

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

Davide Mocci (University of Cagliari; University of Genoa)

1. Background. *Tough*-movement, exemplified by the dependency between *John* and *t* in (1), has A'-movement properties in English: it can in principle cross infinitely many clause boundaries; it can cross the subject of any clause embedded within the complement of the *tough*-predicate; it is blocked by an intervening element in Spec-CP (Chomsky 1977:103-104): see the contrast between (1a) and (1b), slightly adapted from Hicks (2009:542).

(1a) [John]_i is hard [to imagine [**any woman** believing [**she** would agree to marry *t*_i]]]

(1b) ??[John]_i is hard [to imagine [any woman wondering [**why** she would agree to marry *t*_i]]]

However, *tough*-movement also has A-movement properties: it is theta-marked by the infinitive embedded within the *tough*-predicate, rather than by the *tough*-predicate itself; it is blocked by any experiencer PP intervening between the subject of the matrix clause and the first infinitive embedded within the complement of the *tough*-adjective: see (2) from Hartman (2012:87-90).

(2a) It is important (**to Mary**) to avoid cholesterol.

(2b) Cholesterol is important (***to Mary**) to avoid *t*_i

(2c) **To Mary**, cholesterol is important to avoid *t*_i

To account for this twofold nature of *tough*-movement, Hicks (2009:546-552) came up with a smuggling derivation, where a complex null operator A'-moves from the complement of *avoid* to the embedded CP, while *cholesterol* A-moves from within DP to the matrix subject position.

(3) Cholesterol_j is important [_{CP} [_{DP} **D°** [_{NP} **Op** *t*_j]]_i [PRO to avoid *t*_i]]

Indeed, Keine & Poole (2017:321-324) re-interpreted the data in (2) as the result of a semantic type mismatch, which excises the need for the smuggling derivation in (3). Just as Chomsky (1977), they reduce *tough*-movement to the A'-movement of a silent operator to the CP of the infinitive: the subject of the matrix clause externally merges with the AP headed by the *tough*-predicate, and hence does not undergo A-movement, as illustrated in (4) (Keine & Poole 2017:315, 321). The co-indexing of *Op* with *cholesterol* accounts for the illusion of *cholesterol* being theta-marked by *avoid*.

(4) Cholesterol_i is [_{AP} important [_{CP} **Op**_i [PRO to avoid *t*_i]]]]

2. Vedic and beyond. *Tough*-movement does not exhibit A'-movement properties in Romance, where the crossing of any embedded subject and clause boundary (CP) yields ill-formedness (Rizzi 1982:25-27): cf. English (1) with Italian (5), taken from Kayne (2000:52).

(5) *[Questo libro]_i è difficile [da convincere [_{CP} **Mario** a finire *t*_i prima di lunedì]].

'This book is difficult to convince Mario to finish before Monday.'

The old Indo-Aryan language Vedic Sanskrit (Vedic for short), dating back to 1500-500 B.C., provides morphological data that are in line with the Italian evidence in (5), suggesting that *tough*-movement lacks A'-movement properties in both languages. Thus, *tough*-constructions are formed via exocentric (so-called "bahuvrīhi") compounds in Vedic (Whitney 1899:§1287a): e.g., *su-gātave* and *su-hánā* (6). As a rule, the right-hand slot of bahuvrīhis cannot be a full-fledged verb (Lowe 2015:269-273). The element corresponding to the English infinitive (*gātave* 'going'; *hánā* 'killing') is allocated to the right-hand slot of the bahuvrīhi in Vedic *tough*-constructions (6a-b). Hence, such an element is non-verbal.

(6a) ubhé ... [su]-[gā-tav-e]
both.NOM easy-go-NOMINALIZER-DAT
'Both these things are easy to reach.' (*Atharvaveda-Śaunakīya* 6.1.3)

(6b) kṛdhi [su]-[hán-ā] ... vṛtrá
make.IMP.2SG easy-smash-NOMINALIZER.ACC enemies.ACC
'Make [our] enemies easy to smash!' (*Rgveda* 7.25.5)

The non-verbal nature of *gātave* and *hánā* in (6) is confirmed by the fact that they involve suffixes – here dubbed as 'nominalizers' – that are independently known to take as input verbal bases (*gam-* 'to go'; *han-* 'to kill') and yield as output nominal bases: *-tav-* in (6a) and *-a-* (6b) (see Keydana 2013).

Therefore, the occurrences of *gātave* and *hánā* in (6) are on the nominal side of the noun-verb continuum. Now, although the nominal status of an element does not per se exclude the projection of a CP by that element in languages like Hungarian (Szabolcsi 1987), a compound-internal nominal does indeed fail to project a CP in Vedic. This is witnessed by the rigidity of Vedic compounds' word-order and their impossibility of housing sentence-level particles like *evá* and *u*. All in all, *gātave* and *hánā* lack the CP layer when serving as bahuvrīhi-members, as in (6).

3. A puzzle. If no CP is involved in the bahuvrīhis *su-gātave* and *su-hánā* of (6), A'-movement – which presupposes the presence of a CP – must also be banned from these bahuvrīhis. A direct consequence of this is that an analysis à la Keine & Poole (2017) that reduces *tough*-movement to the A'-movement of a silent operator targeting the CP of the infinitive is empirically inadequate not only for Italian (5), but for Vedic, too. We are now faced with a puzzle: how can *ubhé* 'both' and *vṛtrā* 'enemies' be theta-marked by *gātave* 'going' (6a) and *hánā* 'killing' (6b), respectively, granted the unavailability of an operator in Vedic *tough*-constructions?

4. On bahuvrīhis. A closer look at the structure of bahuvrīhis provides the relevant insight into this puzzle. Thus, Mocci (2022), (2024) argued that the derivation of Vedic bahuvrīhis such as *subhāgaḥ* (7) involves the A-movement of a non-case-marked element (*sá*) from the NP headed by the compound's right-hand member (*bhāga-*) to a compound-external position where such an element (*sá*) is case-marked: see (7b), where the compound boundary is represented by a case phrase (KP – Bittner & Hale 1996), with K° spelled out as *-ḥ* (nominative).

(7a) *su-bhāga-ḥ sá [...] astu*
 good-share-NOM he.NOM be.IMP.3SG
 'Let him be with good share!'

(7b) $[_{KP} [_{LP} \text{su-} [_{L'} [_{NP} \text{bhāga- } t_i] L^\circ=\emptyset]] K^\circ=\text{ḥ}] \text{sa}_i$ (*Rgveda* 1.86.7)

The compound's left-hand member (*su-*) is housed in the specifier of the compound head L°, a linker (Eik 2019:182-183) which is silent here but is overtly realized elsewhere (see Mocci 2022 on how this configuration licenses the extraction of *sá* from NP). L° and K° are linearized to the right of their complements in (7) in the light of Vedic's head-finality, while the moved element (*sá*) may freely appear to the left or the right of the compound.

5. Proposal. I contend that *tough*-movement in Vedic belongs to the family of A-movements exemplified in (7). Thus, *ubhé* 'both' of (8), just like Italian *entrambi* 'both' of (9), A-moves from the internal argument position of the infinitive (*gātave/raggiungere* 'to reach'), where it gets theta-marked, to the subject of the matrix clause, thereby crossing the silent external argument (PRO) of the infinitive. Relativized Minimality is not violated because PRO is indeed to be understood as a free variable – existentially closed at the C-I interface – which lacks the relevant features to qualify as an A-intervener, along the lines of Manzini (2017:241).

(8) $ubhé_i [_{KP} [_{LP} \text{su-} [_{L'} [_{NP} \text{PRO } gātav- t_i] L^\circ=\emptyset]] K^\circ=e] = ubhé \text{ su-}gātave ((6a))$

(9) $[Entrambi]_i \text{SONO } [_{LP} \text{facili } [_{L'} L^\circ=da [_{NP} \text{PRO } raggiungere t_i]]] = \text{'both are easy to reach'}$

6. Conclusions. *Tough*-movement is A-movement in Vedic (as well as Italian), and A'-movement in English. This opposition directly reflects the fact that a CP fails to be stacked above the infinitive in Vedic *tough*-constructions like (6): in the absence of a CP, the silent operator à la Keine & Poole (2017) – and hence the A'-movement properties associated with that operator, too – is excluded. All in all, this study bears fresh data from an ancient South-Asian language to bear on the crosslinguistic variation of *tough*-movement. Such variation may be correlated with the nominal (Vedic) vs. verbal (English) nature of the infinitive in *tough*-constructions.

Bittner & Hale 1996 *Linguistic Inquiry* 27(1). **Chomsky** 1977 In Culicover, Wasow, and Akmajian, eds. **Eik** 2019 *The Morphosyntax of Compounding in Norwegian*. **Hartman** 2012 *Varieties of clausal complementation*. **Hicks** 2009 *Linguistic Inquiry* 40(4). **Kayne** 2000 *Parameters and Universals*. **Keine & Poole** 2017 *The Linguistic Review* 34(2). **Keydana** 2013 *Infinitive im Rgveda*. **Lowe** 2015 *Participles in Rigvedic Sanskrit*. **Manzini** 2017 In LaCara, Moulton, and Tessier, eds. **Mocci** 2022 *JSAL* 12(2). **Mocci** 2024 *Proceedings of the Fifty-ninth Annual Meeting of the Chicago Linguistic Society*. **Rizzi** 1982 *Issues in Italian Syntax*. **Szabolcsi** 1987 In I. Kenesei, ed. **Whitney** 1889 *A Sanskrit Grammar*.