

# Vedic Sanskrit compounding as a window into *tough*-movement

DAVIDE MOCCI, *University of Cagliari & University of Genoa*

## ABSTRACT

In the present study, English *tough*-constructions are compared with their analog in Vedic Sanskrit, i.e., the earliest dialectal variety (dating back to ca. 1200-500 BC) of Sanskrit, an old Indo-Aryan language spoken in the Indian subcontinent for over three millennia. It is found out that Vedic Sanskrit *tough*-constructions are codified via a specific type of compound dubbed as ‘bahuvrīhi’. Exploiting recent work on Sanskrit bahuvrīhis, according to which their derivation involves A-movement (more precisely, the movement of an overt phrase from within the bahuvrīhi to a bahuvrīhi-external position where they are case-marked), the present study extends this derivation to the bahuvrīhis that codify *tough*-constructions in Vedic Sanskrit. In this way, the conclusion is reached that *tough*-movement can be reduced to an instance of A-movement in Vedic Sanskrit, thereby contrasting with *tough*-movement in English, which has instead been argued to be a variety of A-bar movement (more precisely, the movement of a silent operator to the left-periphery of the infinitive contained in English *tough*-constructions).

## 1 Background: What is *tough*-movement?

*Tough*-movement has been a classic topic of investigation in generative grammar: not only are there several different analyses of *tough*-movement that reflect the different stages of development of the generative field, but there is also no uncontroversial definition of *tough*-movement that generative linguists may agree upon. That said, generative linguists do agree on the fact that, regardless of the analysis adopted, *tough*-movement yields a *tough*-construction, exemplified in (1) for English.

- (1) These things are easy to reach.

Thus, in *tough*-constructions like (1), a *tough*-type adjective (e.g., *easy*, *difficult*, *tough*, *impossible*) is predicated of a subject (*these things*), which is nonetheless simultaneously interpreted as an internal argument of the infinitive introduced by the *tough*-adjective (e.g., *these things* is interpreted as the theme of the action denoted by *to reach*).<sup>1</sup>

For English, I adopt the so-called ‘base-generation approach’ to *tough*-movement championed by Chomsky (1977), and sanctioned by Keine & Poole (2017, 297-298) and Salzmann (2023, 240). The fundamental ingredients of this approach may be graphically illustrated as follows:

- (2) [These things]<sub>i</sub> are [<sub>AP</sub> easy [<sub>CP</sub> Op<sub>i</sub> [<sub>PRO</sub> to reach *t*<sub>i</sub> ]]]

<sup>1</sup>See Hicks (2017) for a historical overview of *tough*-movement and *tough*-constructions.

According to the analysis in (2), *tough*-movement is the movement of a silent operator (*Op*) from an argument position (where it is assigned a thematic role) to the left periphery of an infinitival clause. Thus, in this analysis, the subject of the matrix clause (*these things*) does not undergo movement: only the silent operator does. More precisely, the remark made above, to the effect that the subject of the matrix clause is understood to be theta-marked by the embedded infinitive (i.e., is interpreted as an internal argument of the embedded infinitive) is nothing more than an illusion: what is theta-marked is again the silent operator, the illusion being ultimately traceable to the fact that the subject of the matrix clause and the silent operator are co-referential (as indicated by co-indexing).<sup>2</sup>

While a detailed explanation of the arguments underpinning the analysis in (2) would take us too far afield, in the present contribution I would like to focus on an aspect of this analysis: i.e., that *tough*-movement is a kind of A-bar movement in (2).<sup>3</sup> Thus, like any instance of A-bar movement, *tough*-movement can cross multiple CPs (3a), and is blocked by an intervening *wh*-element such as *why* (3b):<sup>4</sup>

- (3) a. [A guy like John]<sub>i</sub> is [AP hard [CP *Op*<sub>i</sub> [PRO to imagine [CP *t*<sub>i</sub> [any woman believing [CP *t*<sub>i</sub> [she would agree [CP *t*<sub>i</sub> [PRO to marry *t*<sub>i</sub> ]]]]]]]]]]
- b. ??[A guy like John]<sub>i</sub> is [AP hard [CP *Op*<sub>i</sub> [PRO to imagine [CP *t*<sub>i</sub> [any woman wondering [CP why [she would agree [CP *t*<sub>i</sub> [PRO to marry *t*<sub>i</sub> ]]]]]]]]]]

In what follows I shall test to what extent this conception of *tough*-movement as A-bar movement can be extended to Vedic Sanskrit, i.e., to the earliest dialectal variety (dating back to ca. 1200-500 BC) of Sanskrit, an old Indo-Aryan language constituting the medium of communication in the Indian subcontinent for over three millennia. In this way, the ultimate goal of the present study is a crosslinguistic exercise: can a linguistic phenomenon like *tough*-constructions be given a unified treatment (or - put another way - be treated as a natural class) in languages, like English and Vedic Sanskrit, that vary greatly for date of attestation and geographical area?

## 2 *Tough*-movement and Vedic Sanskrit

In Vedic Sanskrit, *tough*-constructions are codified via compounding. To illustrate this point, let us consider the examples in (4)-(5).

---

<sup>2</sup>For an alternative approach to English *tough*-constructions, whereby *these things* starts out as the complement of *to reach* and subsequently undergoes long-distance movement to the position of subject of *are*, see, among others, Hicks (2009) and Hartman (2012, 84-97).

<sup>3</sup>A-bar movement is the movement of a linguistic expression that terminates in an A-bar position, i.e., a position that does not correspond to any grammatical function (equivalently: a position where no grammatical case is assigned). For ease of exposition, A-bar positions may be identified here with slots in the complementizer field (CP) of the clause, on which see Rizzi (1997).

<sup>4</sup>The examples in (3) are from Hicks (2009, 542). Cf. also Chomsky (1977, 103-104).

- (4) *ubhé* ... [*su*]-[*gátave*]  
 both.NOM.DU.M [easy]-[reach.INF.DAT]  
 ‘Both these things are easy to reach.’ (*Atharvaveda-Śaunakīya* 6.1.3c)<sup>5</sup>
- (5) *kṛdhi* [*su*]-[*hánā*] ... *vṛtrā*  
 make.IMP.AOR.2SG [easy]-[smash.INF.ACC.PL.N] enemy.ACC.PL.N  
 ‘Make [our] enemies easy to smash!’ (*Ṛgveda* 7.25.5c)<sup>6</sup>

In these examples the *tough*-type element, which in languages like English is an adjective (e.g., *easy*), is indeed a bound morpheme (i.e., a morpheme that cannot occur outside compounding) of unclear grammatical category, spelled out as *su-*, which in itself conveys a wide range of meanings: ‘well’, ‘good’, ‘easy’.<sup>7</sup> *su-* combines with an infinitive (*gátave*; *hánā*) to form a compound: *su-gátave* ‘easy to reach’ (4); *su-hánā* ‘easy to smash’ (5). The question now arises as to what empirical motivations support the classification of *su-gátave* and *su-hánā* as compounds. Indeed, Vedic Sanskrit makes a clear distinction between compounds and non-compound words. Let us consider how.

First, while any Vedic Sanskrit non-clitic word bears the pitch accent outside compounding (except for vocatives and verbs of matrix clauses), only one word is accented inside compounding:<sup>8</sup> cf. the following pair.

- (6) a. *amítrāṇām* *sénāh*  
 enemy.GEN.PL.M army.ACC.PL.F  
 ‘The army of [our] enemies.’ (*Atharvaveda-Śaunakīya* 11.9.3c)
- b. *amitra-senām*  
 enemy-army.ACC.SG.F  
 ‘The army of [our] enemies.’ (*Atharvaveda-Śaunakīya* 3.1.3a)

Second, while compound-external nouns and adjectives are inflected in case, gender, and number in Vedic Sanskrit, they lack a case ending - thereby qualifying as stems from a morphological viewpoint - when they are allocated to the left-hand member of a compound (Lowe 2015b, 72, 93): cf. again (6a), where *amitra-* is marked with a genitive plural ending (*amítrāṇām*) and (6b), which features the stem form *amitra-*, with no case-marking.<sup>9</sup>

<sup>5</sup>References to the *Atharvaveda-Śaunakīya* are based on the edition by Kim (2022).

<sup>6</sup>References to the *Ṛgveda* are based on the edition by van Nooten & Holland (1994).

<sup>7</sup>Etymologically, *su-* parallels Ancient Greek *eu-*, which resurfaces in such English words as *eugenics*.

<sup>8</sup>For an overview of accentuation in Vedic Sanskrit, with special reference to the language of the *Ṛgveda*, see Elizarenkova (1995, 250-259).

<sup>9</sup>It is controversial whether the right-hand member of Sanskrit compounds is also to be considered as a stem. If the answer is in the affirmative, the case ending which shows up on the right-hand edge of Sanskrit compounds - e.g., the accusative-marker *-m* in (6b) - should be taken as appended not to the compound’s right-hand member (i.e., [*amitra*]-[*senā*-*m*]), but rather to the entire compound (i.e., [[*amitra*]-[*senā*]-*m*]). For a discussion of the latter analysis, which goes back to the great ancient Indian grammarian Pāṇini (ca. 4th century BC), see Candotti & Pontillo (2019, 31), Mocci (2025, 110-112), and the references quoted therein.

Last, compounds count as one position, just like simple words, for the placement of second position clitics. This means that two sentence-initial nouns or adjectives may precede a second position clitic in Vedic Sanskrit when they combine to form a compound: see (7a), where *asyāḥ* (a second position clitic) follows the compound *dīrghāyus*, made up of *dīrghá-* ‘long’ and *āyus-* ‘life’. Conversely, when they do not form a compound, the clitic must intervene between them, on the understanding that the position following two case-marked nouns or adjectives counts as a third position, hence as an invalid host for the clitic: see (7b), where the clitic *naḥ* is sandwiched between the accusative-marked forms of *dīrghá-* ‘long’ and *āyus-*.<sup>10</sup>

- (7) a. *dīrghá-āyur*      *asyā*      *yáḥ*      *pátir/*      *jīvāti*  
long-life.NOM.SG.M she.GEN.CL wh.NOM.SG.M lord.NOM.SG.M live.SBJV.PRS.3SG  
*śarādaḥ*      *śatám//*  
autumn.ACC.PL.F hundred.ACC.SG.N  
‘(to him) who as her husband will live, long-lived, through a hundred autumns.’  
(tr. Jamison & Brereton 2014; *R̥gveda* 10.85.39cd)
- b. *dīrghám*      *na*      *āyuh*      *prati-búdhyamānā/*      *vayám*  
long.ACC.SG.N we.GEN.CL life.ACC.SG.N back-wake.PTCP.PRS.NOM.PL.M we.NOM  
*túbhyaṃ bali-h̄taḥ*      *syāma//*  
YOU.DAT tribute-bearing.NOM.PL.M be.OPT.PRS.1PL  
‘awakening to meet our long lifetime, may we be tribute-bearers to thee.’ (tr. Whitney 1962; *Atharvaveda-Śaunakīya* 12.1.62cd)

Let us now apply these diagnostics to test the compoundhood of *su-gātave* (4) and *su-hánā* (5). The second diagnostic is unverifiable: although the left-hand member *su-* is not case-marked, the possibility cannot be excluded that *su-* is indeclinable to begin with, which makes it impossible to establish whether *su-* is actually a stem in *su-gātave* and *su-hánā*. The first diagnostic is trivially met: both *su-gātave* and *su-hánā* bear only one accent, namely on their right-hand member.<sup>11</sup> As for the third diagnostic, it cannot be directly verified, because *su-gātave* and *su-hánā* are not attested as co-occurring with second position clitics. Nonetheless, examples like the following make it possible to establish that compounds whose left-hand member is *su-* may occur in sentence-initial position and precede a second position clitic (*naḥ*).

- (8) *su-bhāgān*      *naḥ*      *devāḥ*      *kṛnutā*  
good-share.ACC.PL.M we.GEN.CL god.VOC.PL.M make.IMPV.PRS.2PL

<sup>10</sup>This is indeed an oversimplification. For an overview of the intricacies related to the positioning of clitics in the left-periphery of the Vedic Sanskrit clause, see Hock (1996), Lowe (2014), and Hale (2018).

<sup>11</sup>One may object that *su-* lacks the pitch accent because it is a clitic-like element and not because it is the left-hand member of a compound. However, clitics may cluster around other clitics in Vedic Sanskrit: see Lowe (2014, 35-38) for relevant examples and analysis. Since *su-gātave* and *su-hánā* cannot be interrupted by bona fide clitics, we may safely exclude that *su-* is a clitic in the examples under examination.

*su-rátnān*

good-treasure.ACC.PL.M

‘Do you, O gods, make us have a good share, good treasures.’ (*Rgveda* 10.78.8a)

It is therefore reasonable to maintain that *su-gátave* and *su-hánā*, just like *su-bhāga-* (8), satisfy the third diagnostic. Accordingly, we conclude that *su-gátave* and *su-hánā* are indeed genuine compounds. All in all, the takeaway from this section is that *tough*-constructions are codified via compounding in Vedic Sanskrit, in the sense that the *tough*-type adjective and the infinitive form a compound (*su-gátave*; *su-hánā*) which in turn combines with a compound-external word that is theta-marked by the compound-internal infinitive: *ubhé su-gátave* (4); *su-hánā vṛtrá* (5).

Now, the codification of *tough*-constructions via compounding is expected to be of great importance for the question as to whether *tough*-movement is A-bar movement in Vedic Sanskrit just like in English. This is because the derivation of Vedic Sanskrit compounding has been independently argued to involve a specific type of syntactic movement. The main goal of the present paper is therefore to exploit the derivation of Vedic Sanskrit compounds as a window into the nature of *tough*-movement in this language. Thus, in the next couple of sections (§3-4) I provide a detailed account of how Vedic Sanskrit compounds - including *su-gátave* and *su-hánā* - are derived, which will pave the way for the following reflections on the nature of *tough*-movement as A- vs. A-bar movement (§5).

### 3 On Vedic Sanskrit bahuvrīhi compounds

According to Whitney (1896, §1287a), compounds like *su-gátave* (4) and *su-hánā* (5) are “probably possessives”. In the classification of compounds developed within the Indian grammatical tradition, possessives correspond to so-called ‘bahuvrīhi’ compounds, i.e., exocentric compounds where the relation between the compound members is appositive or subordinate.<sup>12</sup> We assume that Whitney was correct, and categorize *su-gátave* and *su-hánā* as bahuvrīhis accordingly.

In a recent analysis of Vedic Sanskrit bahuvrīhis, Mocci (2022, 2025) argued that the derivation of this compound type involves movement. Thus, below we shall review Mocci’s analysis with an eye to understanding what the instance of movement involved in the derivation of bahuvrīhis may tell us on the nature of *tough*-movement in Vedic Sanskrit.

To begin with, Lowe (2015b) argued that the members of Sanskrit compounds - including bahuvrīhis - are syntactic units of a special sort. He identified these units with an item theorized within Lexical-Functional Grammar: non-projecting words, i.e., heads which, like

<sup>12</sup>The labels used in the indigenous Indian classification of compounds have been adopted in some influential modern linguistic classifications of compounds, including Bauer (2017), who defines bahuvrīhis as ‘compounds which canonically label a part of the whole which the compound denotes’ (Bauer, 2017, 65). For an insightful overview of the Indian grammatical classification of compounds, see Radicchi (1985–1988), Candotti & Pontillo (2019). See instead Pontillo (2021), Candotti & Pontillo (2022) for a critical discussion of the differences and similarities between the indigenous Indian analysis of bahuvrīhis (with special reference to Pāṇini’s grammar) and modern linguistic analyses of this compound class.



b. [ [AP su- ] [NP hánā ] ]

Nonetheless, the discovery of an ordering constraint on Sanskrit bahuvrīhis forces us to complicate the simple structures in (10), which merely consist of the merger of an AP with NP. The ordering constraint states that, if the bahuvrīhi's 'external referent' is interpreted as a dependent of the bahuvrīhi member  $\alpha$ ,  $\alpha$  is allocated to the right-hand slot of the bahuvrīhi (Mocci 2022, 2025). For example, (11a) contrasts with (11b) in that the qualifier precedes the qualificand only in (11a), and this contrast correlates with the fact that the bahuvrīhi's external referent (i.e., *ráthena* and *devāḥ*) is understood as a dependent (qua possessor) of the qualificand (*ásva-* 'horse') in (11a), but of the qualifier (*pitṛ-*) 'father' in (11b).<sup>15</sup> The dependency between the external referent and the qualificand or qualifier is emphasized in bold in (11a-b).

- (11) a. *vṛṣaṇ-ásvéna* ... *ráthena*  
 bull-horse.INS.SG.M chariot.INS.SG.M  
 Qualifier-**Qualificand** Possessor.of.**Qualificand**  
 'With the chariot whose horses are like bulls.' (*Rgveda* 8.20.10ab)
- b. *tvát-pitārah* ... *devāḥ*  
 you-father.NOM.PL.M god.NOM.PL.M  
 Qualificand-**Qualifier** Possessor.of.**Qualifier**  
 'The gods whose father you are.' (*Taittirīya Saṃhitā* 1.5.10.2)<sup>16</sup>

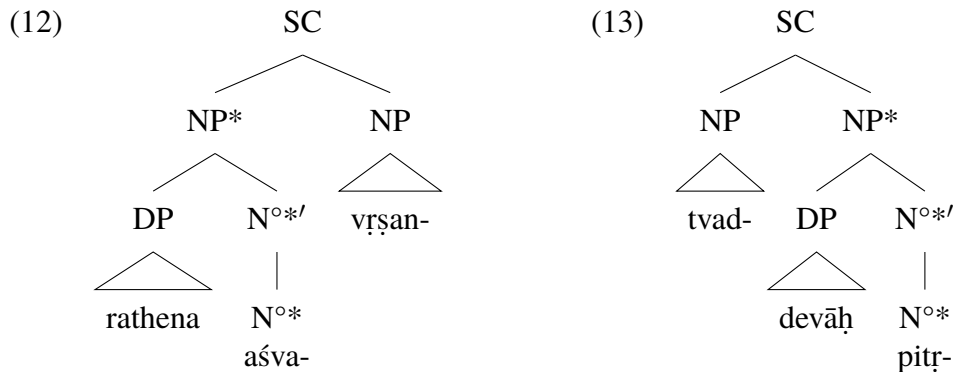
Thus, the structures in (10) must be complicated to capture the ordering constraint at stake. Following Mocci (2022, 2025), the complication required here is a direct consequence of the syntactic codification of two semantic relations involved in Sanskrit bahuvrīhis. Let us consider how, by focusing on the bahuvrīhis of (11a-b).

To start with, there is a qualificand-qualifier relation between the bahuvrīhi members (e.g., between *tvád-* 'you' and *pitṛ-* 'father'), and a possessum-possessor relation between the right-hand member of the bahuvrīhi and the bahuvrīhi's external referent (e.g., between *pitṛ-* and *devāḥ* 'gods'). Following den Dikken (2006, 10-12), the qualificand-qualifier relation is ultimately reducible to a subject-predicate link, which is codified in syntax via a small clause (Moro 2019), notated as SC. Moreover, in keeping with Adger (2013, 113-132), the possessum-possessor relation is codified in syntax via adjunction, which is non-distinguishable from a specifier-head configuration for our purposes (Kayne 1994, 17, Manzini 2017, 237). This yields the structures in (12)-(13) for the bahuvrīhis in (11a-b), respectively: here the possessor is the specifier of the possessum, and the predicate forms a small clause

<sup>15</sup>The possessum-possessor relationship holding between *ráthena* and *ásva-* in (11a), and between *devāḥ* and *pitṛ-* in (11b), respectively parallels the relationship between *ráthasya* and *ásvāḥ* in *ráthasya ásvāḥ* 'the chariot's horses', and between *devānām* and *pitā* in *devānām pitā* 'father of the gods': in these phrases the semantic dependency between the possessor (*ráthasya*; *devānām*) and the possessum (*ásvāḥ*; *pitā*) is overtly (i.e., morphosyntactically) signaled by the genitive case-marking on the possessor.

<sup>16</sup>References to the *Taittirīya Saṃhitā* are based on the digital edition by Fushimi (2012), following Weber (1871–1872).

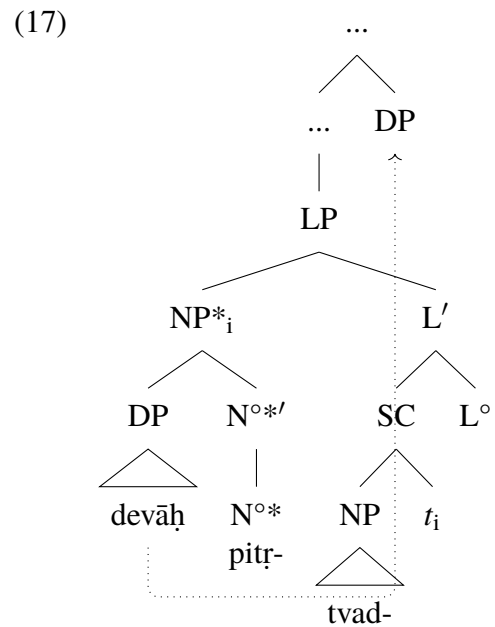
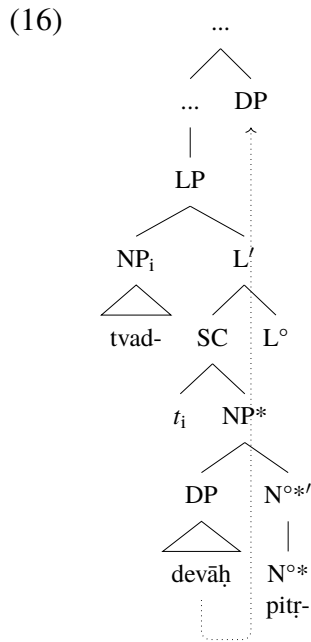
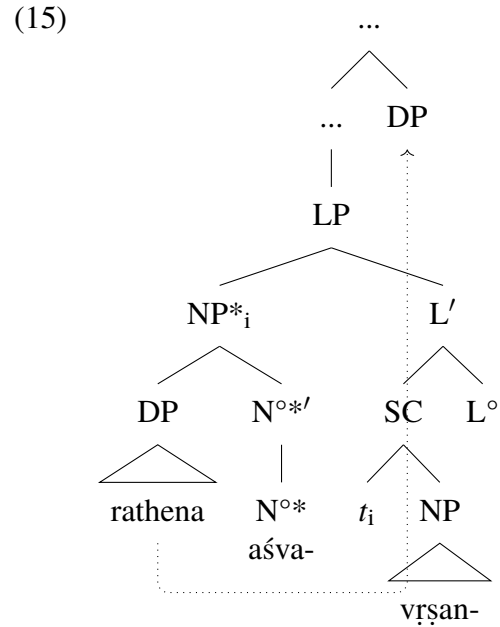
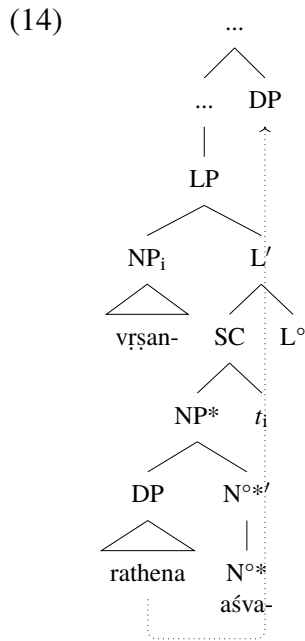
with its subject.<sup>17</sup>



The derivation proceeds along the following lines. First, one of the two components of the small clause moves to the specifier of a functional head, dubbed as  $L^\circ$  (i.e., linker) by Mocci (2022, 2025), in order not to violate a general syntactic principle which prevents the direct merger of two maximal projections (e.g., the NP headed by *tvád-* and the NP\* headed by *pitṛ-* in (13)).<sup>18</sup> This yields the following phrase markers, where the small clause component raised to Spec-LP is either the predicate ((14), (17)) or the subject ((15), (16)).

<sup>17</sup>Following standard practice, the predicate is allocated to the right-hand slot of the small clause in these phrase markers. It should be noted that the bahuvrīhi's external referent is represented as an inflected word (*ráthena*, GEN.SG; *devāḥ*, NOM.PL) and hence as a DP in these phrase markers, whereas the bahuvrīhi members are represented as stems (e.g., *áśva-*; *pitṛ-*) and hence as NPs, under the assumption that inflection belongs in the functional lexicon. In this perspective, the case marker appended to the right-hand member of bahuvrīhis (*[[vṛṣan]-[aśvéna]]*, with *aśvéna* 'horse.INS.SG.M'; *[[tvát]-[pitārah]]*, with *pitārah* 'father.NOM.PL.M') is an optical illusion: the case-marker is indeed appended to the entire bahuvrīhi (*[[vṛṣan]-[aśvá-] + INS.SG.M]*; *[[tvát]-[pitṛ-] + NOM.PL.M]*), which is also to be treated as a stem, albeit a complex one, before it gets case-marked (Candotti & Pontillo 2022). On the empirical and theoretical motivation for this differential morphological and syntactic representation of the bahuvrīhi members vs. the bahuvrīhi's external referent, see Mocci (2025, 201-204).

<sup>18</sup>Different versions of this general principle, which was originally discovered by Moro (2000) under the rubric of 'Dynamic Antisymmetry', have been advanced in the literature. For two recent versions, dubbed as 'Labeling Algorithm' and 'Generalized Dynamic Antisymmetry', see Chomsky (2013) and Moro & Roberts (2024), respectively. For empirical and theoretical arguments in favor of the postulation of  $L^\circ$ , see instead Mocci (2025, 194-209).



Second, the bahuvr̥hi's external referent (i.e., *ráthena* and *devāḥ*) moves from Spec-NP\* - i.e., the position where it was interpreted as the possessor of the subject, as in (14)-(15), or of the predicate, as in (16)-(17) - to an LP-external position, i.e., to a compound-external position, under the assumption that LP is the upper phrasal boundary of bahuvr̥his. This is indicated by the dotted arrows in (14)-(17).<sup>19</sup>

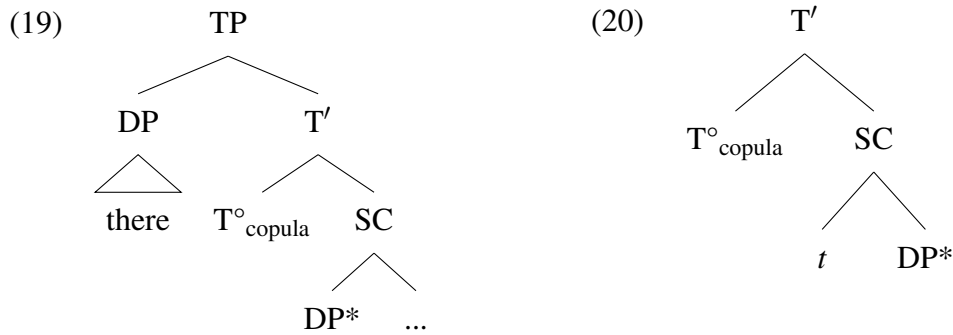
<sup>19</sup>It should be pointed out that, while the ordering constraint exemplified in (11) determines which bahuvr̥hi member (i.e., the qualifier or the qualificand) is allocated to the right-hand slot of the bahuvr̥hi, it does not say anything about the positioning of the bahuvr̥hi with respect to its external referent (e.g., the positioning of

Third, the LP-internal (hence, the bahuvrīhi-internal) material maps into a linear order in keeping with the following general scheme (see Kayne 1994 for a detailed discussion of this scheme): that the specifier of a head (e.g.,  $L^\circ$ ) systematically follows the complement of that head.<sup>20</sup> Accordingly, after the bahuvrīhi's external referent (*rāthena*) has reached an LP-external position, (14)-(15) respectively yield the linear orders *vṛṣaṇ-aśvá-* and *\*aśva-vṛṣan-*, the latter being unattested. Likewise, after the raising of the bahuvrīhi's external referent (*devāḥ*) to an LP-external position, (16)-(17) respectively yield the linear orders *tvát-pitr-* and *\*pitṛ-tvad-*, the latter being - again - unattested.

Last, the unattested orders (*\*aśva-vṛṣan-*, corresponding to (15), and *\*pitṛ-tvad-*, corresponding to (17)) are treated as violations of an independently assumed locality principle. To illustrate this point, let us consider the following pair of English copular sentences. While (18a) exemplifies the extraction of linguistic material from an in-situ (i.e., small-clause-internal) subject, (18b) exemplifies the extraction of linguistic material from an in-situ predicate nominal.

- (18) a. [Which wall]<sub>i</sub> do you think there were [<sub>DP\*</sub> two pictures of  $t_i$  ]?  
 b. [Which riot]<sub>i</sub> do you think two pictures of the wall were [<sub>DP\*</sub> the cause of  $t_i$  ]?

Now, the relevant portions of (18a-b) are given the syntactic representations in (19)-(20), respectively, in Moro's (1997) classic work on predication. Crucially, these phrase markers show: (i) that the extraction from an in-situ subject is licensed by a local relation between a predicative head (the  $D^\circ$  *there*) and the head ( $T^\circ$ ) which locally c-commands that subject ( $DP^*$ ), as depicted in (19); that the extraction from an in-situ predicate is licensed by a head ( $T^\circ$ ) locally c-commanding that predicate, as depicted in (20).



*vṛṣaṇ-aśvéna* with respect to *rāthena* in (11a)). In fact, a bahuvrīhi may either precede or follow its external referent. Here I shall not be concerned with the positioning of the bahuvrīhi's external referent; I will limit myself to representing its landing site to the right of the bahuvrīhi itself, as indicated in (14)-(17).

<sup>20</sup>This, of course, does not mean that any specifier of any head  $X^\circ$  precedes any complement of  $X^\circ$  in Sanskrit: syntactic movement, ultimately related to the expression of scope-discourse semantic values, may disrupt the basic Specifier-Complement order. Nonetheless, this disturbing factor may be disregarded for our purposes, inasmuch as there is no empty slot within LP that may license further instances of movement than those already represented in (14)-(17).

Regardless of the exact formulation of the locality conditions that capture the extraction from the subject and predicate nominal,<sup>21</sup> the following considerations can be made: (i) *vṛṣan-aśvá-* is well-formed because it is the spell-out of a phrase marker (i.e., (14)) which corresponds in all relevant respects to (19), i.e., the phrase marker for the well-formed extraction from an in-situ subject; (ii) *tvát-pitṛ-* is well-formed because it corresponds in all relevant respects to (20), i.e., the phrase marker for the ill-formed extraction from an in-situ predicate. *\*aśva-vṛṣan-* and *\*pitṛ-tvad-* are instead ill-formed because they are the spell-out of phrase markers where the extraction takes place from a non-in-situ subject (15) or a non-in-situ predicate nominal (17), which is in violation of locality.<sup>22</sup>

All in all, the derivation of Sanskrit bahuvrīhis involves the movement of the bahuvrīhi's external referent from the phrase projected by a bahuvrīhi member, which fulfils the function of either subject or predicate. In the next section I shall extend this derivation to Vedic Sanskrit *tough*-constructions, which are codified as bahuvrīhis in this language.

#### 4 The derivation of Vedic Sanskrit *tough*-constructions

Let us go back to the Vedic Sanskrit *tough*-constructions in (4)-(5), repeated below as (21)-(22). Since *su-gātave* and *su-hánā* can be categorized as bahuvrīhis for the reasons specified in the preceding section, it is reasonable to treat (21)-(22) along exactly parallel lines to (11a-b). Accordingly, I take *ubhé* and *vṛtrá* as the external referents of *su-gātave* and *su-hánā*, respectively.

(21) *ubhé* ... [su]-[gātave]  
 both.NOM.DU.M [easy]-[reach.INF.DAT]  
 'Both these things are easy to reach.' (*Atharvaveda-Śaunakīya* 6.1.3c)

(22) *kṛdhi* [su]-[hánā] ... *vṛtrá*  
 make.IMP.AOR.2SG [easy]-[smash.INF.ACC.PL.N] enemy.ACC.PL.N  
 'Make [our] enemies easy to smash!' (*Rgveda* 7.25.5c)

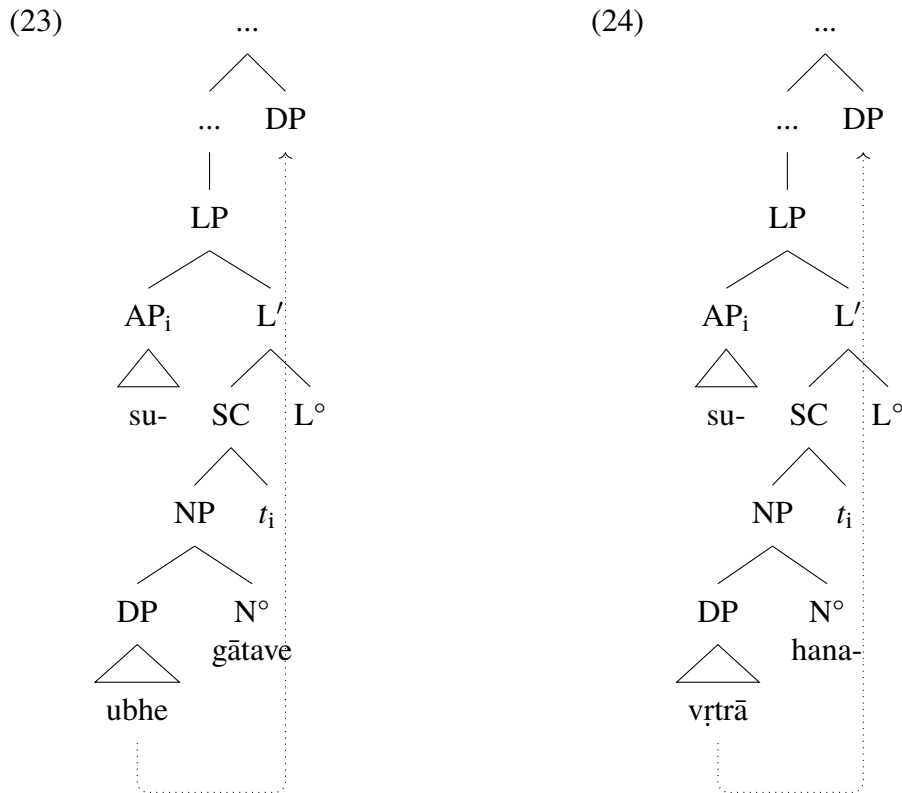
The parallelism with (11a-b) also requires that a subject-predicate link is assumed between the bahuvrīhi members, whereby *su-* 'easy' is the predicate (in both examples) and the right-hand members *gātave* 'reach.INF.DAT' and *hánā* 'smash.INF.ACC.PL.N' the corresponding subjects. This is reminiscent of Salzmann (2023, 233-237), who argues that *tough*-adjectives such as *easy* take an infinitival complement such as *to reach* as their external argument in such English *tough*-constructions as *these things are easy to reach*: considering that the relationship of adjectives (qualifiers) with their external arguments (qualificands) may be reduced to a subject-predicate link at the relevant level of representation (den Dikken

<sup>21</sup>See Moro (1997, 2017a,b) for a formulation based solely on selection and c-command, along the lines of the Subjacency Condition à la Cinque (1990). See instead den Dikken (2006) for a phase-based formulation. Cf. also Mocchi (2025, 38-96) for a critical comparison of Moro's and den Dikken's formulations.

<sup>22</sup>Space constraints prevent me from discussing why locality is violated in these cases. See however Mocchi (2025) for a detailed discussion of this point.

2006, 10-12), it is not unreasonable to assume that, besides being an external argument of *easy*, *to reach* also serves as its subject at the relevant level of representation.

By pursuing the parallelism with (11a-b) further, the following phrase markers are obtained for (21)-(22), where the external referents are represented as internal arguments (bearing the theme theta-role) of the infinitives *gātave* and *hánā*, and hence as part of the phrase projected by one of the bahuvrīhi member, in line with (16)-(19).



That is, the bahuvrīhi-internal predicate *su-* moves to Spec-LP in both (23)-(24) in order not to violate the general ban against the direct merger of two maximal projections (i.e., the SC-internal occurrences of NP and AP). Next, the bahuvrīhi's external referent (*ubhé* 'both'; *vṛtrā* 'enemies') moves outside the bahuvrīhi, to an LP-external position, as indicated by the arrow, in both (23) and (24). Such movement does not violate locality because it is an instance of extraction from an in-situ subject, where  $L^\circ$  (i.e., the head locally c-commanding this subject) is in a local relation with the predicate of the in-situ subject itself (i.e., with the  $A^\circ$  *su-*). In other words, both (23)-(24) run parallel to (19), i.e., to the well-formed extraction from the in-situ subject of *there*-sentences (e.g., *which wall do you think there were two pictures of?*). In this way, the derivation of the bahuvrīhis *su-gātave* and *su-hána* completes successfully, yielding the *tough*-constructions in (21)-(22).

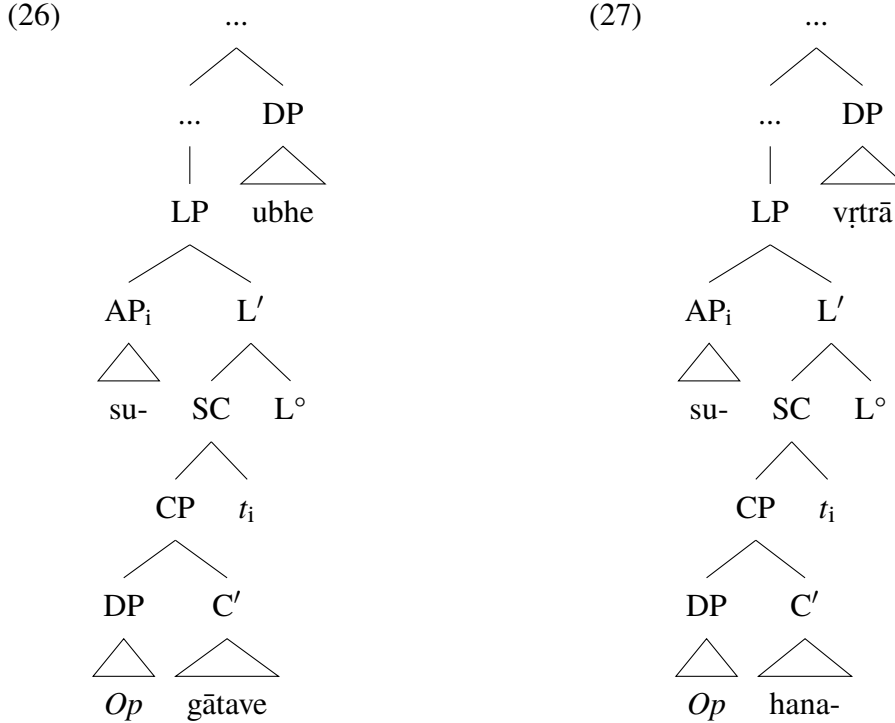
In the next section I shall assess the consequences that the derivation of Vedic Sanskrit *tough*-constructions championed here has for the nature of *tough*-movement.

## 5 Consequences for the nature of *tough*-movement

There are at least two points that should be made with regard to the syntactic representations in (23)-(24). First, the infinitive (*su-gātave*; *su-hānā*) is represented as a noun ( $N^\circ$ ) there. This is because of a well-established ban on the right-hand member of Sanskrit compounds being a full-fledged verb, i.e., an element endowed with tense-aspect morphology (see Lowe 2015a, 269-273 on this ban). Related to this categorization of the infinitive in (23)-(24) as a noun is the second point: namely that no silent operator is housed in the left periphery of the infinitive in these phrase markers, unlike in the representation of English *tough*-constructions (2), repeated below as (25).

(25) [These things]<sub>i</sub> are [<sub>AP</sub> easy [<sub>CP</sub> *Op*<sub>i</sub> [<sub>PRO</sub> to reach  $t_i$  ]]]

That is, while *tough*-movement in English is handled as the movement of a silent operator to the left-periphery of the embedded clause projected by the infinitive (see above, §1), the postulation of a null operator that moves to the left-periphery (Spec-CP) of the infinitives *gātave* and *hānā* would simply be an unnecessary complication. Let us consider why. The movement of the null operator to Spec-CP preempts the movement of the bahuvrīhi's external referent (*ubhé*; *vṛtrā*) from within LP to an LP-external position: the bahuvrīhi's external referent is simply base-generated in a bahuvrīhi-external position (just as *these things* is base-generated in the matrix clause in (25)), along the lines of (26)-(27).



Now, crucially, the null operator does not cross the CP boundary projected by the infinitive in (26)-(27). This implies that no movement of linguistic material outside the phrase pro-

jected by a bahuvrīhi member (i.e., by the infinitives *gātave* and *hānā*) takes place in this case. However, the ordering constraint on Sanskrit bahuvrīhis (according to which if the bahuvrīhi's external referent is interpreted as a dependent of the bahuvrīhi member  $\alpha$ ,  $\alpha$  is allocated to the right-hand slot of the bahuvrīhi) was reduced to a mere consequence of the locality conditions on the extraction of the bahuvrīhi's external referent from within the phrase projected by a bahuvrīhi member to an LP-external position (see above, §3).

Therefore, the movement of the null operator to the left-periphery of the infinitive forces us to state the linear order of the bahuvrīhis *su-gātave* and *su-hānā* as an axiom - not being derivable from independently assumed principles of grammar - thereby setting *su-gātave* and *su-hānā* aside from all other Sanskrit bahuvrīhis. Rather than accepting this treatment of *su-gātave* and *su-hānā* as an exceptional type of Sanskrit bahuvrīhi, I shall stick to the representations in (23)-(24), which replace the movement of a silent operator to the left-periphery of the infinitive (*gātave* and *hānā*) with the movement of the bahuvrīhi's external referent (*ubhé; vṛtrā*) to an LP-external (i.e., bahuvrīhi-external) position. The main motivation for adopting the representations in (23)-(24) is that they make it possible to treat the linear order of virtually all Sanskrit bahuvrīhis - including Vedic Sanskrit *su-gātave* and *su-hānā* - not as an axiom but as a theorem, being derivable from locality (i.e., from an independently assumed principle of grammar).

All in all, the codification of *tough*-constructions via compounding in Vedic Sanskrit suggests treating *tough*-movement not as the movement of a silent operator to the left-periphery of an infinitive, but rather as the movement of an overt phrase (which I have dubbed as 'the bahuvrīhi's external referent') from within the bahuvrīhi-internal phrase projected by the infinitive to a position outside the bahuvrīhi; no left-periphery needs to be assumed for the bahuvrīhi-internal infinitive, which can be simply categorized as a noun. Since the bahuvrīhi-external position targeted by the movement of the bahuvrīhi's external referent is a position where case is assigned (or checked, depending on what theoretical approach to case marking is taken),<sup>23</sup> *tough*-movement in fact qualifies as a type of A-movement in Vedic Sanskrit. Accordingly, Vedic Sanskrit contrasts with English, where *tough*-movement has instead been argued to be a variety of A-bar movement, whereby a silent operator targets the left-periphery of the embedded infinitive of English *tough*-constructions. The present study is thus in line with Tayalati & Danckaert (2020), who brought data from Modern Standard Arabic in support of the conclusion that *tough*-constructions may not actually form a natural class crosslinguistically.

## 6 Concluding remarks

*Tough*-constructions are codified via bahuvrīhi compounding in Vedic Sanskrit. Since the derivation of Sanskrit bahuvrīhis has been independently established to involve the A-movement of the bahuvrīhi's external referent to a bahuvrīhi-external position, in the present contribution this derivation has been extended to the bahuvrīhis that codify *tough*-constructions

<sup>23</sup>See Chomsky (2000, 123-124) and Chomsky (2001, 16-17) for an approach to case-marking based on the notion of checking. See instead Chomsky (1993) for a government-based approach to case-marking.

in Vedic Sanskrit. In this way, *tough*-movement has been identified with the A-movement of the bahuvrīhi's external referent. The main takeaway from this study is therefore that *tough*-movement is A-movement in Vedic Sanskrit, which accordingly contrasts with English, where *tough*-movement has been identified with the A-bar movement of a silent operator to the left-periphery of the embedded infinitive.

## References

- Adger, David. 2013. *A syntax of substance*. Cambridge, MA: MIT Press.
- Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.
- Bošković, Željko. 2015. From the Complex NP Constraint to everything: On deep extractions across categories. *The Linguistic Review* 32(4). 603–669.
- Candotti, Maria Piera & Tiziana Pontillo. 2019. Lexical subordination in derivation and compounding. *Studi e Saggi Linguistici* LVIII(2). 155–175.
- Candotti, Maria Piera & 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.
- Chomsky, Noam. 1977. On wh-movement. In Peter W. Culicover, Thomas Wasow & Adrian Akmajian (eds.), *Formal syntax*, 71–133. New York: Academic Press.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1993. *Lectures on Government and Binding: The Pisa lectures*. Berlin: de Gruyter Mouton 7th edn.
- Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels, Juan Uriagereka & Samuel Jay Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of howard lasnik*, 89–155. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. *Beyond explanatory adequacy* (MIT Occasional Papers in Linguistics 20). Cambridge, MA: Department of Linguistics and Philosophy, MIT.
- Chomsky, Noam. 2013. Problems of projection. *Lingua* 130. 33–49.
- Cinque, Guglielmo. 1990. *Types of  $\bar{a}$ -dependencies*. Cambridge, MA: MIT Press.
- den Dikken, Marcel. 2006. *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Cambridge, MA: MIT Press.
- Elizarenkova, Tatyana J. 1995. *Language and style of the Vedic Ṛsis*. Albany, NY: State University of New York Press.
- Fushimi, Makoto. 2012. Taittirīya-Saṃhitā. Electronic text input by Makoto Fushimi; TITUS version by Jost Gippert, last edited 21.04.2012. <http://titus.uni-frankfurt.de/texte/etcs/ind/aind/ved/yvs/ts/ts.htm>.
- Hale, Mark. 2018. The syntax of indo-iranian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics, 1924–1942*. Berlin: de Gruyter Mouton.
- Hartman, Jeremy. 2012. *Varieties of clausal complementation*: Massachusetts Institute of Technology dissertation.

- Hicks, Glyn. 2009. *Tough*-constructions and their derivation. *Linguistic Inquiry* 40(4). 535–566.
- Hicks, Glyn. 2017. *Tough*-movement. In Martin Everaert & Henk van Riemsdijk (eds.), *The wiley blackwell companion to syntax, second edition*, 1–27. John Wiley & Sons, Ltd. doi:<https://doi.org/10.1002/9781118358733.wbsyncom035>.
- Hock, Hans Henrich. 1996. Who’s on First? toward a Prosodic Account of P2 Clitics. In Aaron L. Halpern & Arnold M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, 199–270. Stanford, CA: CSLI Publications.
- Jamison, Stephanie W. & Joel P. Brereton. 2014. *The Rigveda: The earliest religious poetry of India. an English translation*. Oxford/New York: Oxford University Press.
- Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Keine, Stefan & Ethan Poole. 2017. Intervention in *tough*-constructions revisited. *The Linguistic Review* 34(2). 295–329.
- Kim, Jeong-Soo. 2022. Atharvavedasaṃhitā der śaunakaśākhā: Eine neue Edition unter besonderer Berücksichtigung der Parallelstellen der Paippalādasamhitā. <https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/27703>.
- Lowe, John J. 2014. Accented clitics in the *ṛgveda*. *Transactions of the Philological Society* 112(1). 5–43.
- Lowe, John J. 2015a. *Participles in Rigvedic Sanskrit: The syntax and semantics of adjectival verb forms*. Oxford: Oxford University Press.
- Lowe, John J. 2015b. the syntax of Sanskrit compounds. *Language* 91(3). e71–e115.
- Manzini, Maria Rita. 2017. Passive, smuggling and the *by*-phrase. In Nicholas LaCara, Keir Moulton & Anne-Michelle Tessier (eds.), *A Schrift to Fest Kyle Johnson*, 233–244. Linguistics Open Access Publications.
- Mocci, Davide. 2022. The possessive relation in Sanskrit bahuvrīhi compounds: Ellipsis or movement? *Journal of South Asian Linguistics* 12(2). 1–30.
- Mocci, Davide. 2025. *A configurational approach to Sanskrit bahuvrihi compounds* (RGG Companion Series). Lingbuzz Press. <https://lingbuzz.com/j/rgg/monographs/mocci/>.
- 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. 2017a. Copular sentences. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax. Second Edition*, 1210–1232. Oxford: John Wiley & Sons Ltd.
- Moro, Andrea. 2017b. Existential sentences. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax. Second Edition*, 1483–1508. Oxford: John Wiley & Sons Ltd.
- Moro, Andrea. 2019. The geometry of predication: A configurational derivation of the defining property of clause structure. *Philosophical Transactions of the Royal Society B* 375(1791). 1–7.

- Moro, Andrea & Ian Roberts. 2024. The duality of syntax: Unstable structures, labelling and linearisation. *Natural Language and Linguistic Theory* 42. 609–631.
- van Nooten, Barend A. & Gary B. Holland. 1994. *Rig Veda: A Metrically Restored Text*. Cambridge, Mass.: Harvard University Press.
- Ntelitheos, Dimitrios. 2022. Compounding in greek as phrasal syntax. *Languages* 7(151). 1–22.
- 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.
- Radicchi, Anna. 1985–1988. *La teoria pāṇiniana dei samāsa secondo l’interpretazione delle scuole grammaticali indiane dal quinto all’ottavo secolo d.C.* Firenze: Elite. Two volumes.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar*, 281–337. Dordrecht: Kluwer.
- Salzmann, Martin. 2023. Experiencer intervention in English *tough* movement: Evidence from extraction of the *tough* adjective against syntactic- and semantic-intervention accounts. *Syntax* 26(2). 223–249.
- Tayalati, Fayssal & Lieven Danckaert. 2020. The syntax and semantics of Modern Standard Arabic resumptive *tough*-constructions. *Folia Linguistica* 54(1). 197–238.
- Weber, Albrecht. 1871–1872. *Die Taittirīya-Saṃhitā*. Leipzig: Brockhaus. Two volumes.
- Whitney, William Dwight. 1896. *A Sanskrit grammar: Including both the Classical language, and the older dialects, of Veda and Brahmana*. Leipzig: Breitkopf & Härtel. Third edition.
- Whitney, William Dwight. 1962. *Atharva-Veda Saṃhitā: Translated with a critical and exegetical commentary*. New Delhi: Motilal Banarsidass. Volume 2 (Books VIII to XIX).