). The reason why a small clause cannot be linearised or labelled is that it is a symmetrical structure (technically, a 'point of symmetry'), while labelling and linearisation are defined in such a way as to obtain in asymmetrical structures only: see Kayne (1994:3-6) (also Moro 2000:15-28) on the algorithm that determines the linear order of terminal nodes, and Chomsky (2013:43-44) on the algorithm that determines the label of nonterminal nodes.

[34] In essence, the small clause component that undergoes movement (e.g., the subject-of-predication) is replaced by a trace which, being phonologically null, need not be linearised with respect to the other small clause component (e.g., the predicate), thus solving the problem of linearisation. In addition, the trace thereby created, being unable to label the small clause, allows the member that remained in situ within the small clause to provide a label to the small clause itself, thus solving the labelling problem as well. See Moro (2000; 2004), Chomsky (2013; 2015), Rizzi (2016), and Moro and Roberts (2020) for in depth-discussion.

[35] The idea that compounds are headed by a linking element is advocated for by Di Sciullo (2009) and Delfitto and Melloni (2009). According to Eik (2019:182), L° contributes to the interpretation of the compound in which it is contained by providing "an instruction that specifies how elements should compose" (see Eik 2019:183 for an implementation of this idea in formal semantic terms).

[36] See §2.2 above for the special accentual properties of the bahuvrīhis in (27)-(30).

(27) *víśva-* 'all' + *kárman-* 'work' > *viśvá-karma-* (ṚV 10.166.4) 'whose action is on everything', 'performing all work'.

(28) a. *sú-* 'well' + *gó-/gáv-* 'cow' + *-a-* > *sugáva-* (ṚV 1.116.25) 'whose cows are good'.
    b. *anyá-* 'other' + *udára-* 'womb' + *-ya-* > *anyódarya-* (ṚV 7.4.8) 'whose womb is another', 'born from another womb'.
    c. *trí-* 'three' + *ambā́-* 'mother' + *-ka-* > *tryàmbaka-* (ṚV 7.59.12) 'whose mothers are three'.

(29) *śiti-* 'white' + *kákṣa-* 'lurking place', 'belly' + *-ín-* > *śiti-kakṣ-ín-* (TS 5.5.20.1) 'whose belly is white'.

(30) a. *daśan-* 'ten' + *aṅgúli-* 'finger' > *daśāṅgulá-* (ṚV 10.90.1) '(extension) whose fingers are ten', '(extension that is) ten fingers long'.
    b. *dhūmá-* 'smoke' + *gandhá-* 'smell' > *dhūmágandhi-* (ṚV 1.162.15) 'in which there is smell of smoke', 'smelling of smoke'.

The hypothesis that Sanskrit bahuvrīhis involve a functional head L° that is located to the immediate right of the compound stem can explain the data in (27)-(30). Consider how.

    L° takes the small clause as complement to its left, and the small clause contains the compound-members, as depicted in (31) above. L° can be realised as *-a-*, *-ya-*, *-ka-*, *-ín-*, *-i-*, or Ø. After one of the components of the small clause has raised to Spec-LP, and *áditi-* has moved to P, L° is affixed to the remaining component of the small clause: this component happens to be the right-hand member of the compound, as shown in (31).[37]

(31) [P aditi$_j$-] … [LP [AP ugra-]$_i$ [L' [SC [NP t$_j$ [N' putra-]] t$_i$] L°]].
    = (aditi-)   ugra-putra-L°

    Although we are still far from a full understanding of these facts, it seems reasonable to claim that the affixation of L° to the right-hand member of the bahuvrīhi brings about phonological changes in the final segment of the right-hand member itself: deletion of a stem-final sound, as in (27); addition of a suffix, as in (28)-(29); replacement of a stem-final sound with another stem-final sound, as in (30). If no functional head were present at the end of a bahuvrīhi, the phonological changes in (27)-(30) would be hard to justify.

    Let us now return to the well-known fact that Sanskrit bahuvrīhis agree in Case, gender, and number with the external referent of the compound, e.g., with *áditi-* in (4) (see §3.5 above). We should like to show that a relatively simple explanation emerges for this fact if the presence of the linking element L° is posited in the structure of Sanskrit bahuvrīhis.

    To begin with, if LP constitutes the structure of the bahuvrīhi stem (see (26) and (31)), where should the Case ending of a bahuvrīhi be housed? Following Bittner and Hale (1996), we consider Case endings as being the spell-out of a functional head (the Case head K°) which takes a nominal stem as complement. Simplifying somewhat, a Case-marked nominal such as *sákhy-uḥ* (< *sakhi-uḥ*) 'companion-GEN.SG' is analysed as [KP [NP sakhi] [K° uḥ]]: that is, the stem *sákhi-* is an NP, while the Case ending *-uḥ* is a K°, K° being a full-fledged syntactic head; *-uḥ* takes *sakhi-* as complement, and projects a KP; accordingly, the whole Case-marked nominal *sákhy-uḥ* qualifies as a KP headed by *-uḥ* (see Hale and Bittner 1996:3-6). In the same vein, we consider the Case ending of a bahuvrīhi as the spell-out of a Case head K° which takes the whole bahuvrīhi stem LP as complement. For instance, the structure of *ugraputrā* would be the following (based on a slight modification of (31)), with the NOM.SG.F ending *-ā* of *ugraputrā* sitting under K°:
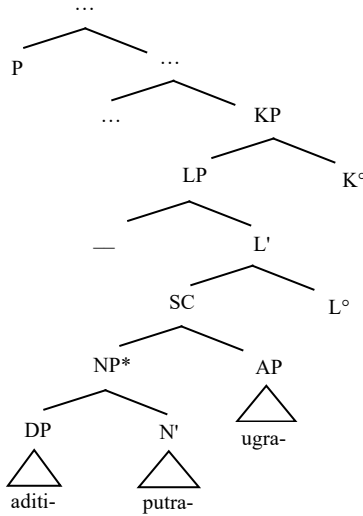
(32) [P aditi$_j$-] … [KP [LP [AP ugra-]$_i$ [L' [SC [NP t$_j$ [N' putra-]] t$_i$] L°]] K°].
    = (Aditi-)   ugra-putra-L°-ā

---

[37] *t$_x$*, which stands for 'trace bearing index x', indicates the position occupied by a coindexed phrase (e.g., XP$_x$) before that phrase moved to another position. See §4.5 below on the mapping between hierarchical structure and linear order.
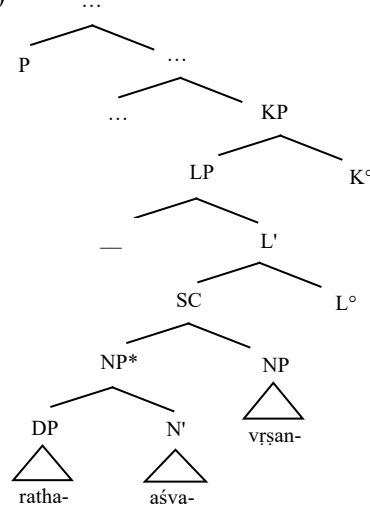
Note now that the Case ending sitting under K° 'views' (i.e., is sensitive to) the gender of the NP stem which K° takes as complement. Thus, when K° takes the masculine NP stem *sákhi-* as complement and K° is GEN.SG, then K° is realised as *-uḥ* (i.e., *sákhy-uḥ*); when instead K° takes the feminine NP stem *sakhī-* as complement and K° is GEN.SG, then K° is realised as *-āḥ* (i.e., *sakhy-āḥ*). However, K° takes LP—and not an NP stem—as complement in (32). Therefore, K° cannot see the gender of any compound-internal NP (e.g., K° cannot see the masculine gender of the NP stem *putrá-* in (32)). Hence the bahuvrīhi's Case ending, which is the spell-out of K°, can acquire the gender of the external referent of the bahuvrīhi: e.g., the Case ending *-ā* acquires the feminine gender of *áditi-* in (32). This explains away the fact that Sanskrit bahuvrīhis have adjectival agreement (i.e., agree in Case, gender, and number with the external referent of the bahuvrīhi). This explanation crucially rests on the intervention of L° between a bahuvrīhi-internal NP stem (e.g., *putrá-* in (32)) and K°: thus, insofar as it is tenable, this explanation lends further support to the postulation of a linking element within Sanskrit bahuvrīhis.

All in all, the proposal that there is a silent functional head L° that takes a small clause as complement in such Sanskrit bahuvrīhis as (4)-(7) is at least compatible with the data found within the Sanskrit language system. Of course, this compatibility does not per se suffice to legitimate the postulation of a silent head: the crucial ingredient needed for legitimating a silent head is that its presence allows us to explain some data (i.e., to decompose complex empirical phenomena into simpler abstract entities). In this case, we shall show that the presence of that head makes it possible to derive the restrictions on the internal orders in (4)-(7) (a complex empirical phenomenon) from the interaction of independently motivated syntactic principles (simpler abstract entities). Thus, we take (26) (= (33)) as the initial stage of the derivation of *ugra-putra-* ((4)).
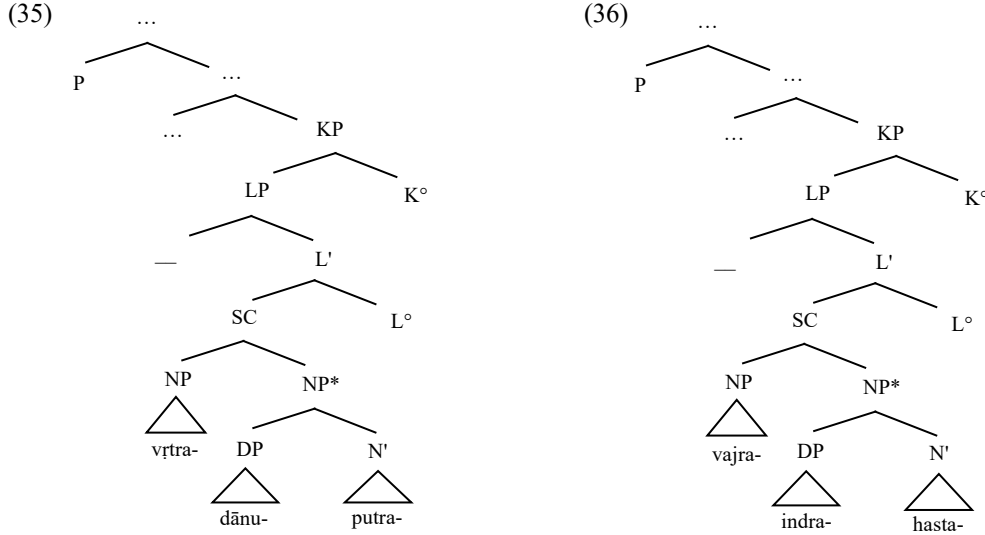
(33)

(34)

Following the same reasoning, we posit an exactly parallel structure for the initial stage of the derivation of *vṛṣaṇ-aśvá-* ((5)), *vṛtrá-putra-* ((6)), and *vájra-hasta-* ((7)).

vṛ́ṣan- 'bull' ascribes the properties typical of bulls (notably, the prominence) to *aśva-* 'horse' in *vṛṣaṇ-aśvá-*. Equivalently, *vṛ́ṣan-* selects *aśva-* in *vṛṣaṇ-aśvá-*: *vṛ́ṣan-* sets up a scenario describable as 'some entity bears the prominence typical of bulls', and specifies that the horses denoted by *aśva-* serve as the prominence bearer in that scenario.

Likewise, *putrá-* 'son' ascribes the property of being a son (specifically, a son of Dānu) to *vṛtrá-* 'Vṛtra' in *vṛtrá-putra-*. Equivalently, *putrá-* selects *vṛtrá-* in *vṛtrá-putra-*: *putrá-* sets up a scenario describable as 'some entity bears the property of being a son of Dānu', and specifies that the individual denoted by *vṛtrá-* serves as the property bearer in that scenario.

By the same token, *hásta-* 'hand' ascribes the property of being in one's hand (specifically, in Indra's hand) to *vájra-* 'mace' in *vájra-hasta-*. Equivalently, *hásta-* selects *vájra-* in *vájra-hasta-*: *hásta-* sets up a scenario describable as 'some entity bears the property of being in Indra's hand', and specifies that the mace denoted by *vájra-* is the property bearer in that scenario.

These bits of semantic information translate, by generalisation (22), into the following structural claims: at the initial stage of the derivation of *vŕṣaṇ-aśvá-*, *vŕṣan-* and *áśva-* form a small clause in which *vŕṣan-* is the predicate and *áśva-* the subject-of-predication; at the initial stage of the derivation of *vṛtrá-putra-*, *putrá-* and *vṛtrá-* form a small clause in which *putrá-* is the predicate and *vṛtrá-* the subject-of-predication; at the initial stage of the derivation of *vájra-hasta-*, *hásta-* and *vájra-* form a small clause in which *hásta-* is the predicate and *vájra-* the subject-of-predication. After the insertion of the linking element L° and the Case head K°, we get the phrase markers (33)-(36); in each of these, the bahuvrīhi's external referent (*áditi-*, *rátha-*, *dā́nu-*, and *índra-*) is represented as a DP to indicate that it bears an unvalued Case feature that needs to be valued later in the derivation.

(35)

```
              ...
             /  \
            P    ...
                /  \
              ...   KP
                   /  \
                  LP   K°
                 /  \
                —    L'
                    /  \
                   SC   L°
                  /  \
                NP    NP*
                 |    /  \
              vṛtra- DP   N'
                     |     |
                  dānu-  putra-
```

(36)

```
              ...
             /  \
            P    ...
                /  \
              ...   KP
                   /  \
                  LP   K°
                 /  \
                —    L'
                    /  \
                   SC   L°
                  /  \
                NP    NP*
                 |    /  \
              vajra- DP   N'
                     |     |
                  indra-  hasta-
```

The trees in (33)-(36) represent the initial stage of the derivation of the bahuvrīhis in (4)-(7). In the remainder of this paper, we shall show that the restrictions on the internal orders in (4)-(7) follow from the constraints at play in the derivation of such bahuvrīhis.

4.4 The Subjacency Condition

Movement has been discovered to obey nontrivial conditions, technically known as locality conditions. One of these is the so-called Subjacency Condition, which constrains movement from maximal projections. We adopt the following version of the Subjacency Condition, due to Moro (1997:49-57, 115; 2017:108, 163), who builds on Cinque (1990:40-43):

(37) Let H be a head nondistinct from [+V]. Then, nothing can move from a maximal projection XP unless one of the following two conditions is met:
   i.   XP is a predicate;
   ii. H locally c-commands and selects XP.

The crucial notions involved in (37) are local c-command and selection. A node α locally c-commands a node β iff: i. α c-commands β; and ii. there is no node γ such that γ c-commands β and is c-commanded by α (see Rizzi 1990:7; Rizzi and Shlonsky 2007:139). An element α selects an element β if α sets up an abstract scenario and specifies the semantic role played by the denotatum of β in that scenario (see §4.1 above).

Crucially, Moro (1997:115-127) has shown that a head nondistinct from [+V] which does not per se select an XP may indeed derivatively select such XP. This happens whenever condition (38) obtains. This condition essentially states that, if a head is close enough to a certain selector, then that head inherits the selectional capacities of that selector.

(38) Let H be a head nondistinct from [+V] and let XP be a maximal projection. Then, H derivatively selects XP iff: the selector of XP locally c-commands H.

Moro (1997:124) capitalises on (38) to explain the contrast between (39) and (40):

(39) a. [DP The [NP* cause of the riot]]i was [SC [a picture of the wall] ti].
　　b. *[Which wall]j do you think [DP the [NP* cause of the riot]]i was [SC [a picture of tj] ti]?

(40) a. [There]i was [SC [a picture of the wall] ti].
　　b. [Which wall]j do you think [there]i was [SC [a picture of tj] ti] ?

The NP* *cause of the riot* is the selector of *a picture of the wall* in (39): NP* sets up a scenario describable as 'some entity bears the property of causing the riot', and specifies that the picture denoted by *a picture of the wall* serves as the property bearer in that scenario; note that in this case the selector (NP*) is contained within the predicate (the whole DP *the cause of the riot*), therefore does not coincide with the predicate. Likewise, *there* is the selector of *a picture of the wall* in (40): *there* sets up a scenario describable as 'some entity bears the property of existing', and specifies that the picture denoted by *a picture of the wall* serves as the property bearer in that scenario (see Moro 1997:142-148 for a precise characterisation of existential meaning); in this case, the selector (*there*) coincides with the predicate. Both (39) and (40) are instances of inverse copular sentences, in the sense that the predicate (the DP *the cause of the riot* and *there*, respectively) raises to the specifier of *was* (see Moro 1997: Ch. 1 and 2 for in-depth discussion). So why is the extraction of *which wall* possible in (40b) but not (39b)?

In a nutshell, *was* is a head nondistinct from [+V] in both (39) and (40), by definition. Moreover, *was* does not select the subject-of-predication (*a picture of the wall*) in (39)-(40), as the subject is already selected by the predicate (*there*), or by a phrase contained within the predicate (the NP* *cause of the riot*). Note however that, while *there* locally c-commands *was* in (40), the NP* *cause of the riot* does not locally c-command *was* in (39): the DP dominating *the* intervenes between NP* and *was* in (39), thereby preventing NP* from c-commanding *was*. If we now bear in mind that the NP* *cause of the riot* and *there* are the selectors of *a picture of the wall* in (39) and (40), respectively, the following picture emerges: only in (40) does the selector of *a picture of the wall* locally c-command *was*. This means that *was* can derivatively select the subject-of-predication *a picture of the wall* in (40) but not in (39), in accordance with generalisation (38).

Since *a picture of the wall* is not a predicate in (38)-(39), movement from it is only possible if *a picture of the wall* is locally c-commanded and selected by *was* (which is a head nondistinct from [+V]), in compliance with the Subjacency Condition. Indeed, *a picture of the wall* is locally c-commanded and (derivatively) selected by *was* in (40) but not (39): consequently, movement from *a picture of the wall* is permitted in (40) but not (39). Thus, the movement of *which wall* in (40b) yields a well-formed sentence because this movement does not violate the Subjacency Condition. On the contrary, the movement of *which wall* in (39b) gives an ill-formed result because it violates the Subjacency Condition (see Moro 1997:115-127 for further details).
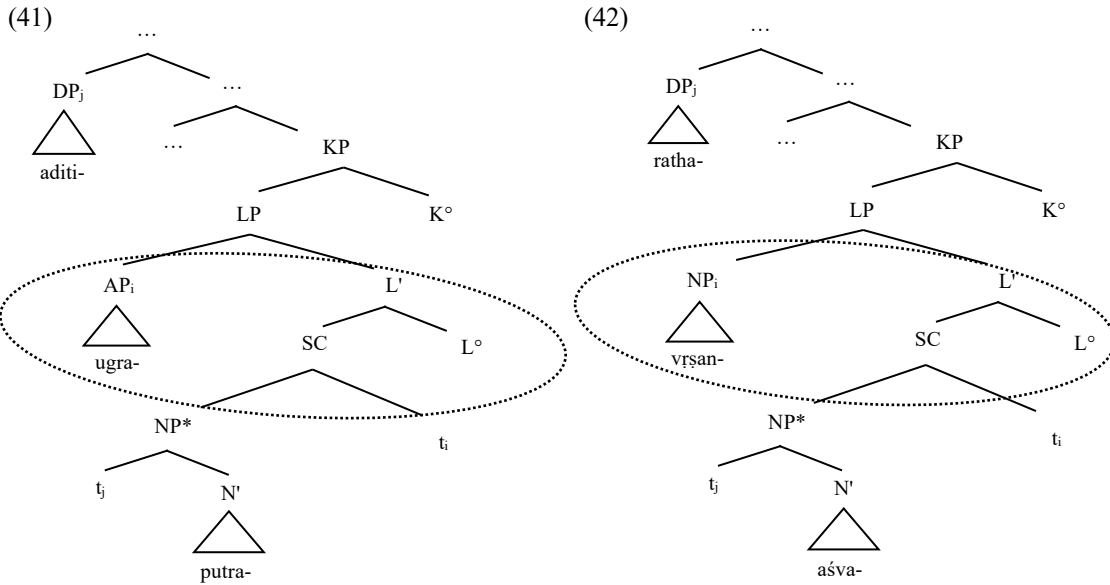
The claim made at the end of the previous subsection, to the effect that the restrictions on the internal orders in (4)-(7) are a consequence of the constraints at play in the derivation of Sanskrit bahuvrīhis like (4)-(7), can now be made more precise: such restrictions follow from the Subjacency Condition. We attempt to show this in the next subsection.

4.5 A unified derivation for Sanskrit bahuvrīhis

The structures in (33)-(36) are unstable in two senses. First, they contain a point of symmetry (i.e., the small clause); second, they contain an element (i.e., the bahuvrīhi's external referent: *áditi-*, *rátha-*, *dā́nu-*, and *índra-*) whose Case feature cannot be valued in situ (i.e., in the specifier of NP*). The problem concerning the point of symmetry is solved if one of the two small clause components (e.g., the NP* *[[aditi-] putra-]* or the AP *ugrá-* in (33)) is displaced to Spec-LP; the problem regarding the unvalued Case feature is instead solved if the bahuvrīhi's external referent (e.g., *áditi-* in (33)) moves to P, which stands for any position in which the Case feature of this external referent can be valued (e.g., Spec-TP).

Now, the bahuvrīhi's external referent cannot directly move to P in (33)-(34). Consider why. The bahuvrīhi's external referent is contained within NP* (a maximal projection) in (33)-(34). Crucially, NP* serves as the subject-of-predication in (33)-(34). Movement from a maximal projection that does not qualify as a predicate complies with the Subjacency Condition (37) only if that maximal projection is selected and locally c-commanded by a head bearing verbal features. We assume that, by virtue of taking a clausal complement (i.e., the small clause), L° bears some verbal feature, on a par with the copula *was* in (39)-(40) above. Now, L° locally c-commands NP*, but does not select it in (33)-(34) (or in (35)-(36)): indeed, NP* is already selected by the small clause predicate, represented as the right-hand component of the small clause (i.e., *ugrá-* and *vŕṣan-*) in (33)-(34). Since L° fails to select NP*, and since NP* is the subject-of-predication (hence, does not qualify as a predicate) in (33)-(34), the direct movement of the external referents *áditi-* and *rátha-* from within NP* to P in (33)-(34) violates the Subjacency Condition. Therefore, this direct movement should be excluded. But if *áditi-* and *rátha-* do not reach P, the bahuvrīhis *ugra-putra-* ((4)) and *vṛṣaṇ-aśvá-* ((5)) cannot be formed: *áditi-* and *rátha-* must end up outside the bahuvrīhis at the end of the derivation, otherwise the Case feature of *áditi-* and *rátha-* cannot be valued. Hence, something must happen in (33)-(34) that enables the movement of *áditi-* and *rátha-* to P. Recall that either component of the small clause must raise to Spec-LP to rescue the small clause. We consider both possibilities in turn.

When it is the right-hand component (i.e., the AP *ugrá-* in (33); the NP *vŕṣan-* in (34)) which raises to Spec-TP, (41)-(42) are obtained.

(41)   (42)

This raising has dramatic consequences. The AP *ugrá-* and the NP *vŕṣan-* now locally c-command L° in their respective trees; moreover, such phrases are the selectors (as well as the predicates) of the NP*s (subjects-of-predication) *[[aditi-] putra-]* and *[[ratha-] aśva-]*, respectively. But this means that L° derivatively selects NP* in both (41) and (42), in accordance with generalisation (38). Since L° also locally c-commands NP* in (41)-(42), the Subjacency Condition is now satisfied (see specifically condition (37ii)): movement from NP* is possible in (41)-(42) insofar as NP* is locally c-commanded and (derivatively) selected by L°, a head which we have assumed is nondistinct from [+V] (i.e., L° bears some verbal feature). Therefore *áditi-* and *rátha-* can now reach P, yielding (41)-(42), which are the structures corresponding to *ugra-putra-* and *vṛṣaṇ-aśvá-*.

We abstract away from the ordering of *áditi-* with respect to *ugra-putra-*, and from the ordering of *rátha-* with respect to *vṛṣaṇ-aśvá-*: *áditi-* may in principle precede or follow *ugra-putra-* in (4), much as *rátha-* may in principle precede or follow *vṛṣaṇ-aśvá-* in (5). Focusing instead on the internal order of the bahuvrīhis *ugra-putra-* and *vṛṣaṇ-aśvá-*, we must make sure that *ugrá-* linearly precedes *putrá-* in (41), and that *vŕṣan-* linearly precedes *aśva-* in (42), otherwise the correct order for *ugra-putra-* and *vṛṣaṇ-aśvá-* is not obtained. This result is achieved without further stipulation

by appealing to the following considerations on Sanskrit word order.

First, in accordance with Hale and Bittner's (1996) analysis, which we have adopted here (see §4.3 above), a Case ending is a syntactic head K°, and the nominal to which the Case ending attaches is the complement of K°. Case endings follow the nominal to which they attach in Sanskrit: e.g., *devá-m* (god-ACC.SG). Therefore, Case heads follow their complements in Sanskrit.

Second, interrogative pronouns occupy the specifier of a functional head in the CP-field in Sanskrit; for simplicity's sake, let us call this position Spec-CP, i.e., specifier of the C° head (see Hale 2018:1933 for a finer representation of the left periphery in Vedic). Moreover, the sentence over which the interrogative pronoun takes scope is projected as the complement of C°. Now, interrogative pronouns ordinarily precede the sentence over which they take scope, as shown below.

(43) kím   ichántī   sarámā   prédám
      what.ACC.SG.N   seek.PRS.PTCP.NOM.SG.F   Saramā.NOM.SG.F   PVB_here
      ānaṭ?
      arrive.IMPF.3SG
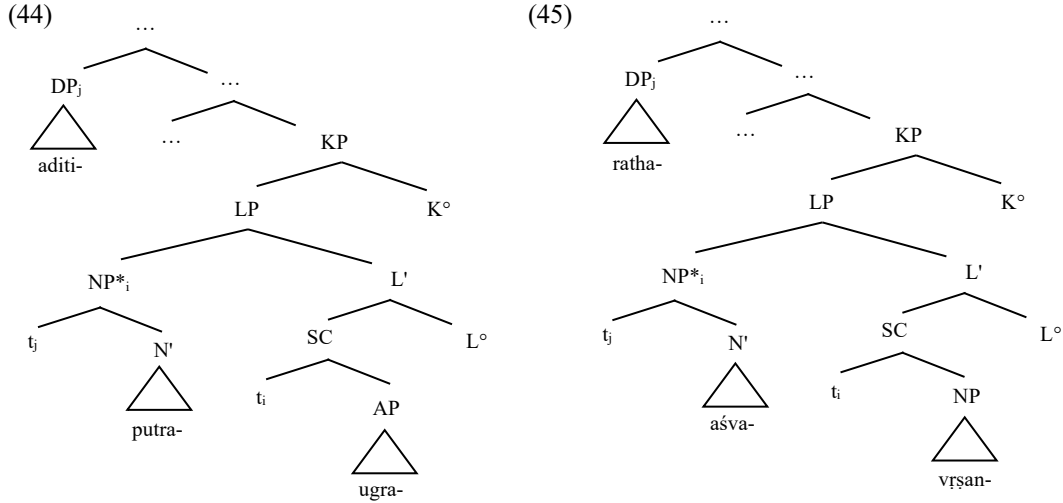   'Seeking what has Saramā arrived here?' (ṚV 10.108.1a; tr. Jamison and Brereton 2014).

Accordingly, the specifier of C° precedes the complement of C° in Sanskrit. We assume that this state of affairs (i.e., the fact that the specifier of a head precedes the complement of that head) also holds for other phrases in which a specifier can be safely detected. The following picture thus emerges: if X° is a Case-like head, X° follows its complement; the specifier of X° precedes the complement of X° generally. To wit, if we confine our attention to phrases headed by a Case-like head, the internal order of these phrases in Sanskrit is: Specifier-Complement-Head.

Let us now come back to (41)-(42). The lexical items contained in Spec-LP (i.e., *ugrá-* and *vṛ́ṣan-*) linearly precede the lexical items contained in the complement of L° (i.e., *putrá-* and *áśva-*, which are the only overt lexical items dominated by the small clause in (41)-(42)), by virtue of the remarks made above: the specifier of a head precedes the complement of that head in Sanskrit.[38] Consequently, we obtain: *ugra-putra-* and *vṛṣaṇ-aśvá-*. These are the correct internal orders of the bahuvrīhis in (4) and (5).

We now turn to the second possible strategy for rescuing the small clause of (33)-(34), i.e., the strategy in which the left-hand component of the small clause (the NP* *[[aditi-] putra-]* in (33); the NP* *[[ratha-] aśva-]* in (34)) raises to Spec-LP. This is depicted in (44)-(45) below. There is no head that locally c-commands and simultaneously selects NP* in (44)-(45): indeed, L° here does not select NP* derivatively either (in the sense of (38)), because the AP *ugrá-* (i.e., the selector of NP* in (44)) and the NP *vṛ́ṣan-* (i.e., the selector of NP* in (45)) do not locally c-command L° in (44)-(45). As a consequence, nothing can be extracted from NP* in accordance with the Subjacency Condition. We have then reached the following conclusion: the external referents *áditi-* and *rátha-* must move to P to get their Case feature valued; yet, the movement of *áditi-* and *rátha-* to P, shown in (44)-(45), violates the Subjacency Condition. Hence, (44)-(45) are ruled out both if *áditi-* and *rátha-* remain in situ, and if they move to P.
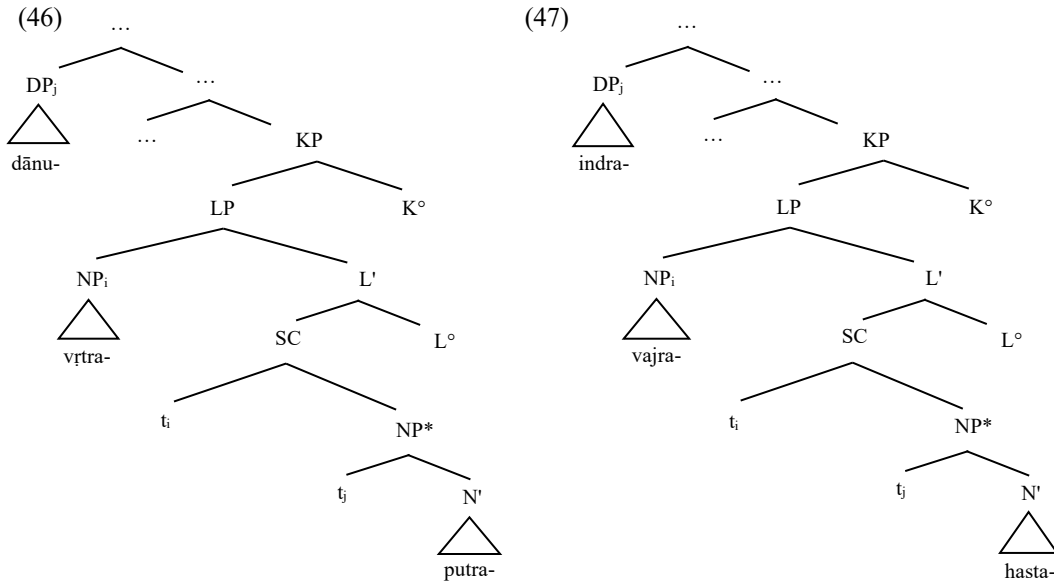
Capitalising on the fact that the specifier of a head precedes the complement of that head, we can map the lexical items populating LP in (44)-(45) to precise linear orders: *putrá-* (which occupies Spec-LP) linearly precedes *ugrá-* (which is contained within the complement of L°) in (44); by the same token, *vṛ́ṣan-* (which occupies Spec-LP) linearly precedes *áśva-* (which is contained within the complement of L°) in (45). To wit, (44)-(45) correspond to the bahuvrīhis **putrogra-* and **aśva-vṛṣan-*, respectively. Since (44)-(45) are illegal structures, **putrogra-* and **aśva-vṛṣan-*, which are derived from (44)-(45), should be ill-formed. This is a welcome result, insofar as **putrogra-* and **aśva-vṛṣan-* are unattested as bahuvrīhis: their non-attestation can now be explained as a result of them violating the Subjacency Condition.

---

[38] When L° is overtly realised (i.e., when L° is spelled out as *-a-*, *-ya-*, *-ka-*, *-ín-*, or *-i-*, as in the examples of §4.3 above), it is akin to a Case ending: L° is a suffix which attaches to a nominal contained in the complement to L°, just as K° is a suffix which attaches to a nominal contained in the complement to K°. Hence, L° linearly follows its complement in (41)-(42), i.e., L° linearly follows the content of the small clause. This may be disregarded here because L° is silent in (41)-(42).

(44)

```
                    ...
         DPj                  ...
         /\              ...        KP
      aditi-                      /    \
                                LP      K°
                              /    \
                          NP*i      L'
                         /   \     /   \
                       tj    N'   SC    L°
                            /\   /  \
                        putra- ti   AP
                                    /\
                                  ugra-
```

(45)

```
                    ...
         DPj                  ...
         /\              ...        KP
      ratha-                      /    \
                                LP      K°
                              /    \
                          NP*i      L'
                         /   \     /   \
                       tj    N'   SC    L°
                            /\   /  \
                        aśva- ti   NP
                                    /\
                                  vṛṣan-
```

Now, (4) and (5) are instances of bahuvrīhis in which the predicate is nominal and the external referent is <u>the Possessor of the subject-of-predication</u>. Clause (i) of generalisation (15) requires that, in such bahuvrīhis, the predicate precede the subject-of-predication; we now have a principled explanation as to why this is so. The internal order in which the predicate precedes the subject-of-predication corresponds to a syntactic configuration that permits the movement of the external referent <u>from within the subject-of-predication</u> not to violate the Subjacency Condition. Consequently, clause (i) of generalisation (15) falls out from the Subjacency Condition.
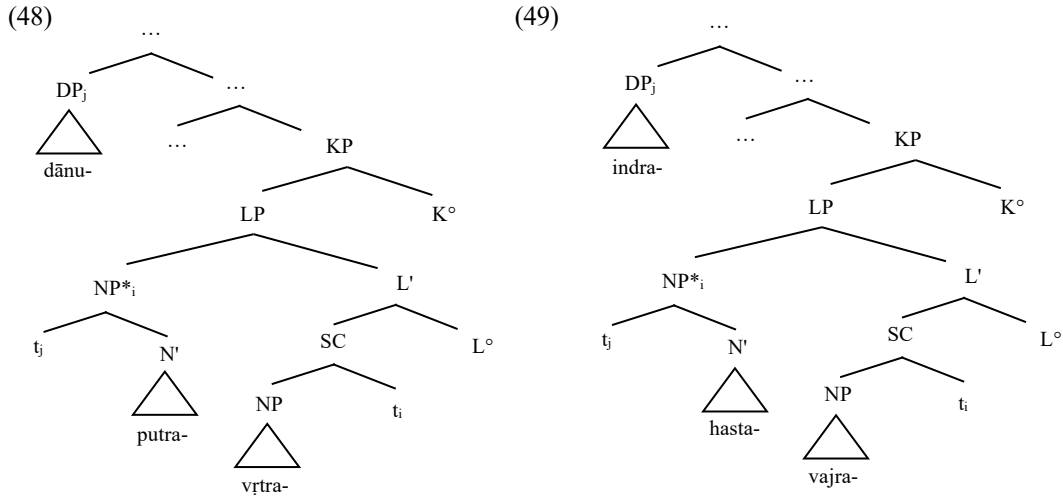
It remains to show how we get from the phrase markers in (35)-(36) to the bahuvrīhis in (6)-(7). To begin with, the external referents *dā́nu-* and *índra-* must move from within NP* to P to get their Case feature valued in (35)-(36). Moreover, either component of the small clause must raise to Spec-LP to rescue the small clause in (35)-(36). Consider the possibility in which the left-hand component of the small clause (i.e., the NP *vṛtrá-* in (35); the NP *vájra-* in (36)) raises to Spec-LP; after the movement of the external referents *dā́nu-* and *índra-* to P, (35)-(36) become as in (46)-(47). The movement of *dā́nu-* and *índra-* from within NP* to P, depicted in (46)-(47), complies with the Subjacency Condition: indeed, NP* fulfills the function of predicate in (46)-(47), and the Subjacency Condition as formulated in (37) permits movement from a maximal projection that qualifies as a predicate.

(46)

```
                 ...
       DPj               ...
       /\          ...        KP
     dānu-                  /    \
                          LP      K°
                        /    \
                     NPi       L'
                     /\       /   \
                  vṛtra-     SC    L°
                            /  \
                           ti   NP*
                               /   \
                             tj    N'
                                   /\
                               putra-
```

(47)

```
                 ...
       DPj               ...
       /\          ...        KP
     indra-                 /    \
                          LP      K°
                        /    \
                     NPi       L'
                     /\       /   \
                  vajra-     SC    L°
                            /  \
                           ti   NP*
                               /   \
                             tj    N'
                                   /\
                               hasta-
```

We now proceed with mapping the lexical items populating LP in (46)-(47) to precise linear orders,

in keeping with our guiding principle that the specifier of a head precedes the complement of that head: (46) maps to *vṛtrá-putra-*, and (47) maps to *vájra-hasta-*. These are the correct internal orders for the bahuvrīhis in (6) and (7).

In contrast, if it is the right-hand component of the small clause (i.e., the NP* *[[dānu-] putra-]* in (35); the NP* *[[indra-] hasta-]* in (36)) that raises to Spec-LP, (35)-(36) become as in (48)-(49) (see below). The phrase markers in (48)-(49) also feature the movement of *dā́nu-* and *índra-* from within NP* to P. Now, NP* is a predicate in (48)-(49) just like in (46)-(47); movement from maximal projections that qualify as predicates is permitted by the Subjacency Condition (37), as we have seen in (46)-(47); accordingly, we would expect the movement of *dā́nu-* and *índra-* from within NP* to P to be Subjacency-compliant in (48)-(49). Note however that NP* is a <u>moved</u> predicate in (48)-(49), unlike in (46)-(47). But movement from a moved predicate is generally difficult (see Moro 1997:50-51; Bianchi and Chesi 2014 and references therein; cf. also Bošković's 2018 study on movement from moved elements); hence, whatever principle bans the movement from a moved predicate will also ban the movement of *dā́nu-* and *índra-* from NP* illustrated in (48)-(49).[39] Since (48)-(49) feature an illegal instance of movement (i.e., the movement of *dā́nu-* and *índra-* from NP*), (48)-(49) are illegal structures.

(48)

```
                ...
          ┌──────┴──────┐
        DPⱼ             ...
      (dānu-)      ┌─────┴─────┐
                  ...          KP
                          ┌─────┴─────┐
                         LP          K°
                     ┌────┴────┐
                   NP*ᵢ        L'
                 ┌──┴──┐    ┌───┴───┐
                tⱼ    N'   SC      L°
                    (putra-) ┌──┴──┐
                            NP    tᵢ
                          (vṛtra-)
```

(49)

```
                ...
          ┌──────┴──────┐
        DPⱼ             ...
      (indra-)     ┌─────┴─────┐
                  ...          KP
                          ┌─────┴─────┐
                         LP          K°
                     ┌────┴────┐
                   NP*ᵢ        L'
                 ┌──┴──┐    ┌───┴───┐
                tⱼ    N'   SC      L°
                    (hasta-) ┌──┴──┐
                            NP    tᵢ
                          (vajra-)
```

The LPs in (48)-(49) map to **putra-vṛtra-* and **hasta-vajra-*, respectively, in keeping with the Specifier-Complement-Head scheme. Since **hasta-vajra-* and **putra-vṛtra-* derive from (48)-(49), which are illegal structures, they should be ill-formed: this is a welcome result, insofar as **putra-vṛtra-* and **hasta-vajra-* are unattested as bahuvrīhis.[40]

To sum up, (6) and (7) are instances of bahuvrīhis in which the predicate is nominal and the external referent is interpreted as <u>the Possessor of the predicate</u>. Clause (ii) of generalisation (15) requires that, in such bahuvrīhis, the predicate follow the subject-of-predication; we now have a principled explanation as to why this is so. Indeed, the internal order in which the predicate follows the subject-of-predication corresponds to a syntactic configuration that permits the movement of the external referent <u>from within the predicate</u> not to violate the Subjacency Condition. As a consequence, clause (ii) of generalisation (15) falls out from the Subjacency Condition.

We have thus reduced both clause (i) and clause (ii) of generalisation (15) to an independently motivated principle: the Subjacency Condition.

4.6 Summary and further remarks

---

[39] It is unclear whether the ban on movement from a moved predicate can be completely reduced to the Subjacency Condition. Such a reduction is carried out in Moro (1997:50-57).

[40] This explanation of the ill-formedness of **putra-vṛtra-* and **hasta-vajra-* is only tenable if the movement of *dā́nu-* and *índra-* from NP* to P in (48)-(49) takes place after the movement of NP* to Spec-LP. It remains unclear, however, what prevents the movement of *dā́nu-* and *índra-* from NP* to P from taking place before the movement of NP* to Spec-LP.

In this section we have proposed a unified derivation of Sanskrit bahuvrīhis like (4)-(7). The main ingredients of this derivation are: first, the assumption that the bahuvrīhi's external referent starts out within the phrase projected by one of the bahuvrīhi-members of (4)-(7), but is moved outside the bahuvrīhi (i.e., outside LP) during the derivation, in accordance with the Movement Strategy (3b); second, the identification of the predicate with the bahuvrīhi-member which selects the remaining bahuvrīhi-member; third, the adoption of a mapping principle whereby the lexical items in Spec-LP linearly precede the lexical items contained in the complement of L°.

Next, we have shown that, under the proposed derivation, the restrictions on the internal orders in (4)-(7)—described in generalisation (15)—reduce to an independently motivated syntactic principle, namely Moro's (1997) version of the Subjacency Condition. Specifically, we have argued that the position occupied by the bahuvrīhi-internal predicate with respect to the bahuvrīhi-internal subject in (4)-(7) merely reflects the syntactic configuration needed for licensing the movement of the bahuvrīhi's external referent from the phrase in which this external referent was generated. Thus, the bahuvrīhi's internal order in which the (nominal) predicate precedes the subject-of-predication corresponds to a syntactic configuration that permits the movement of the bahuvrīhi's external referent from within the subject-of-predication not to violate the Subjacency Condition; on the other hand, the bahuvrīhi's internal order in which the (nominal) predicate follows the subject-of-predication corresponds to a syntactic configuration that permits the movement of the bahuvrīhi's external referent from within the predicate not to violate the Subjacency Condition. The derivation proposed in this section is a specific implementation of the Movement Strategy (3b). Insofar the restrictions on the internal orders in (4)-(7) follow from the Subjacency Condition under this derivation, we may say that the Movement Strategy predicts the internal orders in (4)-(7). On the other hand, the Elided Possessive Unit Strategy (3a) fails to predict the internal orders in (4)-(7) (see §2.3 above). We conclude that the Movement Strategy is empirically superior to the Elided Possessive Unit Strategy.

Furthermore, since the proposed derivation of Sanskrit bahuvrīhis makes crucial use of movement, which is an operation characteristic of syntax as opposed to morphology, the success of the proposed derivation in predicting the internal orders (4)-(7) would also speak in favour of Lowe's (2015) conclusion that Sanskrit compounds are formed by syntactic rules (see §3 above).

Finally, it should be pointed out that generalisation (15) only concerns those bahuvrīhis like (4)-(7) in which the predicate is nominal and the external referent is interpreted as the Possessor of one of the bahuvrīhi-members. Thus, the domain of bahuvrīhis whose predicate is verbal (e.g., *kṛtá-brahman-*, mentioned in §2.1 above), and of bahuvrīhis whose external referent is not interpreted as a Possessor, is still to be investigated.

5 Conclusion

Many Sanskrit bahuvrīhis involve a possessive relation whereby one of the bahuvrīhi-members is the possessum (e.g., *putrá-* in (4)) and the external referent of the bahuvrīhi is the corresponding possessor (e.g., *áditi-* in (4)).

In this work we have shown how this possessive relation can be modelled via the Movement Strategy. That is, at the initial stage of the derivation of the bahuvrīhi, the external referent of the bahuvrīhi is contained within the phrase projected by one of the bahuvrīhi-members (e.g., *áditi-* is housed in the specifier of the NP *putrá-* at the initial stage of the derivation of (4)). In this position the external referent is read as the possessor (in a broad sense) of the bahuvrīhi-member in which it is contained (e.g., *áditi-* is read as the possessor of *putrá-* in (4)). Then, in a later stage of the derivation, the external referent moves from the phrase projected by the possessum and lands in a bahuvrīhi-external position, which we have called P (e.g., *áditi-* moves from within the NP *putrá-* to P in (4)). The movement of the bahuvrīhi's external referent from the phrase projected by the possessum to P is constrained by the Subjacency Condition. The internal order of the bahuvrīhi corresponds to a precise syntactic configuration that may or may not violate the Subjacency Condition. This makes it possible to reduce the restrictions on the internal order of Sanskrit bahuvrīhis like (4)-(7), described in generalisation (15), to the Subjacency Condition: certain internal orders are excluded because they correspond to syntactic configutations that violate the Subjacency Condition, while other internal orders are permitted because they correspond to syntactic configurations that comply with the Subjacency Condition. Accordingly, the Movement Strategy predicts the internal

orders in (4)-(7). By contrast, the Elided Possessive Unit Strategy, according to which a linguistic unit that conveys the meaning 'having' and has undergone ellipsis is affixed to an endocentric compound, fails to predict the internl orders in (4)-(7). This means the Movement Strategy is empirically superior to the Elided Possessive Unit Strategy.

We have concluded from this that the possessive relation holding between the external referent of the bahuvrīhi and one of the bahuvrīhi-members in Sanskrit is established via the Movement Strategy, and not via the Elided Possessive Unit Strategy.

Acknowledgements

References

Primary Literature

A   [Aṣṭādhyāyī]. The Aṣṭādhyāyī of Pāṇini, ed. and tr. by R. N. Sharma, 5 vols. Delhi: Munshiram Manoharlal, 1990-2003.

AVŚ   [Atharvaveda Śaunakīya]. Atharvavedasaṃhitā der Śaunakaśākhā: Eine neue Edition unter besonderer Berücksichtigung der Parallelstellen der Paippalādasaṃhitā, ed. by J-S. Kim. https://www.phil.uni-wuerzburg.de/fileadmin/04080400/2021/Anhang_AVS_Text_06.2021.pdf

BĀU   [Bṛhadāraṇyaka-Upaniṣad]. Bṛhadāraṇyaka-Upaniṣad. In V.P. Limaye and R.K. Vadekar, eds., Eighteen Principal Upaniṣad, Vol. 1. Poona: Vaidika Saṃsodhana Maṇḍala, 1958.

KB   [Kauṣītaki Brāhmaṇa]. Das Kauṣîtaki Brâhmaṇa, ed. by B. Lindner. Jena: H. Costenoble, 1887.

JB   [Jaiminīya Brāhmaṇa]. Jaiminīya Brāhmaṇa of the Sāmaveda, ed. by R. Vira and L. Chandra. Varanasi: Arya Bharati 1954. Reprinted, Delhi: Motilal Banarsidass, 1986.

MBh   [Mahābharata]. The Mahābharata, ed. by V.S. Sukthankar, S.K. Belvalkar, P.L. Vaidya et al., 19 Vols., Poona: Bhandarkar Oriental Research Institute, 1933-1966.

MS   [Maitrāyaṇī Saṃhitā]. Maitrāyaṇī Saṃhitā, ed. by L. von Schroeder, 4 Vols. Leipzig: Brockhaus, 1881-1886. Reprinted, Wiesbaden: Steiner, 1970.

PB   [Pañcaviṃśa Brāhmaṇa = Tāṇḍya-Mahābrāhmaṇa]. The Tāṇḍyamahābrāhmaṇa Belonging to the Sāma Veda with the Commentary of Sāyāṇāchārya, ed. by A. Chinnasvāmi Śāstrī and Paṭṭābhirāma Śāstrī. 2 Vols. Benares: Krishnadas-Gupta, 1935-1936.

ṚV   [Ṛgveda]. Rig Veda: A metrically restored text with an introduction and notes, ed. by B.A. van Nooten and G.B. Holland. Cambridge, MA: Harvard University Press, 1994.

ŚB   [Śatapatha Brāhmaṇa]. The Çatapatha-Brâhmaṇa in the Mâdhynandina-Çâkhâ with Extracts from the Commentaries of Sâyaṇa, Harisvâmin and Dvivedaganga, ed. by A. Weber. Berlin: Dümmler, 1855. Reprinted, Varanasi: Chowkhamba, 1964.

TS   [Taittirīya Saṃhitā]. Die Taittirîya-Saṃhitâ, ed. by A. Weber. Leipzig: Brockhaus 1871-1872.

VaikhGS [Vaikhānasasmārtasūtram]. Vaikhānasasmārtasūtram, the domestic rules of the Vaikhānasa school belonging to the Black Yajurveda, ed. by W. Caland. Calcutta: Baptist Mission Press, 1927.

Secondary Literature

Barcelona, Antonio. 2008. The interaction of metonymy and metaphor in the meaning and form of 'bahuvrihi' compounds. Annual Review of Cognitive Linguistics 6:208-281.

Bauer, Laurie. 2008. Exocentric Compounds. Morphology 18:51-74.

Bauer, Laurie. 2016. Re-evaluating exocentricity in word-formation. In D. Siddiqi and H. Harley, eds., Morphological Metatheory, pages 461-477. Amsterdam: John Benjamins.

Bauer, Laurie. 2017. Compounds and Compounding. Cambridge: Cambridge University Press

Benfey, Theodor. 1852. Vollständige Grammatik der Sanskritsprache. Zum Gebrauch für Vorlesungen und zum Selbststudium. Leipzig: Brockhaus.

Bianchi, Valentina and Cristiano Chesi. 2014. Subject Islands, Reconstruction, and the Flow of the Computation. Linguistic Inquiry 45(4):525-569.

Bittner, Maria and Ken Hale. 1996. The Structural Determination of Case and Agreement. Linguistic Inquiry 27(1):1-68.

Bopp, Franz. 1827. Ausführliches Lehrgebäude der Sanskrita Sprache. Berlin: Dümmler.

Bošković, Želiko. 2018. On Movement out of Moved Elements, Labels, and Phases. Linguistic Inquiry 49(2): 247-282.

Candotti, Maria Piera and Tiziana Pontillo. 2019. Lexical subordination in derivation and compounding: Pāṇini's focusing on the non-head. Studi e Saggi Linguistici. LVIII(2):155-175.

Candotti, Maria Piera and Tiziana Pontillo. 2022. Dispensing with zero in the analysis of Sanskrit bahuvrīhi: Resurfacing, testing and assessing Pāṇini's model. Journal of South Asian Linguistics 12(1):1-22.

Cardona, George. 1997. Pāṇini: His Work and Its Tradition, Vol. 1: Background and Introduction. Delhi: Motilal Banarsidass.

Chomsky, Noam. 1986. Knowledge of language: Its nature, origin and use. New York: Praeger.

Chomsky, Noam. 1995. The Minimalist Program. Cambridge, MA: MIT Press.

Chomsky, Noam. 2013. Problems of projection. Lingua 130:33-49.

Chomsky, Noam. 2015. Problems of projection: Extensions. In E. Di Domenico, C. Hamann, and S. Matteini, eds., Structures, strategies and beyond: Studies in honour of Adriana Belletti, pages 47-66. Amsterdam: John Benjamins.

Chomsky, Noam. 2019. The UCLA Lectures. Lingbuzz/005485.

Cinque, Guglielmo. 1990. Types of Ā-dependencies. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and Its Exceptions. Linguistic Inquiry 36(3):315-332.

Cinque, Guglielmo. 2010. The Syntax of Adjectives: A Comparative study. Cambridge, MA: MIT Press.

Corver, Norbert. 2013. Lexical categories and (extended) projection. In M. den Dikken, ed., The Cambridge Handbook of Generative Syntax, pages 353-424. Cambridge: Cambrisge University Press.

Delfitto, Denis and Chiara Melloni. 2009. Compounds don't come easy. *Lingue e Linguaggio* 8(1):75-104.

Deshpande, Madhav M. 1989. Ellipsis in modern Linguistics and Pāṇini. Annals of Bhandarkar Oriental Research Institute 70:103-124.

Dikken, Marcel den. 1998. Predicate inversion in DP. In A. Alexiadou and C. Wilder, eds., Possessors, Predicates and Movement in the Determiner Phrase, pages 177-214. Amsterdam: John Benjamins.

Dikken, Marcel den. 2006. Relators and Linkers: The Syntax of Predication, Predicate Inversion, and Copulas. Cambridge, MA: MIT Press.

Di Sciullo, Anna-Maria and Edwin Williams. 1987. On the Definition of Word. Cambridge, MA: MIT Press.

Di Sciullo, Anna-Maria. 2009. Why are compounds a part of human language? A view from asymmetry theory. In R. Lieber and P. Štekauer, eds., The Oxford Handbook of Compounding, pages 145-177. Oxford: Oxford University Press.

Eik, Ragnhild. 2019. The Morphosyntax of Compounding in Norwegian. PhD. Thesis, NTNU, Trondheim.

Geach, Peter Thomas. 1980. Reference and Generality: An Examination of Some Medieval and Modern Theories. Ithaca, NY: Cornell University Press.

Geach, Peter Thomas. 1987. Reference and Buridan's Law. Philosophy 62(239):7-15.

Gillon, Brendan. 2008. Exocentric Compounds in Classical Sanskrit. In G. Huet and A. Kulkarni, eds., First International Sanskrit Computational Linguistics Symposium, Oct. 2007, Rocquencourt, France. http://hal.inria.fr/SANSKRIT/fr/. <inria-00202860>

Gippert, Jost, Javier Martínez, Agnes Korn, and Roland Mittmann. 2016. Thesaurus Indogermanischer Text- und Sprachmaterialien (TITUS). https://titus.fkidg1.uni-frankfurt.de/indexe.htm.

Guevara, Emiliano and Sergio Scalise. 2009. Searching for Universals in Compounding. In: S. Scalise, E. Magni, and A. Bisetto, eds., Universals of Language Today. Studies in Natural Language and Linguistic Theory 76:101-128.

Hale, Mark. 2018. The syntax of Indo-Iranian. In J. Klein, B. Joseph, and M. Fritz (in cooperation with Mark Wenthe), eds., Handbook of Comparative and Historical Indo-European Linguistics, Vol. 3, pages 1924-1942. Berlin: de Gruyter.

Hellwig, Oliver. 2010-2021. DCS – The Digital Corpus of Sanskrit. http://www.sanskrit-linguistics.org/dcs/index.php

Jamison, Stephanie W. and Joel P. Brereton, trs., 2014. The Rigveda: The Earliest Religious Poetry of India, 3 Vols. Oxford: Oxford University Press.

Kastovsky, Dieter. 2009. Diachronic perspectives. In R. Lieber and P. Štekauer, eds., The Oxford Handbook of Compounding, pages 323-340. Oxford: Oxford University Press.

Katre, Sumitra M. (translator). 1987. Aṣṭādhyāyī of Pāṇini. Austin: University of Texas Press.

Kayne, Richard S. 1994. The antisymmetry of syntax. Cambridge, MA: MIT Press.

Kiparsky, Paul. 1982. From Cyclic Phonology to Lexical Phonology. In H. van der Hulst and N. Smith, eds., The structure of phonological representations (Part I), pages 131-175. Dordrecht: Foris.

Kiparsky, Paul. 2009. On the architecture of Pāṇini's grammar. In G. Huet, A. Kulkarni, and P. Scharf, eds., Sanskrit Computational Linguistics, ISCLS 2007/2008. Lecture Notes in Computer Science 5402:33-94.

Kiparsky, Paul. 2010. Compositional vs. paradigmatic Approach to Accent and Ablaut. https://web.stanford.edu/~kiparsky/Papers/ucla_IE_09.submitted.new.pdf

Kiparsky, Paul. 2016. The agent suffixes as a window into Vedic grammar. http://web.stanford.edu/~kiparsky/Papers/agentnouns.final.pdf

Longobardi, Giuseppe. 2001. The Structure of DPs: Some Principles, Parameters, and Problems. In M. Baltin and C. Collins, eds., The Handbook of Contemporary Syntactic Theory, pages 562-603. Malden, MA: Blackwell.

Longobardi, Giuseppe and Giuseppina Silvestri. 2013. The Structure of NPs: Some Insights on Case, Empty Categories, and Poverty of Stimulus. In S. Luraghi and C. Parodi, eds., The Bloomsbury Companion to Syntax, pages 88-117. London: Bloomsbury.

Lundquist, Jesse and Anthony D. Yates. 2018. The morphology of Proto-Indo-European. In J. Klein, B. Joseph, and M. Fritz (in cooperation with M. Wenthe), eds., Handbook of Comparative and Historical Indo-European Linguistics, Vol. 3, pages 2079-2195. Berlin: de Gruyter.

Lowe, John J. 2015. The syntax of Sanskrit compounds. Language 91(3):71-114.

Macdonell, Arthur Anthony. 1910. Vedic Grammar. Strassburg: Trübner.

Macdonell, Arthur Anthony. 1916. A Vedic grammar for students. Oxford: Clarendon Press.

Mādhavānanda, Swāmī (translator). 1950. The Bṛhadaraṇyaka Upaniṣad: With the commentary of Śaṅkarācārya. Hollywood, CA: Vedanta Press.

Manzini, Maria Rita. 2017. Passive, smuggling and the by-phrase. In N. LaCara, K. Moulton, and A-M. Tessier, eds., A Schrift to Fest Kyle Johnson, pages 233-244. Linguistics Open Access Publications,.

Marchand, Hans. 1967. Expansion, transposition, and derivation. In D. Kastovsky, ed., Studies in Syntax and Word-Formation. Selected articles by H. Marchand, pages 322-337. Münchener: Fink.

Molina-Muñoz, Adriana. 2013. Sanskrit Compounds and The Architecture of the Grammar. In S-F. Chen and B. Slade, eds., Grammatica and Verba. Glamour and Verve: Studies in South Asian, historical, and Indo-European linguistics in honor of Hans Henrich Hock on the occasion of his seventy-fifth birthday, pages 181-201. Ann Arbor: Beech Stave Press.

Moro, Andrea. 1997. The raising of predicates: Predicative noun phrases and the theory of clause structure. Cambridge: Cambridge University Press.

Moro, Andrea. 2000. Dynamic Antisymmetry. Cambridge, MA: MIT Press.

Moro, Andrea. 2004. Linear compression as a trigger for movement. In H. van Riemsdijk and A. Breitbarth, eds., Triggers, pages 387-429. Berlin: de Gruyter.

Moro, Andrea. 2009. Rethinking symmetry: A note on labeling and the EPP. Snippets 19:17-18.

Moro, Andrea. 2017. A Brief History of the Verb to Be (tr. by B. McClellan-Broussard). Cambridge, MA: MIT Press.

Moro, Andrea. 2019. The geometry of predication: A configurational derivation of the defining property of clause structure. Phil. Trans. R. Soc. B 375(1791):1-7.

Moro, Andrea and Ian Roberts. 2020. Unstable structures and Generalized Dynamic Antisymmetry. Lingbuzz/005515.

Ntelitheos, Dimitrios. 2022. Compounding in Greek as Phrasal Syntax. Languages 7(151):1-22.

Okubo, Tatsuhiro. 2014. Expletives in words: Linking elements as markers of wordhood. Lingvisticæ Investigationes 37(2):225-239.

Pennanen, Esko V. 1982. Remarks in Syntagma and Word-Formation. Folia Linguistica 16:241-261.

Perrin, Jean-Baptiste. 1913. Les atoms. Paris: Alcan.

Pontillo, Tiziana. 2021. Did the Sanskrit model bring «true enlightenment to european scholars» when they analysed and classified the bahuvrīhi compounds?. Studi Classici e Orientali 67:497-514.

Postal, Paul. 1969. Anaphoric islands. Proceedings from the Annual Meeting of the Chicago Linguistic Society 5(1):205-239.

Renou, Louis. 1961. Grammaire Sanscrite. Paris: Maisonneuve.

Rizzi, Luigi. 1990. Relativized Minimality. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2013. The functional structure of the sentence, and cartography. In M. den Dikken, ed., The Cambridge Handbook of Generative Syntax, pages 425-457. Cambridge: Cambridge University Press.

Rizzi, Luigi. 2016. Labeling, maximality and the head-phrase distinction. The Linguistic Review 33(1):103-127.

Rizzi, Luigi and Ur Shlonsky. 2007. Strategies of subject extraction. In H-M. Gärtner and U. Sauerland, eds., Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics, pages 115-160. Berlin: de Gruyter

Scalise, Sergio and Antonietta Bisetto. 2009. The Classification of Compounds. In R. Lieber and P. Štekauer, eds., The Oxford Handbook of Compounding, pages 34-53. Oxford: Oxford University Press.

Scalise, Sergio and Antonio Fábregas. 2010. The head in compounding. In S. Scalise and I. Vogel, eds., Cross-Disciplinary Issues in Compounding, pages 109-126. Amsterdam: John Benjamins.

Scalise, Sergio and Irene Vogel. 2010. Why compounding? In S. Scalise and I. Vogel, eds., Cross-Disciplinary Issues in Compounding, pages 1-18. Amsterdam: John Benjamins.

Selkirk, Elisabeth O. 1982. The syntax of words. Cambridge, MA: MIT Press.

Speijer, Jacob Samuel. 1886. Sanskrit Syntax. Leiden: Brill.

Wackernagel, Jakob. 1905. Einleitung zur Wortlehre. Nominalkomposition (Altindische Grammatik, Band II.1). Göttingen: Vandenhoeck & Ruprecht.

Whitney, William Dwight. 1889. A Sanskrit Grammar, including both the Classical language, and the older Dialects, of Veda and Brahmana. Cambridge, MA: Harvard University Press.

Zamparelli, Roberto. 2000. Layers in the Determiner Phrase. New York: Garland.

An acoustic study of voiceless stops in Indian English

JAHNAVI NARKAR, *University of California–Los Angeles*
PETER STAROVEROV, *independent scholar*

ABSTRACT

This paper analyzes phonetic corpus data showing that speakers of Indian English produce word-initial and post-/s/ stops that do not differ in terms of voice onset time (VOT). They systematically produce both categories with short positive VOT which is typical of unaspirated stops. By contrast, speakers of British English do distinguish aspirated and unaspirated stops in terms of VOT in the same contexts. These results are based on VOT measurements for word-initial and post-/s/ bilabial and dorsal stops produced by 102 bilingual speakers of one of nine Indic languages and Indian English, and by 28 monolingual British English speakers. Given that Indian English is influenced by Indic languages which have phonological voiceless aspirated stops, this pattern is particularly surprising. We suggest that this cross-language category mismatch might be driven by the phonetic implementation of aspirated stops in Indic languages. More generally, if our results are representative of the stage in which Indian English and British English were in direct contact, they suggest that the existence of a richer laryngeal system in borrowing languages cannot in itself prepare the speakers to perceive and produce borrowed laryngeal contrast more faithfully. Finally, we also measured vowel onset f0 in our corpus, but found no significant differences by context – word-initial and post-/s/ stops were followed by comparable onset f0. These results suggest that intonational context may need to be further controlled for more sensitive f0 analysis.

1 Introduction

This paper establishes the acoustic correlates of voiceless stops in Indian English. It shows that English speakers of Indic (also known as Indo-Aryan) languages, which have phonological voiceless aspirated stops in their inventory, systematically produce English stops with short-lag, rather than long-lag VOT. Although this seems unexpected from a systemic phonological point of view, we argue that the phonetic implementation of aspirated stops in Indic languages likely contributes to this cross-language category mismatch (see also Sailaja (2009) on how orthography may contribute to this phenomenon).

Although our study does not directly address loanword adaptation, this pattern of adaptation in Indian English is also reported to apply when English words are borrowed into Indic languages (Paradis and LaCharité 1997). It is, therefore, a potential challenge for theories of loanword adaptation based on featural closeness (Paradis and LaCharité 1997, LaCharité and Paradis 2000, 2005), as aspirates in English and Indic are often thought to have the same features (e.g. Iverson and Salmons 1995). We speculate that the existence of a richer laryngeal system in the borrowing language does not in itself prepare the speakers to perceive and produce the English laryngeal contrast more faithfully.

Existing reports of the production of voiceless stops in Indian English are largely intuitive and for the most part, do not cite detailed acoustic data. A notable exception is Awan and Stine (2011) who compared the VOT of bilabial stops in the words *path* and *apparently* and coronal stops in the words *token* and *foretell* in Indian English and American English. They found that Indian English speakers' production was characterized by significantly shorter VOTs compared to American English speakers' production of the same stops. Awan and Stine (2011) did not include stops following /s/ in their study. Given the paucity of detailed quantitative data, reliance on intuitive reports is potentially problematic since such reports of phonological patterns do not always survive closer scrutiny (Zhang and Lai 2008, Alderete and Kochetov 2009). Moreover, cross-language adaptation patterns might not be directly accessible to intuition, and intuitions themselves can be subject to bias (Kawahara 2011). In light of these findings, our goal here is to test the existing intuitive reports

(Kelkar 1957, Nair 1996, Coelho 1997, Sailaja 2009) and document the production of English stops in Indian English with robust acoustic data.

Present-day Indian English is a conventionalized variety of English spoken by English speakers in the Indian subcontinent (Hoffmann et al. 2011). In addition to providing a phonetic description of this widely spoken yet under-studied variety of English, our study also contributes to the understanding of the processes that were involved in the formation of Indian English phonetics, when learners of British English were presumably influenced by their native languages. Thus we hope that this study will stimulate future research on synchronic cross-language adaptation patterns where English aspirates adapt to native phonetics.

Labial	Dental	Retroflex	Palatal	Velar
p b	t̪ d̪	ʈ ɖ	c ɟ	k g
pʰ bʰ	t̪ʰ d̪ʰ	ʈʰ ɖʰ	cʰ ɟʰ	kʰ gʰ

Table 1. Inventory of stop consonants in relevant Indic languages.

The Indic languages that we investigate all have the stop inventory shown in Table 1 (see section 2.1 for a list of references). Dravidian languages which also have associated Indian Englishes were excluded from this study. Some Dravidian languages like Tamil do not have a four-way contrast among stops. Some others which do, like Kannada and Malayalam, do not employ it across the lexicon or registers. In Kannada, for example, voiced aspirated stops are only found in Sanskritic words, and even those are neutralized to plain voiced stops in most registers (Upadhyaya 1972). To avoid including the effect of phonetic grammars that do not encode the four-way contrast in a similar manner, we restrict our study to Indic languages only. This is not meant to serve as a claim about which languages have historically contributed to the development of present-day Indian English(es), but as a methodological simplification. Throughout this paper, we use the term "Indian Englishes" to refer to the varieties of English spoken by speakers who are bilingual in English and an Indic language.

Our study focuses on labial and dorsal stops since these two places of articulation are shared between English and Indic languages. We did not include coronal stops in our study for several reasons. First, Indic languages have two kinds of coronal stops – dental and retroflex, and neither of these places of articulation matches the English alveolar stops. Although the short-lag stop pattern likely applies to coronal stops in Indian English, documenting this phenomenon will be hindered by potential additional adjustments in place. Second, as we discuss in section 2.1, the context where word-initial /t/ appears in our dataset is likely to introduce additional variability in its production, making it even harder to assess the cues to aspiration.

	k	kʰ	p	pʰ
Bengali (Mikuteit and Reetz 2007)	34	73	14	-
Gujarati (Rami et al. 1999)	41	75	-	-
Hindi (Lisker and Abramson 1964)	16	84	12	63
Marathi (Lisker and Abramson 1964)	21	73	0	35
Nepali (Poon and Mateer 1985)	27	84	6	69
Pahari (Khan and Bukhari 2011)	27	77	19	66
Urdu (Hussain 2018)	35	85	18	73
Indic (average)	29	79	12	61
British English (average)	25	62	11	44
British English (Docherty 1992)	27	62	15	42
British English (Whiteside and Irving 1997)	22	62	7	45

Table 2. A comparison of the existing results on VOT (ms) in Indic languages and in British English.

Table 2 gives an overview of the VOT of voiceless aspirated and unaspirated dorsal and labial

stops for Indic languages and for British English reported in the literature. Aside from British English, this table includes all languages in our dataset (see section 2.1) for which we could find data.[1]

In examining Table 2, a few important caveats should be kept in mind. First of all, the VOT values averaged across Indic languages may only be used to get a very coarse-grained view of laryngeal phonetics since each of those languages may differ from the others considerably. Second, cross-linguistic phonetic comparisons are hindered by the fact that the studies in question differ in methodology, and hence the values listed here are not obtained under comparable conditions. That said, the data in Table 2 support the idea that Indic languages have longer VOT in voiceless aspirated stops than British English, while showing similar VOT in voiceless unaspirated stops. A more direct comparison of American English with Hindi and Marathi is presented by Lisker and Abramson (1964), where the same method is used for each language. Although Lisker and Abramson have only one speaker of each Hindi and Marathi, at least for the former their results are likely representative since two other studies of Hindi yielded much higher VOT values for voiceless aspirated stops (Benguerel and Bhatia 1980, Shimizu 1989). Table 3 lists the VOT values pertaining to utterance-initial voiceless aspirated stops in Lisker and Abramson (1964). Here again, aspirated stops have a longer VOT in Indic languages than in English, usually with a substantial difference.

	/pʰ/	/kʰ/
Hindi	63	84
Marathi	35	73
English	28	43

Table 3. Summary of the average VOT (ms) for Hindi, Marathi, and English reported by Lisker and Abramson (1964).

The fact that Indic languages have a larger stop inventory may also suggest that aspirated stops will exhibit less variability in VOT than their English counterparts. However a recent detailed study of Hindi by Hauser (2019) did not find a distinction in VOT variability between these two languages.

Since Indic languages have a richer set of laryngeal contrasts than English, perceptually these contrasts may be supported by multiple kinds of cues. First, there is evidence that all four categories of Indic stops cannot be distinguished based on VOT alone, and the literature proposes a number of other durational cues to laryngeal contrasts (Mikuteit and Reetz 2007, Rami et al. 1999, Davis 1994, Dutta 2007, Berkson 2012, Dmitrieva and Dutta 2020, Clements and Khatiwada 2007, Schwarz et al. 2019). In this article, one of the cues we investigate is VOT. While this cue is the best studied cue to voicing cross-linguistically, it should be kept in mind that the actual durational cues employed by Indic speakers may be more complex.

The second cue we investigate is onset f0 following the stops. f0, which is a secondary cue to voicing and/or aspiration is some languages, may be pointing in different directions in Indic languages versus in British English. Tonal effects of aspiration are less studied than VOT, so we can only draw some preliminary conclusions from the literature here. Compared to voiced unaspirated stops in English, voiceless aspirates are reported to be associated with rising f0 in the following vowel (House and Fairbanks 1953, Lehiste and Peterson 1961, Hombert 1976, Ohde 1984, Kingston and Diehl 1994). However, in these studies, aspiration co-varies with voicing, so the conclusions we can draw are only preliminary. Studies of Indic languages comparing effects on f0 of aspirated versus unaspirated voiceless stops report a relatively small local f0 *lowering* with aspiration. This is found for f0 of vowels following /k/ versus /kʰ/ in Marathi (Dmitrieva and Dutta 2020), Hindi/Urdu (Dutta 2007, Schertz and Khan 2020), and Nepali (Clements and Khatiwada 2007). Thus, although tonal effects of aspiration are understudied, the existing literature suggests that aspiration in English voiceless stops may tend to raise f0 of the following vowel, while in Indic languages it may have a lowering effect.

To summarize, existing results on the acoustics of Indic voiceless stops suggest two important differences compared to English. In terms of VOT, Indic aspirated stops are more aspirated than English aspirated stops, while Indic unaspirated stops, like their English counterparts, are short-lag

---

[1] Note that Mikuteit and Reetz (2007) did not measure the VOT of [pʰ] since their speakers consistently produced these as [f].

and have comparable VOTs. English aspirated stops are also reported to correspond to higher f0 values in the following vowel, while Indic aspirated stops correspond to lower f0 values. Assuming this current situation is reflective of the state of British English and Indic languages during the first substantial stages of language contact, we may hypothesize that when British English aspirated stops were perceived by native Indic speakers, they may not have been 'aspirated enough' in terms of VOT, and may have had f0 effects inconsistent with native effects in Indic languages. For these reasons, English aspirated stops may have been categorized with native unaspirated stops, as models of non-native speech perception would predict (Best 1995, Best and Tyler 2007, Flege 1995). If this is correct, we expect that word-initial voiceless stops in Indian English have shorter VOT values compared to other varieties of English, and they may lack lowering f0 effects on the following vowel. The aim of our study is to test if these acoustic predictions hold of actual Indian English productions, and hence to document the acoustics of voiceless stops in a corpus of Indian English. The paper is organized as follows. Section 2 presents our methods, followed by results (section 3), discussion (section 4), and concluding remarks (section 5).

2 Method

2.1 Materials and participants

In this study, we analyzed audio recordings of English speech from the Speech Language Archive (Weinberger 2014) – a web-based corpus containing recordings of the same English passage read by multiple speakers of different English varieties, both native and non-native. From the archive (accessed on October 10, 2019), we selected all recorded speech samples from adult speakers of Indian English who were also native speakers of one of nine Indic languages listed in Table 4.[2] These languages were selected because they all have the same inventory of stop consonants, as shown in Table 1.

Twenty-eight adult monolingual speakers of British English (BE), sixteen males and twelve females, functioned as the comparison group. These speakers were all speakers of the same variety, southern British English. The speakers selected were English speakers on the archive who were born and resided in England. Sex, present residence, and birthplace were self-reported by the speakers.

The archive uses self-reported native languages, and hence distinguishes Hindi and Urdu. Following this, we treat Hindi and Urdu as separate languages in our analysis, while not necessarily supporting the claim of these being two different languages. Although multilingualism is common in India, we unfortunately do not have enough information on other languages spoken by the subjects. However, speech samples produced by speakers who spent their childhood and/or adolescence (anytime under the age of eighteen years) in a primarily English-speaking country other than India were excluded.

Language	Number of speakers	Grammatical description
Bengali	14	Khan (2010)
Gujarati	11	Mistry (1997)
Hindi	28	Ohala and Ohala (1972)
Konkani	3	Almeida (2003)
Marathi	8	Dhongde and Wali (2009)
Nepali	14	Acharya (1991)
Oriya	2	Majumadāra (1970)
Pahari	2	Jouanne (2014)
Urdu	13	Ohala and Ohala (1972)
Total	102	

Table 4. Number of speakers and the referenced grammatical description for each language analyzed.

---

[2] Other major Indic languages like Kashmiri, Punjabi and Sinhala were not included since these languages do not have the four-way laryngeal contrast.

Each participant read the same passage, and the recordings are therefore comparable to each other, although speech rate and recording conditions are not controlled for. The passage read by each participant is given in (1). The Speech Accent Archive also provides transcriptions for some of the recordings, but our study did not rely on these transcriptions.

(1) Passage read by participants on the Speech Accent Archive
     "Please call Stella. Ask her to bring these things with her from the store: Six **spoons** of fresh snow **peas**, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the **kids**. She can **scoop** these things into three red bags, and we will go meet her Wednesday at the train station."

The words analyzed in this study include *kids* and *peas* (word-initial stops), and *spoons* and *scoop* (post-/s/ stops). These words are highlighted in bold in (1). These words were chosen since they are all content words occurring in prominent contexts and were least likely to be reduced. All the target words are nouns, except for the verb *scoop*, which is the only instance of unaspirated [k] before a stressed vowel.

We decided not to include the word-initial /t/ of the word *toy* (comparing it, for example, to /t/ in *store*) for two reasons. First, as discussed in section 1, Indic languages do not have alveolar stops, and hence Indian English productions of /t/ might differ not only in aspiration, but also of place, creating a confound. Second, a preliminary examination of a subset of the recordings revealed that the word *toy* occurred in a garden path context and elicited unnatural productions from a number of speakers. In other words, a number of speakers appeared to first parse *toy* as head of its NP, then noticed that in fact *frog* was the head, and *toy* the specifier. This often resulted in elongated and hesitant pronunciations of *toy* and a pause between *toy* and *frog*. These mispronunciations would create high variability in our phonetic data, and make the results for coronals less comparable to other places of articulation. For these reasons, coronal stops were excluded from consideration.[3]

2.2 Annotations and measurements

The words selected for this study allowed us to compare stops that are aspirated in British English (*peas*, *kids*) to unaspirated stops (*spoons*, *scoop*). Each speaker contributed one token of the words in each category in our study. In six cases, a speaker did not produce a given word or misspoke (for example, produced *swoop* instead of *scoop*). Such tokens were excluded.

Our statistical analyses involved calculating the VOT and f0 difference between the initial stops and stops after /s/. For those calculations, we had to also exclude all tokens that didn't have pairs. For example, if a speaker produced *swoop* instead of *scoop*, we excluded the token *swoop*, and also did not analyze the word *kids* from that speaker since there was no #s_ token to compare it to. We followed a similar procedure for words that had to be excluded from f0 measurements because of creaky voice.

	BE	IE-Ben	IE-Guj	IE-Hin	IE-Kon	IE-Mar	IE-Nep	IE-Ori	IE-Pah	IE-Urd	Total
Dor	56	26/24	22	52/50	5	16/14	22	4	4	26/24	176
Lab	56	28	22	52	6	18/16	24	4	4	26	184

Table 5. Number of tokens analyzed for VOT (first number) and f0 (second number, if different), excluding speaker errors and instances of creaky voice.

Table 5 shows the total number of tokens analyzed for each context (recall that each speaker produced just two tokens for each place of articulation). Here and in what follows, we abbreviate the language names: BE – British English; Ben – Bengali; Guj – Gujarati; Hin – Hindi; Kon – Konkani; Mar – Marathi; Nep – Nepali; Ori – Oriya; Pah – Pahari; Urd – Urdu.[4] Note that for the

---

[3] Our preliminary acoustic measurements based on a subset of Indian English data, nonetheless, confirm that the VOT of /t/ in *toy* is, like /p/ and /k/, consistently short-lag.

[4] Recall that first language is reported by the speakers themselves and hence Hindi and Urdu are considered different languages.

Indic languages, the language being analyzed is the variety of English associated with a particular Indic language. Whenever the token counts for VOT and f0 are different, the count for f0 is given as a second number in Table 5.

Acoustic measurements were done in Praat (Boersma and Weenink 2019). VOT was measured manually, as the interval between the beginning of the release burst and the onset of quasi-periodicity.

Since f0 is one of the cues to laryngeal distinctions, fundamental frequency was measured for the vowel immediately following the target stops. f0 measurements were taken using a modified version of a Praat script by Katherine Crosswhite[5] starting at the first point immediately at the onset of voicing at which periodicity could be detected, and then at four additional equally spaced points within the first half of vowel duration. This procedure was adopted for two reasons. First, f0 depression has been found to be local; that is f0 following voiced and aspirated stops is lowered immediately following the stop with lowering effects quickly tapering off beyond this point. Second, measuring f0 at equidistant points in the vowel meant differences in the vowel lengths produced by different speakers in different contexts did not affect the f0 measurement.

Self-reported speaker sex from the Speech Accent Archive was used to set the measurement parameters in the script. In a few cases, f0 could not be measured at a given point because of creaky voice – these cases were inspected manually to confirm the instance of aperiodicity before excluding these individual measurements. Automatic f0 measurements were corrected manually to fix errors which typically result from pitch-halving and pitch-doubling errors of Praat's autocorrelation algorithm (Boersma 1993). f0 outliers were identified by visually inspecting the f0 distribution at each measurement point. For the outlier tokens, pitch was measured manually by taking the duration of a single glottal pulse as the duration of one cycle of the periodic waveform and taking its inverse (see also Dmitrieva and Dutta 2020).

2.3 Summary

To summarize, this study investigated the VOT and f0 of voiceless stops in a corpus of Indian English, that is, the English of speakers whose native languages are Indic. The measurements for Indian English speakers were compared to those from speakers of British English. We expected that British English speakers would produce robustly different VOTs in voiceless stops following /s/ vs. word-initially. On the other hand, we did not expect a robust difference in VOT for Indian English speakers. For VOT, the expectation thus involved the distinction between initial voiceless stops vs. stops after initial /s/. To address these distinctions, we subtracted the VOT of stops after /s/ (in spoons, scoop) from the VOT of initial stops (in peas, kids) for each place of articulation and for each speaker. The resulting VOT difference was our dependent variable in what follows.

Our expectations regarding f0 were complicated by the fact that fundamental frequency may vary with intonational context as well as with laryngeal setting of the preceding consonant. Intonational context of each word is not fully controlled for in our corpus. The target words with initial stops are phrase-final (*peas*, *kids*), but the others (*spoons*, *scoop*) occur in phrase-medial position. Comparing the words with a word-initial stop to words with a stop after /s/, we therefore expected to see both intonational differences and, potentially, differences based on aspiration. The latter effect may be highly local and small in size (cf. Dmitrieva and Dutta 2020, for Marathi) or masked entirely by the intonational effects.

From the point of view of intonation, we may expect phrase-final words to have lower pitch than phrase-medial ones across the board. As discussed in section 1, the effects of aspiration on f0 of the following vowel are often quite subtle but based on the studies of British English we may preliminarily expect that aspirates will have a local raising effect on f0 of the following vowel in this language. This expectation is only tentative since most studies to date compare English voiced unaspirated stops to voiceless aspirated ones, and hence the raising effect on f0 may follow from voicing, not just aspiration (House and Fairbanks 1953, Lehiste and Peterson 1961, Hombert 1976, Ohde 1984, Kingston and Diehl 1994). On the other hand, for Indian English our expectation was that the stops would not differ from each other in aspiration, and therefore the two categories of stops would not show a difference in effects on the f0 of the following vowel.

---

[5] The script is available at: http://phonetics.linguistics.ucla.edu/facilities/acoustic/formant_logging.txt

In sum, we expected to find intonational differences based on context, and these differences were expected to be similar across languages. By contrast, due to effects of aspiration, f0 in the initial portion of the vowel was expected to be raised after aspirates in British English but not in Indian English. Like VOT, this prediction is best addressed by calculating f0 difference between vowels after initial stops vs. stops after /s/. This f0 difference was our dependent variable for statistical modeling.

3 Results

3.1 VOT

Table 6 shows the mean VOT we obtained for British English compared to findings from two previous studies on British English. Given that these other studies included fewer subjects and employed different methodologies, the mean VOT values we obtained for British English are a decent match with them, serving as a sanity check for our measurements and methodology.

Phone	Present study (n=28)	Docherty (1992) (n=5)	Whiteside and Irving (1997) (n=10)
k	25	27	22
k^h	74	62	62
p	14	15	7
p^h	58	42	45

Table 6. Comparison of mean British English VOT (in ms) with previous studies.

Table 7 shows the mean VOT we obtained for Indian English, and compares these values to previous results on VOT in Indic languages. The only comparable study of Indian English itself (Awan and Stine 2011) reports a mean VOT of 33ms for [p^h] – a result that matches our data.[6]

	k	k^h	p	p^h
Bengali (Mikuteit and Reetz 2007)	34	73	14	-
Gujarati (Rami et al. 1999)	41	75	-	-
Hindi (Lisker and Abramson 1964)	16	84	12	63
Marathi (Lisker and Abramson 1964)	21	73	0	35
Nepali (Poon and Mateer 1985)	27	84	6	69
Pahari (Khan and Bukhari 2011)	27	77	19	66
Urdu (Hussain 2018)	35	85	18	73
Indic (average)	29	79	12	61
	post-/s/ k	word-initial k	post-/s/ p	word-initial p
Indian English (present study)	47	47	18	28
British English (present study)	25	74	14	58

Table 7. A comparison of the existing results on VOT (ms) in Indic languages and in British English.
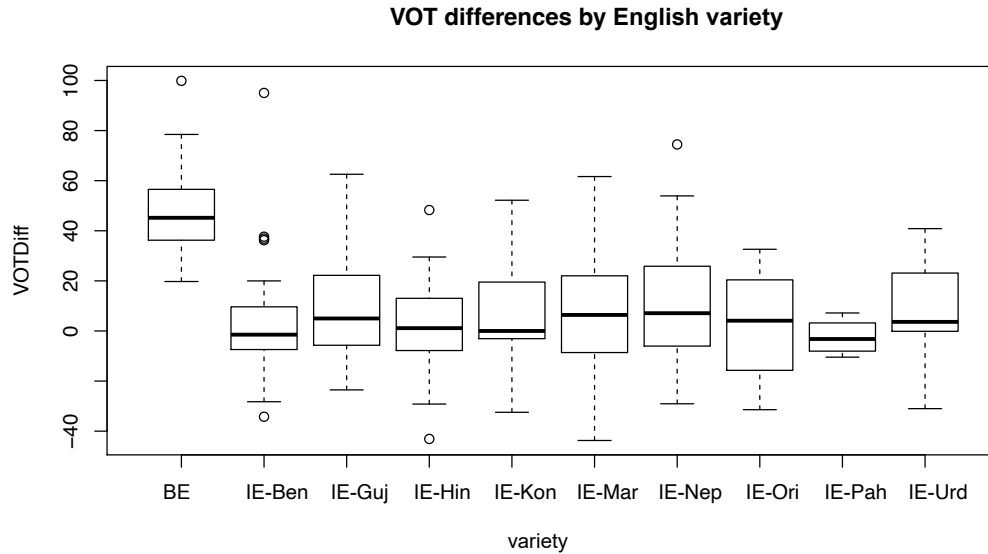
The VOT values for Indian English word-initial stops were close to the VOT of unaspirated stops for a given place of articulation in the Indic languages. Thus, Table 7 shows that when speaking Indian English, Indic speakers produce stop VOTs that more closely approximate Indic VOTs of the unaspirated category.

Furthermore, the differences between initial and post-/s/ stops in Indian English were very small. Figure 1 plots these differences in VOT for each variety of Indian English associated with the corresponding Indic language. It can be observed that all Indian Englishes have a difference between

---

[6] Note that Awan and Stine's (2011) study included speakers of a slightly different set of languages – their set included Punjabi, Kannada, and Telugu but did not include Oriya and Pahari.

initial and post-/s/ stops that is close to zero, while the difference in British English is quite substantial (mean = 46.7ms.; s.d. = 15.3ms).

Figure 1. VOT differences between initial stops and stops after /s/ by English variety.
See section 2.2 for IE variety abbreviations.

VOT differences by English variety

We analyzed the VOT differences with a linear mixed effects regression model using consonant place (dorsal, labial), and speaker language (as self-reported in our dataset) as fixed effects and the speaker as a random effect. Recall that each speaker produced only one token of each word, so we could not separate the effect of speakers from the effect of individual tokens by adding item as a random effect. Our first model, Model 1, summarized in Table 8, included all the Indian Englishes as well as British English. Taking British English VOT differences as baseline, Model 1 found a significant effect for each Indian English variety.

	β	SE	df	t-value	$p(>|t|)$
LanguageBen	−43.3	6.9	203.7	−6.2	<0.001
LanguageGuj	−43.9	7.4	202.5	−6	<0.001
LanguageHin	−51.4	6	205.6	−8.5	<0.001
LanguageKon	−69.1	15.1	208.1	−4.6	<0.001
LanguageMar	−46.7	8.3	204.6	−5.6	<0.001
LanguageNep	−40.4	7.4	204	−5.5	<0.001
LanguageOri	−60.8	15.2	202.5	−4	<0.001
LanguagePah	−52.4	15.2	202.5	−3.5	<0.001
LanguageUrd	−40.9	7.1	203.8	−5.7	<0.001
placeLab	−4.997	5	99.5	−1	n.s.
LanguageBen:placeLab	0.3	8.8	102	0.03	n.s.
LanguageGuj:placeLab	13.3	9.4	99.5	1.4	n.s.
LanguageHin:placeLab	13.3	7.5	106.9	1.8	<0.1
LanguageKon:placeLab	48.8	18.1	119	2.7	<0.01
LanguageMar:placeLab	13	10.4	104.1	1.2	n.s.
LanguageNep:placeLab	13.4	9.3	102.7	1.4	n.s.
LanguageOri:placeLab	32.9	19.4	99.5	1.7	<0.1
LanguagePah:placeLab	6.6	19.4	99.5	0.3	n.s.
LanguageUrd:placeLab	4.4	9	102.3	0.5	n.s.

Table 8. Model 1 results for VOT differences among British English and Indic languages.

In addition, taking dorsal place as baseline, Model 1 found a significant interaction between labial place and the language being Konkani. These interactions were marginal for Hindi and Oriya. We only have two speakers for Oriya and three for Konkani, so the effects for these languages still need to be tested with a larger dataset. A marginal interaction between language being Hindi and labial place was somewhat unexpected, but note that VOT differences in Hindi were still very small (mean –3.4ms for dorsals; 6.1ms for labials).

Although Model 1 supported the difference between British and Indian English as two groups, the potential differences between individual Indian Englishes were harder to assess with this model. Therefore, we ran a second model which included the same factors but addressed only data from Indian English speakers, excluding British English speakers. Model 2, in which Bengali was the baseline, is summarized in Table 9. Unlike Model 1, this model found no significant main effect of language. In fact, in Model 2 only the interaction of language being Konkani with labial place turned out to be significant. Thus, we did not find robust differences in VOT between the different Indian Englishes based on speaker language or consonant place. The significant interaction for Konkani should be interpreted with caution since we only have three speakers for this language.

	β	SE	df	t-value	$p(>|t|)$
LanguageGuj	–0.53	9.1	150.3	–0.1	n.s.
LanguageHin	–8	7.9	152.2	–1	n.s.
LanguageKon	–26	16.8	154.2	–1.5	n.s.
LanguageMar	–3.3	10	151.5	–0.3	n.s.
LanguageNep	2.9	9.1	151.2	0.3	n.s.
LanguageOri	–17.5	16.9	149.8	–1	n.s.
LanguagePah	–9	16.9	149.8	–0.5	n.s.
LanguageUrd	2.4	8.9	151.1	0.3	n.s.
placeLab	–4.7	7.7	76.1	–0.6	n.s.
LanguageGuj:placeLab	13	11.5	74.6	1.1	n.s.
LanguageHin:placeLab	13	9.8	78.8	1.3	n.s.
LanguageKon:placeLab	48.7	20.1	86.9	2.4	<0.05
LanguageMar:placeLab	12.7	12.5	77.1	1	n.s.
LanguageNep:placeLab	13.1	11.4	76.4	1.1	n.s.
LanguageOri:placeLab	32.6	21.4	73.7	1.5	n.s.
LanguagePah:placeLab	6.2	21.4	73.7	0.3	n.s.
LanguageUrd:placeLab	4.1	11.2	76.2	0.3	n.s.

Table 9. Model 2 results for VOT differences among Indic languages.

To check if the model with only Indian Englishes would have found a significant difference with a different baseline language, we ran a model comparison of Model 2 with one that omitted that effect of language. This model comparison is shown in Table 10. This comparison shows that including the effect of language does not improve the model fit. That is, the model with language as a main effect is not significantly different from one in which the effect of language is not included. This indicates that VOT difference is not affected by the linguistic background of the Indian English speakers.

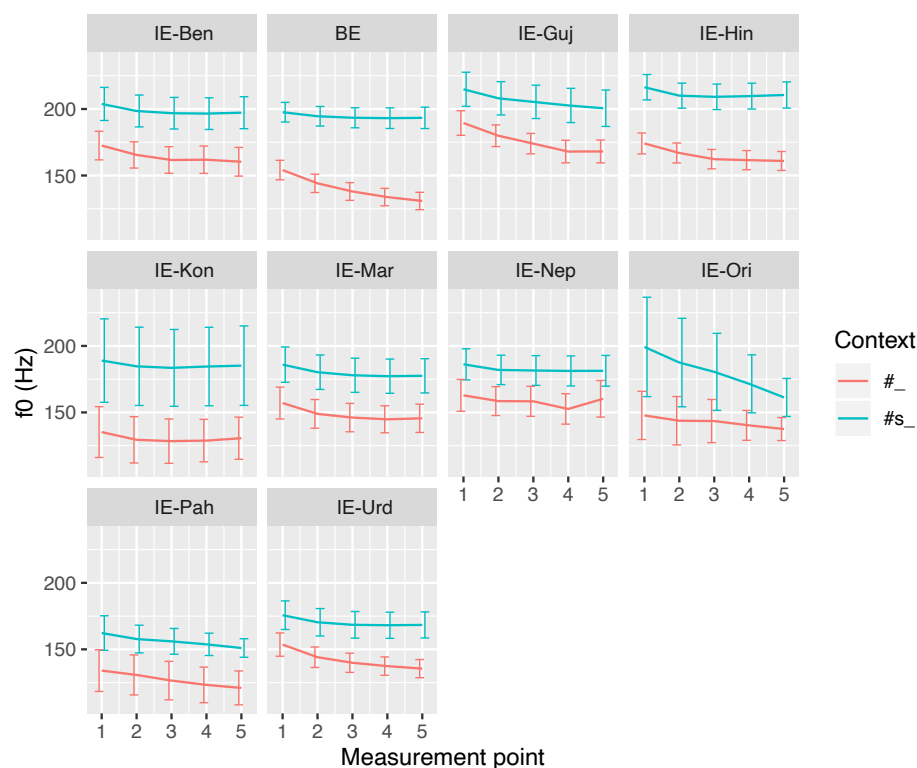| df2	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
model2	4	1564.75	1577.37	–778.38	1556.75	*NA*	*NA*	*NA*
baseline	20	1582.93	1645.99	–771.56	1542.93	13.83	16	0.61

Table 10. Model comparison results for VOT differences among Indic languages.

To summarize, the VOT distinction between aspirated and plain stops was sizeable in our British English data but close to zero in Indian English data (see Figure 1). British English was also significantly different from Indian English as a whole, but the Indian Englishes did not differ from each other in terms of stop VOT distinctions. Our complete dataset also exhibited an interaction of place with the language being Konkani.
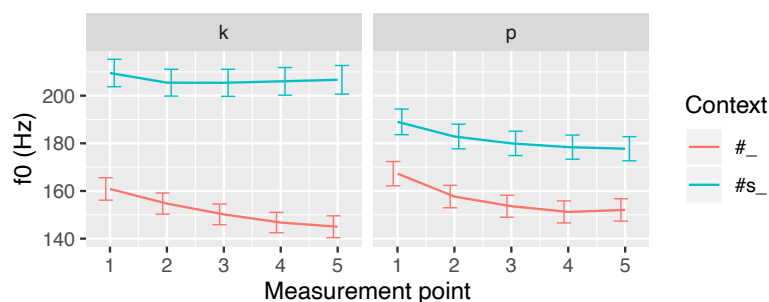
3.2 Fundamental frequency

Figure 2 summarizes mean f0 within the first half of vowel duration by language. Here British English did not appear to differ from Indian English in how vowel f0 patterns with aspiration. For all languages, the overall mean f0 after word-initial stops (aspirated in British English) was lower than that after post-/s/ stops (unaspirated in British English), and the size of the difference was comparable across languages. Relatively large standard error (seen in error bars) for Oriya, Konkani, and Pahari follows from a small size of our corpus for these languages.

Figure 2. Mean f0 values for each measurement point by language. Error bars show standard error. See section 2.2 for language name abbreviations.

The variation in vowel f0 by place of the preceding consonant is examined in Figure 3. Here, the difference between f0 after the two kinds of stops seems larger after /k/ than after /p/. f0 in the words scoop and spoons is also higher than in the words kids and peas respectively. It is likely that this is due to the effect of prosodic context. The words kids and peas are phrase-final in the passage we analyzed, and therefore it is expected that these words may have a lower pitch than the phrase-medial scoop and spoons (Pierrehumbert 1980, Beckman and Pierrehumbert 1986).

Figure 3. Mean f0 values for each measurement point. Error bars show standard error.

The greater f0 difference between vowels following /k/ on the one hand and /p/ on the other, however, was surprising. It is possible that small differences in the prosodic properties of the contexts in which the words occurred led to this difference. The word *kids* was almost always produced at the end of an intonational phrase (IP) while *peas* was variably produced at the end of an intonational or intermediate phrase (ip). There were many instances where *peas* was produced with a rising ip boundary tone typically associated with listed items. Moreover, the word *scoop* was produced with a pitch accent in most cases, but the word *spoons* did not always receive a pitch accent. These are post-hoc speculations and need to be confirmed with quantitative analyses of our data. The conclusion we want to draw at this point is that, predictably, intonational effects had a notable effect on f0 differences.

Additionally, as expected, intonational effects had an influence on the entire first half of the vowel, rather than a localized effect at the onset of the vowel. Thus, an examination of the data suggests that the intonational context of our items, perhaps unsurprisingly, affected vowel f0. On the other hand, there was no clear indication of an effect of context on f0 – the patterning of f0 by context did not seem to differ substantially by language, although for some varieties of Indian English the difference between vowel f0 after initial and post-/s/ stops appeared to be smaller than in English. It may also be the case that local f0 effects based on context were masked by larger-scale intonational effects. However, note that since we did not find any differences in aspiration by context (word-initial versus post-/s/), comparable f0 following these stops was expected in the case of Indian English.

Recall that we expected a potential local raising of f0 with aspiration in British English. This, however, was not borne out. Figure 2 shows that vowel f0 after initial stops was consistently lower than vowel f0 after post-/s/ stops. We also predicted a smaller difference in f0 by aspiration in Indian English than in British. In order to assess this latter prediction, we ran a linear mixed effects regression model (Model 3) analyzing the f0 differences by aspiration. The dependent variable for this model is the difference in f0 between vowels after initial stops versus vowels after post-/s/ stops. This difference is negative for all languages analyzed, as seen in Figure 2. Our model targets f0 differences at the very first measurement point since previous research shows that aspiration effects on f0 are local, appearing to be the largest early on and quickly fading away (Dmitrieva and Dutta 2020). The model included speaker native language (with British English as baseline) and place (with /k/ as baseline) as main effects and individual speakers as a random effect. Model 3 results are presented in Table 11.

	β	SE	df	t-value	$p(>|t|)$
LanguageBen	23.5	12.8	204.3	1.8	<0.1
LanguageGuj	30.8	13.2	200.7	2.3	<0.05
LanguageHin	−1.8	10.2	202.2	−0.2	n.s.
LanguageKon	16.4	27	209.5	0.6	n.s.
LanguageMar	15.3	15.7	204.3	1	n.s.
LanguageNep	24.9	13.2	203	1.9	<0.1
LanguageOri	−10.4	27.2	200.7	−0.4	n.s.
LanguagePah	27.1	27.2	200.7	1	n.s.
LanguageUrd	24.6	12.8	202.7	1.9	<0.1
placeLab	32.1	8.7	103.8	3.7	<0.001
LanguageBen:placeLab	−18.1	15.6	109	−1.2	n.s.
LanguageGuj:placeLab	−25.4	16.4	103.8	−1.6	n.s.
LanguageHin:placeLab	7.9	12.7	106.9	0.6	n.s.
LanguageKon:placeLab	−16.3	31.6	122.7	−0.5	n.s.
LanguageMar:placeLab	−9.8	19.1	109	−0.5	n.s.
LanguageNep:placeLab	−7.4	16.2	106.8	−0.5	n.s.
LanguageOri:placeLab	4.9	33.7	103.8	0.1	n.s.
LanguagePah:placeLab	−23.6	33.7	103.8	−0.7	n.s.
LanguageUrd:placeLab	−7.8	15.7	106.5	−0.5	n.s.

Table 11. Model 3 results for differences in vowel f0 at the first measuring point.

Model 3 found a significant effect of labial place, as expected based on Figure 3 since the f0 difference for /p/ was smaller than for /k/. A significant main effect of language was found for Gujarati; Bengali, Nepali, and Urdu are marginal. These effects seem to go in the predicted direction since the f0 difference was smaller (closer to zero) for these languages than for British English. However, it seems premature to conclude that a smaller f0 difference in some Indian Englishes is due to the implementation of aspiration because f0 differences are affected by a number of additional factors in our dataset, and because there is no consistent effect across the Indian Englishes.

4 Discussion

We compared the VOT of word-initial and post-/s/ stops in Indian English, as well as the influence of these stops on f0 of the following vowel. Both of these acoustic variables may be affected by aspiration, and in addition f0 is affected by a number of contextual factors. VOT is a primary cue to aspiration, and it was found that British English exhibited a substantial difference in VOT between aspirated (word-initial) and unaspirated (post-/s/) stops while all Indian Englishes seem to have a very small VOT difference (Figure 1). In line with this, Model 1 (Table 8) that included British English as baseline and compared it to Indian Englishes found a significant effect of language while Model 2 (Table 9) that included only Indian varieties of English found no main effect of language. These findings are in line with our initial expectation that word-initial stops are unaspirated in Indian English. In this way, our study serves to document the implementation of Indian English stops with corpus data.

In our dataset, the size of aspiration difference by place also seems to vary, and in particular Model 1 (Table 8) finds a significant place x language interaction for Konkani and a marginal one for Hindi and Oriya. We believe that more data would be needed to properly interpret these results, since especially for Konkani and Oriya our corpus has limited data.

f0 of the vowel following stops was found to be lower for word-initial stops than for post-/s/ stops for all varieties of English in our study (see Figure 2). This was particularly unexpected for British English where previous studies lead one to expect a potential raising effect of aspiration on f0. We believe our findings are likely explained by contextual effects, and this is indirectly confirmed by the fact that f0 in the words *kids* and *peas* is lower than for *scoop* and *spoons* at all measurement points, not just at the onset of the vowel. The words *kids* and *peas* are phrase-final in the passage we analyzed, and they may have a lower pitch than the phrase-medial *scoop* and *spoons* (Pierrehumbert 1980, Beckman and Pierrehumbert 1986).

In sum, the direction of f0 difference between vowels after aspirates and vowels after unaspirated stops in British English was not consistent with the expected aspiration effect. We suspect that the contextual influences described above may be masking the more subtle effect of aspiration. More generally, the f0 difference between initial stops and post-/s/ stops varies with place of articulation of the stop and, to some extent, with the Indic language in our dataset. Figure 3 suggests a larger difference for vowels after /k/ than for vowels after /p/, and this difference corresponds to a significant effect of place on the f0 differences at the first measurement point in Model 3 (Table 11). We do not have an immediate explanation for this place effect, and leave this as a topic of investigation for future research.

As for the other language spoken by the participants, several Indian Englishes show a smaller f0 difference than British English. Model 3 (Table 11) found a significant effect of language for Gujarati, and a marginal one for Bengali, Nepali, and Urdu. A smaller difference in vowel f0 would be expected based on the fact that Indian English stops are unaspirated, but given the overall patterning of our f0 data we do not think these effects support this conclusion.

To summarize, our analysis of vowel f0 did not find support for the expected differences. We take this null result to suggest that aspiration effects on f0 are very subtle, and hence likely undiscoverable in a corpus study with lots of additional variables like ours. The effect of aspiration on f0 needs to be reevaluated in a study that controls for intonational context and vowel quality.

In terms of VOT, present-day Indian English exhibits an extremely interesting pattern where both categories of English stops were historically adapted to the native unaspirated category, despite the fact that Indic languages also have aspirates. Existing impressionistic reports also suggest that this pattern extends to loan adaptation (Paradis and LaCharité 1997). In parallel to this pattern, the native VOT of stops in Indic languages differs substantially from that of most varieties of English:

Indic aspirated stops have a much longer VOT than English aspirated stops, while Indic unaspirated stops have VOT values comparable to those in English (see section 2.3). While the VOT of aspirated English stops does not perfectly match those of unaspirated Indic stops, it seems plausible that English aspirated stops could have historically been perceived as unaspirated by a native Indic listener. That is, such a listener might have found English stops to not be 'aspirated enough' by Indic acoustic benchmarks. In other words, it is possible that the category boundary for aspirated versus unaspirated stops in Indic was, at least historically, higher than that in English.

Sailaja (2009) suggests that the lack of aspiration in Indian English voiceless stops may also be conditioned in part by the lack of aspiration marking in English orthography. While this explanation is plausible, it is unlikely to be the only factor involved. There is evidence suggesting that perceptual assimilation plays a role in the early stages of second language acquisition (Best 1995, Flege 1995, Best and Tyler 2007) and loanword adaptation (Silverman 1992, Peperkamp 2005, Peperkamp et al. 2008, Boersma et al. 2009). While orthography has also been shown to be a contributing factor (Vendelin and Peperkamp 2006), it has been shown to be a factor when perception is indeterminate (Daland et al. 2015). Thus, it is possible that both our account and Sailaja's are correct. The acoustic mismatch we observed in our study likely contributed to a perceptual mismatch which led to Indian English speakers referencing the orthography to resolve this resulting ambiguity.

It might be useful to consider the adaptation of English dental fricatives in Indian English to evaluate the effect orthography has on adaptation patterns. The English interdental fricatives /θ/ and /ð/ are adapted as dental stops /t̪/ or /t̪ʰ/ and /d̪/ respectively (Kelkar 1957, Sailaja 2009). Notice that both these phonemes are represented in the orthography with "th" (e.g., as in *think* and *this* respectively). While the adaptation of /θɪŋk/ as [t̪ʰɪŋk] can be explained by the orthography, the explanation for /ðɪs/ being adapted as [d̪ɪs] is potentially problematic for this account. An acoustic account, however, can explain both these adaptation patterns. In the adaptation of /θ/ as [t̪ʰ], aspiration was employed by Indic speakers to preserve the acoustic friction present in the English fricative. This is also observed in the adaptation of /f/ as [pʰ] in some Indian Englishes (Kelkar 1957). Additionally, the reason that /ð/ isn't adapted as [d̪ʰ] likely has to do with conflicting f0 cues. Lowering of f0 is especially prominent following breathy stops like [d̪ʰ] (Dmitrieva and Dutta 2020, Schertz and Khan 2020), and could be a major contributing factor to the perceptual identification of breathy stops (Dmitrieva and Dutta 2020, Schertz and Khan 2020). This acoustic difference between Indic plain voiced and breathy voiced stops leads to the adaptation of /ðɪs/ as [d̪ɪs]. Thus, in this case, due the greater acoustic mismatch, there was no resulting perceptual ambiguity. Listeners, therefore, did not consult the orthography leading to the lack of aspiration in such words that are spelled with <h>.

This account, which assumes assimilation originating in acoustic similarity, can also explain why Indian English speakers who are native speakers of languages that do not have contrastive aspirates such as Ao, Angami, Mizo, and Tamil do produce aspiration in the appropriate contexts where it appears in "native" varieties of English (Wiltshire 2005, Wiltshire and Harnsberger 2006). The unpredictable adaptation of /θ/ as either aspirated or unaspirated is reflective of the competing acoustic parameters – f0, whose lowering is less pronounced after voiceless, aspirated stops, and the attempt to preserve friction. In summary, while we do not completely rule out the influence of orthography on Indian English, we do point out that specific phonetic differences could have contributed to the adaptation pattern seen in Indian English.

Thus, native Indic phonetics seems to have affected Indian English in a predictable way. As noted previously, this pattern also reportedly extends to loanword adaptation, and here the observed correlation between native phonetics and borrowed adaptations is particularly surprising from the point of view of the feature-based theories of loanword adaptation (Paradis and LaCharité 1997, LaCharité and Paradis 2005). If cross-language adaptation is based on features and contrast, we might expect that the native Indic contrasts would be mapped onto the English contrast system in loans, and in Indian English. Thus, a case where this apparently does not happen is a potential challenge to the feature-based theories. On the other hand, the perception-based approaches to loanwords and non-native phonetics (Silverman 1992, Yip 1993, Peperkamp et al. 2008) make predictions about the perceptual processing of English aspirates in Indic speakers, and these predictions can be tested more directly in future perceptual experiments.

Given our results, which are undeniably only half the puzzle, we propose that in the early stages of Indian English, listeners encountered an acoustic mismatch between their native Indic categories

and those in the source language (some variety of British English).[7] Perceptual ambiguity resulted from the mismatch in VOT of word-initial stops in the source language and the VOT of aspirated stops in the listeners' native Indic languages. Since phonological parsing by means of perceptual factors alone was indeterminate, English orthography might have influenced the adaptation of voiceless stops in Indian English as proposed by Daland et al. (2015). Needless to say, this proposed perceptual adaptation pattern needs to be confirmed by a perceptual experiment. It must also be noted that since this pattern presumably occurred in the early stages of Indian English, it is possible that the perceptual category boundaries of the present-day Indian English speaker have shifted, and are, therefore, not a reliable metric to assess this proposed adaptation mechanism.

5 Conclusion

This study documents the acoustics of voiceless stops in Indian English in comparison to British English. Our main finding, in line with previous reports and findings, is that speakers of Indian English show a very small difference in VOT between initial stops and stops after /s/ while British speakers show quite a substantial difference for these stops in the same context. The absolute VOT values of Indian English stops approximate the values for unaspirated stops, and thus our corpus shows evidence of Indian English voiceless stops being always produced without aspiration as suggested by some impressionistic and other acoustic studies. On the other hand, we did not find any evidence to suggest that f0 effects of aspiration may be different in British vs. Indian English.

Our study also shows how naturalistic corpus data can be used to document subtle phonetic phenomena in cross-language adaptation and phonetics. Of course, the design of our study and our reliance on corpus data comes with certain limitations. We hope that this research will stimulate further studies addressing these issues. First, we acknowledge that we do not know if the adaptation pattern described here happens as part of active synchronic grammar of Indian English or exists as a historical pattern established and conventionalized for the existing words but not necessarily implemented grammatically. (Flege and Eefting 1987 make a similar point for Puerto Rican Spanish.) Indian English is a conventionalized variety, and many present speakers acquire it without being exposed to British or American English, and thus it may be that the adaptation of aspirated stops as unaspirated was present as a grammatical pressure only when Indian English was established, and is now carried over as a property of this variety. Future studies with non-words may clarify this.

Second, the acoustic data may not be sufficient to establish the correct featural representation of voiceless stops in Indian English. It could be, for example, that Indian English distinguishes two categories of stops based on the feature [voice] rather than on aspiration features, but in order to confirm this hypothesis we may need additional data on the acoustics of Indian English voiced stops, and potentially on the perception of this contrast.

Finally, although our data suggest that Indic laryngeal phonetics affects the implementation of VOT in Indian English, our study does not allow us to exclude other potential factors such as orthography. This may need to be investigated further, as it is a potential contributing factor along with native phonetics. Our study also highlights the complex interactions between linguistic and extra-linguistic factors in loanword adaptation and non-native perception.

Acknowledgements

---

[7] It might be appropriate to view this early stage through the predictions made by models of non-native speech perception, such as the Perceptual Assimilation Model (or PAM) (Best 1995), and L2 acquisition, such as the Speech Learning Model (Flege 1995) and PAM-L2 (Best and Tyler 2007). See Narkar (2021) for a discussion of the Indian English adaptation pattern in terms of PAM.

References

Acharya, Jayaraj. 1991. *A descriptive grammar of Nepali and an analyzed corpus.* Georgetown University Press.

Alderete, John and Alexei Kochetov. 2009. Japanese mimetic palatalization revisited: Implications for conflicting directionality. *Phonology* 26(3):369–388.

Almeida, Matthew. 2003. *A description of Konkani*. Ph.D. thesis, UMI Ann Arbor.

Awan, Shaheen N. and Carolyn L, Stine. 2011. Voice onset time in Indian English-accented speech. *Clinical linguistics & phonetics* 25(11-12):998–1003.

Beckman, Mary E. and Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. Ph*onology* 3:255–309.

Benders, Titia, Sujal Pokharel, and Katherine Demuth. 2019. Hypo-articulation of the four-way voicing contrast in Nepali infant-directed speech. *Language Learning and Development* 15(3):232–254.

Benguerel, A.-P. and T. K. Bhatia. 1980. Hindi stop consonants: an acoustic and fiberscopic study. *Phonetica* 37:134–148.

Berkson, Kelly. 2012. Capturing breathy voice: Durational measures of oral stops in Marathi. *Kansas Working Papers in Linguistics* 33.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. *Speech perception and linguistic experience*:171–206.

Best, Catherine T. and Michael D. Tyler. 2007. Nonnative and second-language speech. *Language experience in second language speech learning: In honor of James Emil Flege* 17:13.

Boersma, Paul. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences* 17:97–110.

Boersma, Paul and Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. *Loanword Phonology*: 11–58.

Boersma, Paul and David Weenink. 2019. Praat: Doing phonetics by computer. Software.

Clements, George N. and Rajesh Khatiwada. 2007. Phonetic realization of contrastively aspirated affricates in Nepali. In *Proceedings of the 16th International Congress of Phonetic Sciences*: 629–632.

Coelho, Gail M. 1997. Anglo-Indian English: A nativized variety of Indian English. *Language in Society* 26(4):561–589.

Daland, Robert, Mira Oh, and Syejeong Kim. 2015. When in doubt, read the instructions: Orthographic effects in loanword adaptation. *Lingua* 159:70–92.

Davis, Katharine. 1994. Stop voicing in Hindi. *Journal of Phonetics* 22(2):177–193.

Dhongde, Ramesh Vaman and Kashi Wali. 2009. *Marathi*, vol. 13. John Benjamins Publishing.

Dmitrieva, Olga and Indranil Dutta. 2020. Acoustic correlates of the four-way laryngeal contrast in Marathi. *Phonetica* 77:209–237.

Docherty, Gerard J. 1992. *The timing of voicing in British English obstruents*. New York: Foris.

Dutta, Indranil. 2007. *Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt*. PhD dissertation, University of Illinois at Urbana-Champaign Urbana, IL.

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research* 92:233–277.

Flege, James E. and Wieke Eefting. 1987. Production and perception of English stops by native Spanish speakers. Journal of Phonetics 15(1):67–83. Hauser, Ivy. 2019. Effects of phonological contrast on within-category phonetic variation. Doctoral Disser- tations. 1673.

Hoffmann, Sebastian, Marianne Hundt, and Joybrato Mukherjee. 2011. Indian English an emerging epicentre? A pilot study on light verbs in web-derived corpora of South Asian Englishes. *Anglia-Zeitschrift für englische Philologie* 129(3-4):258–280.

Hombert, Jean-Marie. 1976. The effect of aspiration on the fundamental frequency of the following vowel. In *Annual Meeting of the Berkeley Linguistics Society*, vol. 2: 212–219.

House, Arthur S. and Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25(1):105–113.

Hussain, Qandeel. 2018. A typological study of voice onset time (VOT) in Indo-Iranian languages. *Journal of Phonetics* 71:284–305.

Iverson, Gregory and Joseph Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12(3):369–396.

Jouanne, Thomas Christian Nicolas Marie. 2014. *A preliminary analysis of the phonological system of the western Pahari language of Kvar.* Master's thesis, University of Oslo.

Kawahara, Shigeto. 2011. Japanese loanword devoicing revisited: A rating study. *Natural Language and Linguistic Theory* 29(3):705–723.

Kelkar, Ashok R. 1957. "Marathi English": a study in foreign accent. *Word* 13(2):268–282.

Khan, Abdul Qadir and Nadeem Haider Bukhari. 2011. An acoustic study of VOT in Pahari stops. *Kashmir Journal of Language Research* 14(1).

Khan, Sameer ud Dowla. 2010. Bengali (Bangladeshi Standard). *Journal of the International Phonetic Association* 40(2):221–225.

Kingston, John and Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70(3):419–454.

LaCharité, Darlene and Carole Paradis. 2000. Phonological evidence for the bilingualism of borrowers. In *Proceedings of the 2000 Annual Conference of the Canadian Linguistics Association*: 221–232. Cahiers linguistiques d'Ottawa Ottawa, Ontario.

LaCharité, Darlene and Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36(2):223–258.

Lehiste, Ilse and Gordon E. Peterson. 1961. Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America* 33(4):419–425.

Lisker, Leigh and Arthur Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20:384–422.

Majumadāra, Pareśacandra. 1970. *A Historical Phonology of Oṛiyā*, vol. 66. Calcutta: Sanskrit College.

Mikuteit, Simone and Henning Reetz. 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech* 50(2):247–277.

Mistry, Purushottam J. 1997. Gujarati phonology. *Phonologies of Asia and Africa* 2:653–673.

Nair, N. Gopalakrishnan. 1996. *Indian English phonology: A case study of Malayalee English*. Prestige Books.

Narkar, Jahnavi. 2021. The curious absence of aspiration in Indian English: The role of phonetics in adaptation. In *37th West Coast Conference on Formal Linguistics*: 189–196. Cascadilla Proceedings Project.

Ohala, Manjari and John Ohala. 1972. The problem of aspiration in Hindi phonetics. *Project on Linguistic Analysis Reports*: 63–70.

Ohde, Ralph N. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America* 75(1):224–230.

Paradis, Carole and Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33(2):379–430.

Paradis, Carole and Antoine Tremblay. 2009. Nondistinctive features in loanword adaptation. In C. A. Wetzels (Ed.), *Loan Phonology. Current Issues in linguistics Theory* 307:211–224.

Peperkamp, Sharon. 2005. A psycholinguistic theory of loanword adaptations. In *Annual Meeting of the Berkeley Linguistics Society*, vol. 30: 341–352.

Peperkamp, Sharon, Inga Vendelin, and Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: experimental evidence from Japanese. *Phonology* 25(1):129–164.

Pierrehumbert, Janet Breckenridge. 1980. *The phonology and phonetics of English intonation*. Ph.D. thesis, Massachusetts Institute of Technology.

Poon, Pamela G. and Catherine A. Mateer. 1985. A study of VOT in Nepali stop consonants. *Phonetica* 42:39–47.

Rami, Manish K., Joseph Kalinowski, Andrew Stuart, and Michael P. Rastatter. 1999. Voice onset times and burst frequencies of four velar stop consonants in Gujarati. *The Journal of the Acoustical Society of America* 106(6):3736–3738.

Sailaja, Pinjali. 2009. *Indian English*. Edinburgh: Edinburgh University Press.

Schertz, Jessamyn and Sarah Khan. 2020. Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. *Journal of Phonetics* 81:100979.

Schwarz, Martha, Morgan Sonderegger, and Heather Goad. 2019. Realization and representation of Nepali laryngeal contrasts: Voiced aspirates and laryngeal realism. *Journal of Phonetics* 73:113–127.

Shimizu, Katsumasa. 1989. A cross-language study of voicing contrasts of stops. *Studia phonologica* 23:1–12.

Silverman, Daniel. 1992. Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology* 9(2):289–328.

Upadhyaya, U. Padmanabha. 1972. *Kannada Phonetic Reader*. Mysore, Central Institute of Indian Languages.

Vendelin, Inga and Sharon Peperkamp. 2006. The influence of orthography on loanword adaptations. *Lingua* 116(7):996–1007.

Weinberger, Steven. 2014. The speech accent archive. Data corpus. George Mason University.

Whiteside, Sandra P. and Caroline J. Irving. 1997. Speakers' sex differences in voice onset time: Some preliminary findings. *Perceptual and Motor Skills* 85(2):459–463.

Wiltshire, Caroline R. 2005. The "Indian English" of Tibeto-Burman language speakers. *English World-Wide* 26(3):275–300.

Wiltshire, Caroline R. and James D. Harnsberger. 2006. The influence of Gujarati and Tamil L1s on Indian English: A preliminary study. *World Englishes* 25(1):91–104.

Yip, Moira. 1993. Cantonese loanword phonology and optimality theory. *Journal of East Asian Linguistics* 2(3):261–291.

Zhang, Jie and Yuwen Lai. 2008. Phonological knowledge beyond the lexicon in Taiwanese double reduplication. In Y.-E. Hsiao, H.-C. Hsu, L. Wee, and D.-A. Ho, (eds.), *Interfaces in Chinese Phonology: Festschrift in Honor of Matthew Y. Chen on His 70th Birthday*: 183–222. Taiwan: Academia Sinica.

Erratum for 'An experimental study of Hindi and English perfective interpretation'

SUDHA ARUNACHALAM, *New York University*
ANUBHA KOTHARI, *independent scholar*

The authors of the article entitled above have notified the editors of an error in the version published in volume 4 of *JSAL*. The error is found in Table 9. The following version should be taken as the revised and final version of the table.

Predicate	Partial Completion		Full Completion	
	Simple Verb	Particle (*up*)	Simple Verb	Particle (*up*)
cover	83	67	100	83
eat	83	17	100	100
fill	100	83	100	100
wake (return to initial state)	100	83	100	100

Table 9. Experiment 2b (English): Percent "true" responses, by condition and trial.