

Tracing the history of the comitative in Indo-Aryan

Aryaman Arora

December 17, 2022

Contents

1	Introduction	2
1.1	Case to no case and back again	2
1.2	What is a comitative?	5
2	Drawing isoglosses	6
2.1	OIA	6
2.2	MIA	7
2.3	NIA	8
2.3.1	Sanskrit <i>samám</i> × <i>sahá</i>	8
2.3.2	Sanskrit <i>sárt^ha</i>	10
2.3.3	Sanskrit <i>saṅgá</i>	11
2.3.4	Indo-Aryan <i>*nāla</i>	12
2.3.5	Other	12
3	Putting it all together	13

1 Introduction

The Indo-Aryan (IA) language family is the branch of the Indo-European language family that is spoken in the northern half of the Indian subcontinent. By consensus estimates, the modern Indo-Aryan family contains the largest number of languages out of all the Indo-European branches, and yet the historical development of Indo-Aryan is relatively understudied. Indo-Aryan has undergone widespread restructuring of its morphosyntax, from non-configurational syntax and synthetic morphology to more rigid word order and analytic inflection. For example, the large case system of Old Indo-Aryan (OIA; c. 1700 BCE–600 BCE) was gradually degraded in Middle Indo-Aryan (MIA; 600 BCE–1000 CE) and largely replaced by case markers and postpositions in New Indo-Aryan (NIA; 1000 CE–present). This process has been studied in previous work, including Butt and Ahmed [2011], Reinöhl [2016], Kulikov [2006]. A unique feature of the NIA case system is that it largely retains the original semantic distinctions of the OIA system despite its radical restructuring.

In this paper, I will zoom into the historical development of the strategy to indicate the comitative case, i.e. the semantic notion of accompaniment. The semantics and polysemous usage of the comitative postposition in NIA has already been studied to an extent by Khan [2009]. I will examine the diachronics of comitative marker development in Indo-Aryan, and analyse patterns in the grammaticalisation of such forms. This necessitates an overview of the relevant overarching morphosyntactic changes from OIA to NIA, as well as defining what the comitative is, both of which I will now turn to.

1.1 Case to no case and back again

Vedic Sanskrit had numerous tools at its disposal to mark spatio-temporal relations and semantic roles. For the latter, it had an extensive system of 8 cases, indicated morphologically on nominals. These were semantically bleached; the exact semantics are licensed by the argument structure that is lexically encoded in the governor.

Case	Form
Nominative	<i>deváh</i>
Vocative	<i>déva</i>
Accusative	<i>devám</i>
Instrumental	<i>devéna</i>
Dative	<i>deváya</i>
Ablative	<i>devát</i>
Genitive	<i>devásya</i>
Locative	<i>devé</i>

Table 1: Declension of *devá* ‘god’ in Vedic Sanskrit in the singular.

For more concrete spatio-temporal relations, Vedic Sanskrit used preverb particles (listed in table 2), which function similarly to adverbs. In Classical Sanskrit, most of these are lost as independent words and only retained as derivational prefixes on verbs, or syntactically pattern as adverbs and are not confined to preverbal position. Preverb particles are indeclinable, and thus belong to a third morphological class of word besides verbs and nominals.

Concurrent with the degradation of this case system and morphological inflection in general, the overarching post-Sanskrit syntactic change across Indo-Aryan was the **rise of configurationality** [Reinöhl, 2016]. Vedic Sanskrit was a truly non-configurational language; besides the argument



Figure 1: Indo-Aryan languages.

Preverb	Meaning
<i>ati</i>	beyond, over
<i>ad^{hi}</i>	above, besides
<i>anu</i>	after, along, alongside
<i>antar</i>	interior, within
<i>apa</i>	down, off, back
<i>api</i>	unto, close upon or on
<i>ab^{hi}</i>	to, towards, into, over, upon
<i>ava</i>	off, away, down, down from
<i>ā</i>	near, near to, towards- change of direction
<i>ud</i>	up, upwards, upon, on, over, above
<i>upa</i>	towards, near to, by the side of, with
<i>ni</i>	down, in, into
<i>nis</i>	out, forth
<i>parā</i>	away, forth
<i>pra</i>	forward, onward, forth, fore
<i>prati</i>	back to, in reversed direction
<i>pari</i>	round about, around
<i>vi</i>	apart, asunder, away, out
<i>sam</i>	along, with, together

Table 2: Vedic Sanskrit preverbs [Papke, 2010].

ordering flexibility (SOV preferred but all attested), all clauses show extreme levels of permissible discontinuity, to the point that there is no canonical NP-internal structure. For example, genitive-case NPs do not have to be continuous with their governors, even in prose texts without metrical constraints:

- (1) **manó** ha vaí dēvá **manuṣyasya** ājananti
mind.ACC.SG PRT PRT god.NOM.PL man.GEN.SG know.3PL
‘The gods know the **mind of man.**’ (Sanskrit, *Śatapathā Brāhmaṇa* 1.1.1.7)

A confluence of factors led to the rise of configurationality in NIA, which have developed internally-structured NPs and PPs while permitting argument ordering flexibility at the VP level. NIA is far more rigid than Sanskrit, dispreferring discontinuity besides limited sentence-fronting topicalisation.

(2) Hindi:

- a. **us** **kā** **c^hōṭ-ā** **b^hāī** āy-ā t^h-ā
3SG.OBL GEN little-M.SG brother.NOM come-PFV.M.SG be-PST.M.SG
‘His little brother had come.’
b. ***us** **kā** **b^hāī** **c^hōṭā**
c. ***b^hāī** **c^hōṭā** **us** **kā**

A concurrent development in modern Indo-Aryan is the rise of postpositions, which head the newly structured PPs [Butt and Ahmed, 2011]. These serve as a new layer of the lexicon, replacing Sanskrit’s morphologically-fused cases (e.g. genitive *-asya*), preposition–adverbs (*pāri* ‘around’), and some spatial nouns (*mād^{hya}* ‘middle; inside’). As a result, NIA has three layers of lexical material that convey relational semantics [Masica, 1993]:

- **Layer I:** Morphological case, greatly degraded from the large inventory of Sanskrit (to a minimum of 2 cases, NOM and OBL).
- **Layer II:** Case markers/clitics, which indicate theta roles and other core arguments, generally with predicate-licensed semantics.
- **Layer III:** Complex postpositions, which govern a case clitic or morphological case and are often transparently derived from nouns or other POS categories.

The boundary between Layer III and PPs governed by other categories can be fuzzy (not unlike English *outside of*, *inside of*, or German dative-licensing adpositions). For example, the Hindi Layer III postposition *ke andar* ‘inside of’ can be analysed as a complex PP or a noun governing a genitive PP:

- (3) a. vo [pānī ke andar]_{PP} hai
 3SG water GEN.OBL.M inside COP.3SG
 ‘He is underwater.’ (Hindi)
- b. vo [[pānī ke]_{PP} andar]_{NP} se nikl-ā
 3SG water GEN.OBL.M inside ABL exit-PFV.MSG
 ‘He came out from under the water.’ (Hindi)

Also, like nouns the content part of the complex PP also is marked for gender, reflected on the case marker it licenses (e.g. *andar* in *ke andar* ‘inside’ is masculine, whereas *taraf* in *kī taraf* ‘towards’ is feminine). However, distributionally the complex PPs behave like Layer II case-marked PPs: both can occupy verb-adjunct positions and the complement slot of motion verbs. It should be apparent now that postpositions are a complicated category in IA, at the unstable edge between nominals and case markers.

1.2 What is a comitative?

The **comitative** relation is usually described in English as ‘together with’ [Haspelmath, 2009]; a comitative case marker indicates the secondary co-participant of an event. Arkhipov [2009] more formally defines the comitative as ‘pluralising’ participants, noting its semantic equivalence to coordination and plurals:

- (4) a. Tom came **with** Anna. (comitative)
 b. Tom **and** Anna came. (NP-coordination)
 c. **They** came. (plural)
 d. Tom came, **and** Ana came. (clausal coordination)

A comitative participant is defined to be non-core and thus non-obligatory no matter the predicate. Arkhipov is careful to draw a distinction with **polyadic** participants, which can take comitative marking on core, obligatory arguments in many languages. Compare the following:

- (5) a. Tom met **with** Anna. (polyadic)
 b. → *Tom met.
- (6) a. Tom came **with** Anna. (comitative)
 b. → Tom came.

Thus, whether the comitative-marked participant is obligatory is a test for true comitatives that is effective cross-lingually, unlike the informal semantic definitions many previous works relied on. Arkhipov acknowledges that it is not perfect, since some predicates really are semantically ambiguous (e.g. English *fight* can take both comitative or polyadic *with*-participants).

2 Drawing isoglosses

Having clearly defined a true comitative as marking a non-obligatory and secondary co-participant in an event, we can move towards examining the diachronic and synchronic variety of forms marking such an argument in Indo-Aryan languages. For reference, in Hindi, the comitative is indicated with the complex genitive-licensing postposition *ke sāt^h*:

- (7) *rām kabīr ke sāt^h ga-yā*
 Ram Kabir GEN COM go-PFV.M.SG
 ‘Ram went with Kabir.’ (Hindi)

The overall story of comitative marking over the history of Indo-Aryan turns out to be very complicated, as summarised in the diagram below (Modern NIA innovations without any older history are not shown). The complex situation depicted in the diagram asks us to consider the semantic range and history of each of these forms—for example, how did multiple comitative forms coexist in the same stage?

OIA	MIA			NIA			
Vedic	Classical	Early	Middle	Late	Early	Medieval	Modern
<i>sa, sām</i>							
<i>sākām</i>							
	[INS]						
<i>samām</i>		<i>samam</i>		<i>samaū</i>	<i>savem</i>	<i>saum, syūm</i>	<i>sā, se</i>
<i>sahā</i>		<i>saha</i>		<i>sahu(m), saū, saī</i>	<i>saha</i>	× ↑	
<i>sārd^ham</i>		<i>sadd^him</i>					
	<i>sahita</i>			<i>sahiya</i>		<i>saheta</i>	<i>sīh</i>
				<i>sarisa</i>	<i>sariseṁ</i>		
				<i>satt^hihi</i>		<i>sāt^he, -iim</i>	<i>sāt^h(e)</i>
				<i>saṁgerṁ</i>		<i>saṁge, -ā</i>	<i>saṁg</i>
						<i>nāli</i>	<i>nāl</i>

2.1 OIA

Briefly, let us examine how Sanskrit expresses the comitative using data from the Digital Corpus of Sanskrit [Hellwig, 2010–2021]. Vedic Sanskrit most commonly uses the instrumental by itself to mark accompaniers; animate nouns likely bear too much agency to be interpreted as actual instruments. This strategy is attested in the earliest Vedic compositions.

- (8) *dev-ó dev-éb^hir á gam-at*
 god-NOM.SG god-INS.PL toward go-AOR.SBJV.3SG
 ‘...the God, come hither with the Gods.’ (Vedic Sanskrit; RV 1.1.5)

Vedic Sanskrit also attests the use of the preverbs *sa* and *sām*, both meaning ‘together’ or ‘with’, as well as the verbal root *sac-* ‘to accompany, follow’.

- (9) *sām uṣád-b^hiḥ a-jā-yathāḥ*
 with dawn-INS.PL IPFV-be.born-IND.MED.2SG
 ‘...wast born together with the Dawns’ (Vedic Sanskrit; RV 1.6.3)
- (10) *sác-asvā naḥ svast-áye*
 be.with-PRS.IMP.MED.2SG 1PL.ACC/DAT/INS health.DAT.SG
 ‘...be with us for our weal’ (Vedic Sanskrit; RV 1.1.9)

Late Vedic Sanskrit and Classical Sanskrit usually expresses the comitative using the adposition *sahá*, which is derived from the Vedic preverb *sá* + *d^hā* ‘to place’, and thus builds on the older comitative strategy.

- (11) ahaṁ mama kuṭumb-ena saha gataṁ
 1SG 1SG.GEN family-INS.SG COM go.PST.PTCP
 ‘I went with my family.’ (Sanskrit)

Lexical equivalents for *sahá* proliferate beginning with late Vedic Sanskrit, when the preverbs begin to fossilise into verb prefixes and a vacuum opens for comitative markers to be enlisted into. The Rigveda attests *sākám* < *sá* + *añc* ‘to move’ for temporal simultaneity, which secondarily serves as a comitative marker. Also found in the Rigveda is *samám* ‘in like manner, similar’ which is an adverbial derivation from the preverb *sám* and like *sākám* develops a comitative sense in late Vedic. The Śatapatha-Brāhmaṇa (c. 7th century BCE) has the first use of *sārd^ham* ‘together’ < *sá* + *ārd^ha* ‘half’, which is an interestingly literal indicator of accompaniment given it is an adverbial derivation from *sārd^ha* ‘plus a half’, i.e. plus an additional participant. Classical Sanskrit frequently uses *sahita* < *sahá* + the *-ita* adjectivaliser. All of these *sá*-derived adpositions license the instrumental case on their object.¹

The varying distributions and semantic differences between these strategies is worth investigating further, but it suffices for this paper to note the use of the instrumental case for accompaniment along with alternative adpositions which all have forms beginning in *sa-*.

2.2 MIA

Pali and the Prakrits. In Pali (starting in the 3rd century BCE), the Sanskrit strategies of the plain instrumental and the instrumental with *saha*, *sahita*, or *sadd^hiṃ* (< Skt. *sārd^ham*) continue to be used to indicate accompaniment [Wijesekera, 1936]. The Dramatic Prakrits (3rd century CE onwards) have basically the same situation, excepting the sound change of *sahita* > *sahi(y)a* [Sheth, 1923–1928]. There is no pressure to innovate new markers given the relative preservation of the case system and the postpositions from late Vedic and Classical Sanskrit.

Apab^hraṁśa. The late Middle Indo-Aryan languages (c. 6th–13th centuries CE) are collectively known as Apab^hraṁśa to Indian grammarians. Bubenik [1998]’s grammar describes no less than five known strategies to express accompaniment, which I attempt to etymologise below:

INS + <i>samaü</i> , <i>samāṇu</i>	< Skt. <i>samám</i>
(INS) + <i>sahu(m)</i> , <i>saiü</i> , <i>sai(m)</i>	< Skt. <i>sahá</i>
<i>satt^hihi</i>	< Skt. <i>sārd^ha</i> ‘shared goal → caravan’
GEN + <i>saṁgem</i>	< Skt. <i>saṅgá</i> ‘contact’
INS + <i>sahiya</i>	< Skt. <i>sahita</i>

The first two strategies dominate all historical stages of Apab^hraṁśa. The variant *samāṇu* probably reflects Sanskrit *samānám* ‘like’ which is often synonymous with *samám*. The third and fourth strategies are innovative grammaticalisations of nouns into new postpositions which increase in use in late Apab^hraṁśa. The first four are also polysemous; besides marking the comitative, they also mark polyadic arguments and instruments. The last one is probably an anachronistic usage due to its rarity, reflecting a formal register.

¹This grammatical rule is covered in the Aṣṭādhyāyī, sūtra 2.3.19: *sahayukte ‘prad^hāne*, i.e. ‘*sahá* (and its synonyms) license the third case’.

The instrumental as a separate morphologically-indicated case is no longer used to mark accompaniment at this point. We also encounter the first instances of true case clitics in this stage of Indo-Aryan; late Western Apabhraṃśa (centered around Gujarat), represented by the *Sarṇdeśa-Rāsaka* (c. 13th century), attests *saii* and *satt^{hi}hi* as non-case licensing clitics which attach directly to the noun, like the Layer I case markers of NIA which they are a predecessor to.

- (12) *viviha-viakk^haṇa satt^{hi}hi jai pavas-ii ṇiru*
 various-clever COM if enter-3SG continuously
 ‘If in the company of clever persons one takes a stroll in the city’ (Sarṇdeśa-Rāsaka 43)

Since *saii* is attested continuously from Vedic *sahá*, this means it was a Layer III postposition before being grammaticalised into a Layer II case clitic. On the other hand, *satt^{hi}hi* is not attested previous but found to be like a Layer II clitic, in that it does not license a morphological case or clitic. The form *satt^{hi}hi* is actually an oblique derived from the noun *sārt^ha* ‘shared goal → caravan, company’ < *sa* ‘same, together’ + *art^ha* ‘goal, aim’. It is attested in this Sanskrit form in the Campaka-śreṣṭhī-kathānaka, a late Sanskrit prose text with known manuscripts from c. 1500 [Turner, 1962–1966, Hertel, 1911]. This suggests two separate pathways to developing a Layer II case marker in NIA; either it can be grammaticalised from a historical Layer III postposition, or it can be integrated directly from a noun.

2.3 NIA

I found equivalents in a representative sample languages from all subfamilies of Indo-Aryan; below I discuss the historical development of these.

2.3.1 Sanskrit *samám* × *sahá*

Until Late MIA, the reflexes of the comitative postpositions *samám* and *sahá* were kept separate as *samaü* and *sahu*, each with a plethora of variants. Beginning in Early NIA, the situation gets messier. Old Marathi (up to 14th c. CE) attests *savem* and *saha* as separate comitative markers, but the latter could be a learned reborrowing from Sanskrit [Tulpule and Feldhaus, 1999]. The Old Bengali Caryā-pada (10th–12th c. CE) attests *sama* as well as *sane*, *sayem* which are claimed by Chatterji [1926, p. 774] to be both derived from *samám*. Meanwhile, in Medieval Hindi (15th c.) one strategy which indicates the comitative is *saum* and its variants *so*, *som*, and *syūm* which can be argued to derive from either Sanskrit *samám* or *sahá* [Strnad, 2013, p. 355].

In general, Early and Medieval NIA show conflation of *samám* × *sahá* due to phonological degradation. Nevertheless, the resulting form continues to serve as a comitative postposition. However, its semantic breadth expands significantly. In Old Bengali and Medieval Hindi it is already primarily an instrumental and polyadic argument marker. Modern Hindi has the instrumental/ablative case clitic *se* as the direct descendant of this form, and it has entirely lost the comitative sense. In essence, we observe this chain of grammaticalisations:

1. OIA: COMITATIVE
2. Late MIA and Early NIA: COMITATIVE > INSTRUMENT
3. Modern NIA: COMITATIVE > INSTRUMENT > ABLATIVE

Nevertheless, some NIA languages preserve this form as a comitative marker, including Shina, Mewati, Sadri, and Sindhi. The geographically spotty continuation of its use points to these being archaic retentions rather than shared innovations.

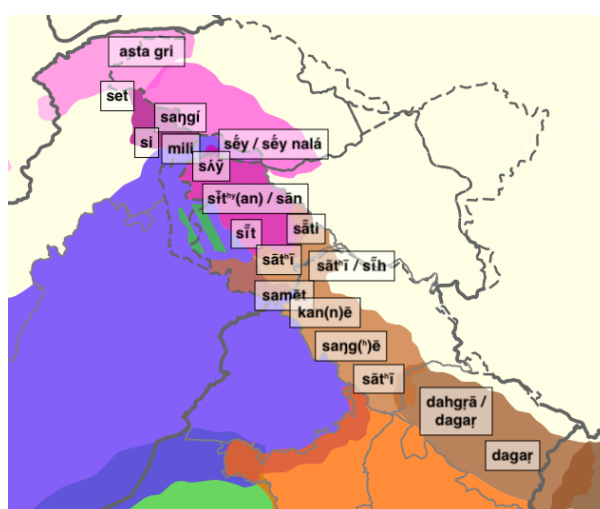
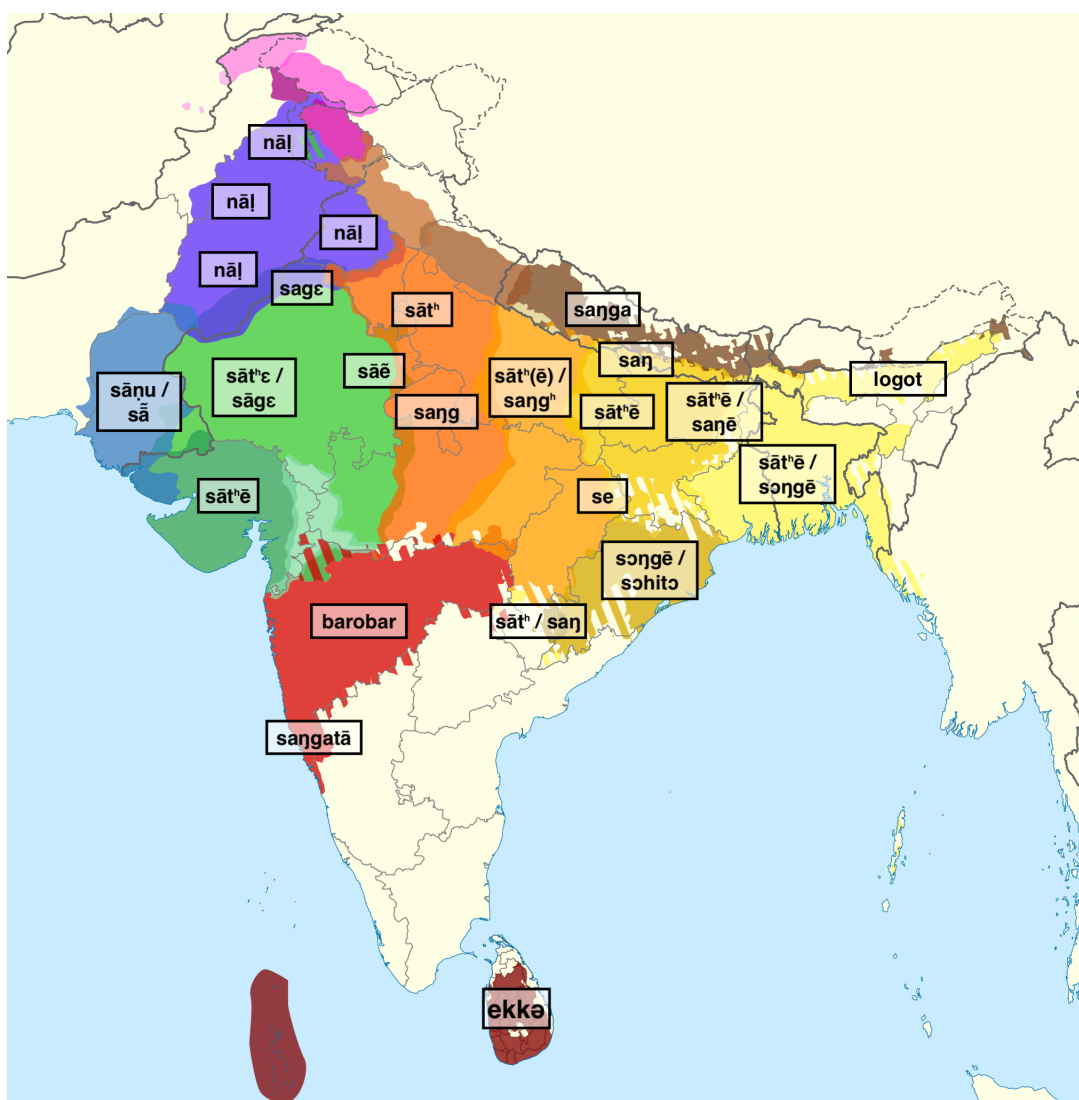


Figure 2: An overview of comitative markers in the modern Indo-Aryan languages.

- (13) mo śahra-žóō žava-sěy áalus
1SG city-ABL brother-COM come.PFV
'I came with my brother from the city.' (Gurezi Shina)
- (14) ke ke mo-r sãě jāvegpo
who who 1SG-GEN COM go.FUT
'Who all will go with me?' (Mewati)
- (15) mõe tor se jā-e rihi
1SG 2SG.GEN COM go-LNK PFV.PST.1PL
'I went along with you.' (Sadri)
- (16) hun-ani sã
3PL-OBL COM
'with them' (Sindhi)

2.3.2 Sanskrit *śārtḥa*

In the Gangetic plain spanning from Western Hindi to Bengali, most languages use a reflex of Sanskrit *śārtḥena* 'in the company of' as a monosemous comitative case marker.² However, as discussed above, its use as a postposition is not attested until late Middle Indo-Aryan, in a Western Apabhramśa text from the 13th century and a Sanskrit prose text from c. 1500. Therefore, it must be a relatively recent development.

In NIA, its first attestation is in the Old Gujarati Upadeśamālābālāvabodha (1487 CE) as *sātḥiim*. Gujarati is the only major NIA language that does not attest any other comitative markers in its history; apparently *sātḥiim* has had no competition and has persisted until modern Gujarati *sātḥe*, undergoing semantic broadening to a polyadic argument marker (but not an instrumental). Curiously, Medieval Hindi does not attest any reflex of *śārtḥa*, and neither does even Middle Bengali (16th c.). It is found in the Medieval Punjabi Guru-Granth-Sāhib (16th c.) as *sātḥi*, which does not conform to expected sound changes in Punjabi (where geminates are preserved, i.e. **sattḥi*) and thus is likely borrowed.

Among modern Indo-Aryan, it is continued as a comitative case marker (from east to west) in the Hindi belt, the Bihari languages, and Bengali. The Kashmiri comitative marker *sītḥy* may also fall in this isogloss, but the vowel does not fit expected sound changes. The late attestation of this form and its limited geographical range suggests that it was lexically diffused northwards from Gujarat. This is not surprising, given the similar history proposed for the ergative marker *ne* by Butt and Ahmed [2011] and others.

- (17) āmi mā-r sātḥe jābo
1SG mother-GEN COM go.FUT
'I will go with my mother.' (Bengali)
- (18) larkā māi ke sātḥe bajāra gail
boy mother GEN COM market go.PFV
'The boy went to market with his mother.' (Bhojpuri)
- (19) m^hē t^hā-rē sātḥē kām-sū jāūlā
1SG 2SG-GEN COM work-to go.FUT
'I will go to work with you.' (Marwari)

²As Gumperz [1957] notes, the Indus and Gangetic plains together form a dialect chain: "the local dialects form a continuous chain from Sind to Assam, the speech of each area shading off into that of the adjoining one." The portion from Bengal to Haryana (i.e. the core Gangetic plain) shares reflexes of *śārtḥena*.

- (20) bi gōs aslam-as sīt^{hy}/sīt^{yan} cakr-as
 1SG go.PFV Aslam-DAT COM walk-DAT
 'I went for a walk with Aslam.' (Kashmiri)

2.3.3 Sanskrit *saṅgá*

In various (discontinuous) regions across nearly the entire Indo-Aryan-speaking Indo-Gangetic plain, the comitative marker is indicated with reflexes of Sanskrit *saṅgá* 'battle' → 'contact with'. The semantic change in *saṅgá* from a concrete noun meaning 'battle' (attested in the oldest Sanskrit text, the Ṛg-veda) to a more abstract contact relation, is already found in the Vedic Sanskrit Taittirīya-Saṁhitā recension of the Yajur-veda dated to the early Indian Iron Age (c. 1200–800 BCE; Witzel, 2001).

- (21) vṛśced yad akṣa-saṅgam
 cut.OPT.3SG if axle-contact.ACC.SG.N
 'If he were to cut it so that it would touch the axle...' (Sanskrit, TS 6.3.3.4)

While modern standard varieties of Hindi, Punjabi, and Marathi do not use it, it is found in older varieties of all three (pre-16th century).

- (22) so tana jar-ai kāṭ^{ha} kai saṅgā
 3SG body burn-PRS.3SG wood GEN COM
 'This body will burn down together with the firewood.' (Old Hindi; Kabīr)
- (23) kah-u nānaka piru merai saṅge tā mai nava nid^{hi} pā-ī
 say-3SG Nanak husband 1SG.GEN COM then 1SG nine treasure obtain-3SG
 'Says Nanak, when my Husband Lord is with me, I obtain the nine treasures.' (Old Punjabi; GGS)

All NIA languages that use a derivative of Sanskrit *sārtha* also attest a derivative of *saṅgá* for the comitative. In some cases (e.g. Hindi), *saṅgá* is seen as archaic or poetic, while other languages (e.g. Bengali) use both *saṅgá* and *sārtha* regularly and interchangeably. Generally, *saṅgá* has not undergone as much grammaticalisation to the INSTRUMENT relation, but some languages (such as Nepali) do use it for polyadic argument marking.

- (24) tũ kĩ gẽ sage ga-yo ho
 2SG what GEN COM go-PFV.M.SG COP.PST
 'Who did you go with?' (Bagri)
- (25) āmi mā-r saṅge jābo
 1SG mother-GEN COM go.FUT
 'I will go with my mother.' (Bengali)
- (26) me-re saṅg kōn ā-ygo
 1SG-GEN COM who come-FUT
 'Who will come with me?' (Braj)
- (27) ek dzaṇ-e-re dev-ā saṅge edz-a si
 one Jana-OBL-GEN.PL god-OBL COM come-IPFV COP.PRS.PL
 '...the other comes from Jana with the deity.' (Kullui)
- (28) m^hẽ t^hã-re sāge kām-sũ jāũlā
 1SG 2SG-GEN COM work-to go.FUT
 'I will go to work with you.' (Marwari)

- (29) báab-a saṅgí śáhr-a t^he ba-áa-nu
 father-OBL COM city city-OBL to go-PRS-M.SG
 ‘I am going to the city with my father.’ (Palula)
- (30) tai hu mir saṅ b^hādz pat^hts^hu se
 2SG EMPH 1SG.GEN COM run behind ABL
 ‘You run with me behind me.’ (Rana Tharu)
- (31) hun-ani sāṇu vañu
 3PL-OBL COM go.IMP
 ‘Go with them.’ (Sindhi)

2.3.4 Indo-Aryan **nāla*

In the northwestern fringe of the Indo-Aryan family, some languages attest a comitative form phonologically derivable from **nāla*. It is unclear what attested Sanskrit form could correspond to this; Turner [1962–1966] suggests Sanskrit *aṅkapāli* ‘embrace’ but this is phonologically difficult due to the velar cluster. The oldest attestation of the form is in the holy texts of Sikhism, written in an early form of Punjabi in the 16th century CE.

- (32) jī^hai bahi samaj^hāi-ai tī^hai koi na cal-io nāli
 where sit.CNV explain-PRS.3SG there anyone NEG go-? COM
 ‘He sits and examines the accounts, there where no one goes along with anyone.’ (Old Punjabi; GGS)

Among modern Indo-Aryan languages it is found in Punjabi, Saraiki, Hindko, and Gurezi Shina; it is very likely that other languages in the upper Indus attest it as well.

- (33) mo śáhra-žóž zava-séy nalá áalus
 1SG city-ABL brother-COM COM come.PFV
 ‘I came with my brother from the city.’ (Gurezi Shina)
- (34) salīm sād-e nāl ā-ve-g-ā
 Salim 2PL.GEN-SG.M.OBL COM come-SBJV.3SG-FUT-SG.M
 ‘Salim will come with us.’ (Punjabi)

2.3.5 Other

Many languages attest idiosyncratic forms for the comitative, which undoubtedly are recent innovations. I will loosely group these by the grammaticalisation paths they fall under.

Assamese and Dameli have the development ‘contact’ > COM. For Assamese this is from the Sanskrit root *lag-* ‘to stick, touch, contact’ and for Dameli it is *mil-* ‘to meet, encounter, touch’.

- (35) mo-r tā-r logot nāsibor mon
 1-GEN 2-GEN COM dancing wish
 ‘I wish to dance with him.’ (Assamese)
- (36) sapun sootii āstrakaa mili nanavaat ge-n
 all together women COM mercy take-INF
 ‘they all [went], with their women, to seek mercy’ (Dameli)

Some of the Western Pahari languages, an extremely diverse branch of Indo-Aryan, use forms derived from Sanskrit *karṇe* ‘ear.LOC’ to indicate the comitative: Churahi *kanē*, Pangwali *kēṇī*, Dogri *kannē*,

Chambeal *kanē*, etc. In other NIA languages *karṇe* is usually the source of the accusative or ergative cases [Bailey, 1908].

Marathi and Odia both use loanwords with the development ‘together’ > COM. For Marathi, the loan is from Persian *bar-ā-bar* ‘equal’ and for Odia it is from Sanskrit *sahita* ‘together’ (which is historically attested as a comitative marker as well).

- (37) lili āpl-yā mul-ā **barobar** ā-l-ī
 Lili self-GEN.M.SG son-OBL COM come-PFV-F.SG
 ‘Lili came with her son.’ (Marathi)

- (38) pue-ti tā mā **sahita** bājār gōlā
 boy-CLF DEF? mother COM market go.PFV
 ‘The boy went to market with his mother.’ (Odia)

The remaining forms I found are of uncertain origin. Sinhala *ekka* may be from Sanskrit *éka* ‘one’ but the semantics are hard to explain. Garhwali *dahgrā* is suggested by Montaut [2022] to be derived from a grammaticalisation ‘path’ > ‘companion’ > COM, with the last step similar to Skt. *sārt^hena*.

- (39) buhḍrī gō kī lohṛiũ **dahgrā** dāl pihsṇũ jāli
 old.woman village GEN.F girl.PL COM daal grind.for go.FUT
 ‘The old woman will go with the girls of the village to grind daal.’ (Garhwali)

- (40) te **asta gri** par-au
 3SG along with go-PFV.3SG
 ‘He left with them.’ (Kalasha)

- (41) apē puta-t **ekkə** ennə
 1PL.GEN son-? COM come
 ‘Come with our son.’ (Sinhala)

3 Putting it all together

As I established earlier, the rise of configurationality in Middle Indo-Aryan completely reworked the syntax of Indo-Aryan languages, and the concurrent erosion of case morphology led to the development of new layers of case clitics (Layer II; marking core arguments) and postpositions (Layer III; more spatio-temporal and adjunct uses).

Having seen a large amount of cross-lingual data, we can move on to the actual questions that began this paper: how did these comitative markers develop, and why do they straddle the boundary between Layer II and Layer III?

To summarise the descriptive parts of this paper, I showed that there while there are three major isoglosses (*sārt^hena*, *sangá*, **nāla*), there are many more regional innovations (and even loans) with no apparent relation in form to each other. I also showed that *sangá* has a much wider reach in early New Indo-Aryan languages (e.g. Old Hindi). In terms of the semantics of these developments, we can look to the known cross-linguistic grammaticalisation pathways for the comitative marker [Heine et al., 2002].

Putting it all together, this is the story I construct for the history of comitatives in Indo-Aryan.

First, configurationality gradually arose in Middle Indo-Aryan in conjunction with the erosion of the morphological case system and increasing grammaticalisation of new postpositions (e.g. *madhye* ‘inside’). The instrumental case was more resilient than some of the others so Pali and other Middle Indo-Aryan language continued to keep it for the comitative. But, *saha* was also available to express the comitative.

Second, the instrumental case was entirely lost by late Middle Indo-Aryan. A profusion of forms were enlisted as comitative markers. All of these etymologically contain the prefix *sa-* or *saṁ-* ‘with’, and undoubtedly the phonological similarity among all of these aided their adoption. Semantically, common grammaticalisations here are COMRADE > COMITATIVE and ACCOMPANY > COMITATIVE.

Stage	Form	Sanskrit	Etymology
OIA	INS + <i>sahá</i>	< <i>sahá</i>	< <i>sa-</i> + <i>d^ha</i>
Early MIA	INS + <i>sadd^hiṃ</i>	< <i>sārd^hám</i> ‘jointly’	< <i>sa-</i> + <i>árd^ha</i> ‘half’
Late MIA	<i>samau</i>	< <i>saṁmuk^há</i> ‘facing’	< <i>saṁ-</i> ‘together’ + <i>muk^há</i> ‘face’
Late MIA	<i>satt^hihi</i>	< <i>sārt^hena</i>	< <i>sa-</i> + <i>ártha</i> ‘aim; wealth’
Late MIA	GEN + <i>saṁgem</i>	< <i>saṅgá</i>	< <i>sañj-</i> ‘cling’ × <i>saṁ-</i> + <i>ga</i> ‘come’
Late MIA	INS + <i>sahiya</i>	< <i>sahita</i>	< <i>sa-</i> + <i>d^hita</i> ‘placed’

Thirdly, some languages (mostly on the fringes of the Indo-Aryan-speaking area) innovated entirely different forms for the comitative that lack the *sa-* commonality of Middle Indo-Aryan. There is no clear reason for this; innovations could have occurred for many reasons (incl. social differentiation, substratum influence) which are not easily accessible now. One driver may be grammaticalisation of the comitative into other uses which necessitated a new, semantically-narrower form to replace the old one. Nevertheless, many of the innovations share the previous semantic development of COMRADE > COMITATIVE. Garhwali’s unique form has the cross-linguistically attested grammaticalisation FOLLOW > COMITATIVE.

Finally, some languages underwent the next step in the grammaticalisation chain of COMITATIVE > INSTRUMENT > ABLATIVE [Narrog and Ito, 2007], which moved this case form into the role of marking core arguments. This explains why the comitative is a Layer II case marker in languages like Punjabi, and doesn’t have to license the genitive. Compare Punjabi and Hindi’s way to express the instrumental:

- (42) *mẽ ne ām ko cākū se kā-ṭā*
 1SG ERG mango ACC knife INS cut-PFV.M.SG
 ‘I cut the mango with a knife.’ (Hindi)
- (43) *mẽ ne amb nū cākū nāl kaṭṭ-īā*
 1SG ERG mango ACC knife INS cut-PFV.M.SG
 ‘I cut the mango with a knife.’ (Punjabi)

Hindi uses a different case marker for the instrumental than that of the comitative. Punjabi has the same form for both the comitative and the instrumental. Both use Layer II markers for the instrumental, since it is a core argument; it just so happens that Punjabi uses this for comitatives too. That concludes the story of the comitative in Indo-Aryan languages—a story of language innovation trying to keep up the grammaticalisation treadmill and phonological degradation.

Sources

Language	Source
Angika	Lahiri [2021]
Assamese	Lahiri [2021]
Bagri	Lakhan [2000]
Bengali	Lahiri [2021]
Bhojpuri	Lahiri [2021]
Braj	Sharma [2011]
Dameli	Perder [2013]
Garhwali	Chandola [1966], Montaut [2022]
Gurezi Shina	Ahmed [2016]
Hindko	Bashir and Connors [2019]
Kalasha	Heegård Petersen [2015]
Kashmiri	Koul and Wali [2004]
Kullui	Renkovskaya et al. [2019–2021]
Maithili	Lahiri [2021]
Marathi	Dhongde and Wali [2009]
Marwari	Sharma [2011]
Mewati	Sharma [2011]
Odia	Lahiri [2021]
Old Hindi	Strnad [2013]
Palula	Liljegren [2019]
Punjabi	Bashir and Connors [2019]
Rana Tharu	Dhakal [2015]
Sadri	Peterson and Baraik [2021]
Saraiki	Bashir and Connors [2019]
Sindhi	Jetley [1964]
Sinhala	Fairbanks [1968]

References

- Musavir Ahmed. A descriptive study of the phonology of gurezi shina. *Acta Orientalia Academiae Scientiarum Hungaricae*, 69(1):87–106, 2016.
- Alexandre Arkhipov. Comitative as a cross-linguistically valid category. *New challenges in typology: Transcending the borders and refining the distinctions*, pages 223–246, 2009.
- Thomas Grahame Bailey. *Languages of the Northern Himalayas: Being Studies in the Grammar of Twenty-Six Himalayan Dialects*. Cambridge University Press, 1908. URL <https://archive.org/details/languagesnorthe00bailgoog>.
- Elena Bashir and Thomas J Connors. A descriptive grammar of hindko, panjabi, and saraiki. De Gruyter Mouton, 2019.
- Vit Bubenik. A historical syntax of late middle indo-aryan (apabhramsa). *A Historical Syntax of Late Middle Indo-Aryan (Apabhramsa)*, pages 1–291, 1998.
- Miriam Butt and Tafseer Ahmed. The redevelopment of Indo-Aryan case systems from a lexical semantic perspective. *Morphology*, 21(3):545–572, 2011.

- Anoop Chandra Chandola. *A syntactic sketch of Garhwali*. PhD thesis, The University of Chicago, 1966.
- Suniti Kumar Chatterji. *The Origin And Development Of The Bengali Language*. 1926.
- Dubi Nanda Dhakal. Notes on rana tharu grammar. 2015. URL https://www.academia.edu/43908543/Notes_on_Rana_Tharu_grammar.
- Ramesh Vaman Dhongde and Kashi Wali. *Marathi*. John Benjamins Publishing, 2009.
- Gordon H. Fairbanks. *Colloquial Sinhalese*. ERIC, 1968.
- John J. Gumperz. Language problems in the rural development of North India. *The Journal of Asian Studies*, 16(2):251–259, 1957.
- Martin Haspelmath. Terminology of case. *The Oxford handbook of case*, pages 505–517, 2009.
- Jan Heegård Petersen. Kalasha texts—with introductory grammar. *Acta Linguistica Hafniensia*, 47 (sup1):1–275, 2015.
- Bernd Heine, Tania Kuteva, et al. *World lexicon of grammaticalization*. Cambridge University Press, 2002.
- Oliver Hellwig. Dcs - the digital corpus of sanskrit, 2010–2021. URL <http://www.sanskrit-linguistics.org/dcs/index.php>.
- Johannes Hertel. The story of merchant campaka. *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, 65(1):1–51, 1911.
- Murlidhar Kishinchand Jetley. *Morphology of Sindhi: A descriptive analysis of Vicholi, the standard Sindhi dialect*. PhD thesis, Deccan College Post Graduate and Research Institute Pune, 1964.
- Tafseer Ahmed Khan. *Spatial Expressions and Case in South Asian Languages*. PhD thesis, University of Konstanz, 2009. URL <http://kops.uni-konstanz.de/handle/123456789/12508>.
- Omkar Nath Koul and Kashi Wali. *Modern Kashmiri Grammar*. Dunwoody Press, 2004.
- Leonid Kulikov. Case systems in a diachronic perspective. In *Case, valency and transitivity*, pages 23–47. 2006.
- Bornini Lahiri. *The Case System of Eastern Indo-Aryan Languages: A Typological Overview*. Taylor & Francis, 2021.
- Gusain Lakhan. Bagri. *München: Lincom Europa. Languages of the World/Materials*, 385, 2000.
- Henrik Liljegren. Palula dictionary. *Dictionaria*, 2019.
- Colin P. Masica. *The Indo-Aryan languages*. Cambridge Language Surveys. Cambridge University Press, Cambridge, 1993. ISBN 9780521299442.
- Annie Montaut. On the non-lexical categories of *avyay* ‘invariables’ and their grammaticalization in Pahari languages, with a comparison to Standard Hindi. 2022.
- Heiko Narrog and Shinya Ito. Re-constructing semantic maps: The comitative-instrumental area. *STUF-Language Typology and Universals*, 60(4):273–292, 2007.

- Julia Kay Porter Papke. *Classical Sanskrit preverb ordering: a diachronic study*. PhD thesis, The Ohio State University, 2010.
- Emil Perder. *A grammatical description of Dameli*. PhD thesis, Department of Linguistics, Stockholm University, 2013.
- John Peterson and Sunil Baraik. *Chotanagpuri Sadri: An Indo-Aryan lingua franca of eastern central India*. Central Institute of Indian Languages, Mysore, 2021. In press.
- Uta Reinöhl. *Grammaticalization and the Rise of Configurationality in Indo-Aryan*. Oxford University Press, 2016.
- E. Renkovskaya, J. Mazurova, and A. Krylova. The corpus of Kullui, 2019–2021. URL <https://kullui.polycorpora.org/>.
- J. C. Sharma, editor. *Linguistic Survey of India: Rajasthan (Part I)*. Language Division, Office of the Registrar General Census Commissioner, India, 2011.
- Hargovind Das T. Sheth. *Paia-sadda-mahannavo; a comprehensive Prakrit Hindi dictionary, with Sanskrit equivalents, quotations and complete references*. Calcutta, 1923–1928. URL <https://dsal.uchicago.edu/dictionaries/sheth/>.
- Jaroslav Strnad. *Morphology and syntax of Old Hindī: edition and analysis of one hundred Kabīr vānī poems from Rājasthān*. Brill, 2013.
- Shankar Gopal Tulpule and Anne Feldhaus. *A dictionary of old Marathi*. Popular Prakashan, Mumbai, 1999. URL <https://dsal.uchicago.edu/dictionaries/tulpule/>.
- Ralph Lilley Turner. *A comparative dictionary of the Indo-Aryan languages*. Oxford University Press, 1962–1966. URL <https://dsal.uchicago.edu/dictionaries/soas/>.
- OH de A Wijesekera. *Syntax of the Cases in the Pali Nikayas*. University of London, School of Oriental and African Studies (United Kingdom), 1936.
- Michael Witzel. Autochthonous Aryans? the evidence from old Indian and Iranian texts. 7(3):1–115, 2001.