e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

Towards a Typology of Negation in South Asian Languages

John Peterson
Christian-Albrechts-Universität zu Kiel, Deutschland

Lennart Chevallier
Christian-Albrechts-Universität zu Kiel, Deutschland

Abstract The present study is a preliminary typology of negation in South Asian languages, based on a database of 25 structural features for 39 languages from three language families and two language isolates. The documented features include the form of the negative marker, the relation of the negative construction to the corresponding affirmative form, whether there are different negative constructions used in different TAM categories, and whether these constructions are symmetric or asymmetric. This study also provides a first analysis of the distribution of these different negative-marking strategies throughout the subcontinent and suggests that a combination of both family bias and areal pressure are needed to account for many of the observed distributions. In some cases of language contact, the data also allows us to determine with some certainty the type of contact situation which has led to the negative-marking patterns documented in the database.

Keywords South Asia. Typology. Negation. Historical linguistics. Language contact. Areal linguistics.

Edizioni Ca'Foscari

Peer review
Submitted 2021-12-10
Accepted 2022-03-02
Published 2022-04-29

Citation Peterson, J.; Chevallier, L. (2022). "Towards a Typology of Negation in South Asian Languages". *Bhasha*, 1(1), 17-62.

DOI 10.30687/bhasha/8409-3769/2022/01/005

1 Introduction

In his study of negation in South Asian languages, Bhatia (1995, 13) provides a few examples of negative marking in Hindi and five other South Asian languages. For example, consider his Hindi examples, given here in (1) (gloss and order of sentences altered).

(1a) *vo* *nahĩ̃* *jā-egā.*
3SG NEG.IND GO-FUT.3SG.M
'He won't go'.

(1b) *tū* *mat* *jā.*
2SG NEG.NH.IMP GO.NH.IMP
'Do not go'.

(1c) *kyā* *vo* *na* *jā-e?*
Q 3SG NEG.SUBJ GO-SUBJ.3SG
'May he not go?'

What all three examples in (1) have in common is that they are all negated by a preverbal particle whose form depends on the mood of the clause: indicative negation is indicated by the particle *nahĩ̃* (1a), which Bhatia (1995, 16) derives from a fusion of the negative marker *na* and the copula *āhī*.[1] The non-honorific imperative is negated by *mat* (1b) while the subjunctive is negated by the negative particle *na* (1c). This is similar in many respects to what we find in Sanskrit (Old Indo-Aryan), where the preverbal particle *mā* is used to negate the imperative, with the likewise preverbal particle *na* found elsewhere (e.g. Whitney 1889, 413, § 1122c). However, this type of negation, with a non-inflecting negative particle preceding a verb and two or three modally determined distinctive forms, is by no means the only negating strategy in South Asian languages, as we will show in the following pages.

The primary goal of the present study is to document as much of the impressive array of negative marking in the languages of the South Asian mainland as possible, based on our current database.[2]

1 Although from a purely formal viewpoint it could also derive from *na* 'NEG' and *=hī* '=FOC', with nasalisation later spreading into the second syllable as this form lexicalized, yielding *nahĩ̃*. For a similar development already in Sanskrit, see Whitney 1889, 413, § 1122e.

2 This work represents a continuation of our ongoing areal-typological research of the languages of South Asia, originally sponsored by the German Research Council (DFG). The earlier project, whose database has been extended here to include a negation, was "Towards a linguistic prehistory of eastern central South Asia (and beyond)",

For practical reasons, in the present work we exclude languages of the Trans-Himalayan (Tibeto-Burman) and Tai-Kadai groups as well as languages spoken outside of the mainland (e.g. Dhivehi, Sinhala, Nicobarese etc.), although these will eventually be added to our database. The study is therefore still very much a work in progress and as the database increases and takes further languages and features into account, the picture will undoubtedly change somewhat. However, as we show below, this study already provides a detailed overview of negation strategies and their relation to the corresponding affirmative categories in ca. 10% of the languages of sub-Himalayan mainland South Asia, so that we believe that many of the distributional tendencies outlined in the following pages are of substantial relevance.

We restrict ourselves in this study to formal features of negation such as the form of the negative marking itself, the relation of the negative construction to the corresponding affirmative form, and also which TAM categories the respective negative constructions are found in. What we will not deal with here, however, are the semantic and pragmatic aspects of negation, which e.g. Bhatia (1995) deals with in his study. As negation is such a complex topic, it is not feasible to begin by investigating all aspects of it at the same time, at least not if the goal is to conduct a more-or-less representative survey. Thus, while Bhatia (1995) deals in considerable detail with marking patterns but also with semantic and pragmatic aspects of negation, his study is restricted to six languages – five Indo-Aryan and one Dravidian. In contrast, we deal here only with formal aspects of negation but in 39 languages from four major stocks and two isolates, allowing us to give a much broader picture of the various negative strategies found in these languages, albeit at the expense of pragmatic and semantic aspects.

The second goal of the present study is to use this information on marking strategies, to the extent possible, to help us identify past areas of language contact and the different types of contact situations which likely underlie these patterns. Innovations in the field of language typology since the early 1990s now allow us to use areal-typological methods to delve much deeper into linguistic prehistory than was previously possible (e.g. Nichols 1992; 1997), and more recent works in fields such as sociolinguistic typology (e.g. Trudgill 2011) and others often allow us to determine what type of contact likely prevailed in earlier times, e.g. prolonged societal bilingualism, language learning by large numbers of adult learners etc.

This study is structured as follows: § 2 presents a brief discussion of language contact in South Asia, which is often referred to as a *Sprachbund* or 'linguistic area', somewhat incorrectly in our view. Instead,

DFG project 326697274. We gratefully acknowledge our indebtedness to the German Research Council for funding this research.

we take a more differentiated view of language contact here and argue that the type of contact phenomena which is generally thought to constitute a linguistic area is in fact only one possible outcome of language contact, one which however is not supported by the data in South Asia. This is followed in § 3 by a brief discussion of our sample in § 3.1 and a detailed discussion of the features in the database in § 3.2, which largely follows the distinctions made in the crosslinguistic typological study of negation in Miestamo (2005), although we deviate occasionally from the methods in that study, as our goals here differ somewhat from Miestamo's. Then, in § 3.3, we briefly address 'zero negation', found in Dravidian.

In § 4 the results of our study are discussed, concentrating primarily on the various language clusters in the data. The significance of this data is assessed in § 5, where we discuss which clusters are likely the result of language contact and what type of contact may be responsible for the patterns we observe. Finally, § 6 provides a summary of the present study and mentions a number of topics for future research.

2 Language Contact. South Asia as a 'Linguistic Area'?

Typological similarities among South Asian languages belonging to different stocks were noted at least as early as Bloch (1934, 322-8), although the real momentum in research on language convergence in South Asia began with Emeneau (1956), who brought the spread of a number of features throughout much of the subcontinent to the attention of a larger linguistic audience. In the years that followed, numerous further features were suggested by various authors, many of which are summarised in Masica's (1976) landmark work on South Asia as a linguistic area. Masica's study expands the scope of research on South Asia as a "linguistic area" to include all of Eurasia and much of Africa, in order to determine to what extent South Asia differs linguistically from neighbouring regions. This is important since assuming that South Asia is a linguistic area in any meaningful sense of the term implies that it exhibits linguistic traits which distinguish it from its neighbours, something that the data however does not support.

Ebert (2006) comes to a very similar conclusion and also calls attention to a typological division of South Asia into two different zones, an eastern and a western, with the line of divide at about the 84th meridian, cutting Bihar, Jharkhand, Odisha and northeastern India off from the western subcontinent. More recent work on language contact in South Asia confirms this major typological schism, although not necessarily based on the same features as Ebert uses. For example, Peterson (2017); Ivani, Paudyal, Peterson (2021) and Borin et al. (2021) all call attention to structural differences distinguishing eastern and western Indo-Aryan from one another, which Peterson (2017)

refers to as the "Indo-Aryan east-west divide", with e.g. split-ergativity found in most western languages, whereas it is largely lacking in eastern IA languages; similarly, arbitrary gender is typically found in western languages, while eastern languages usually lack it etc. Such a divide, cutting right through the subcontinent and creating two typologically distinct regions, of course contradicts the very notion of a homogeneous linguistic area in the subcontinent.

If in fact anything like a South Asian linguistic area really does exist it would seem that it best fits what Campbell (2017, 27) refers to as a "trait-sprawl area" or "TSA". In this type of contact area, some features are found

> crisscrossing some languages while others crisscross other languages, with some extending in one direction, others in another direction, with some partially overlapping others in part of their distribution but also not coinciding in other parts of their geographical distribution.

This is in stark contrast to the "linguistic area *sensu stricto*" or "LASS", in which features are shared across the languages of a clearly delimited geographical area (Campbell 2017, 28).

Many researchers of language contact in South Asia appear to be looking for a list of features with which they can define a "LASS"-type area, in which ideally all South Asian languages share all of these traits. However, the facts clearly support a more "TSA"-like language area, in which certain features are found in many languages but the individual features do not all show the same geographical distribution. There may be "LASS"-type areas in South Asia, but if so these are likely to be found at the micro-level, which has been the focus of studies on language contact in South Asia in recent years (e.g. Abbi 1997; Ebert 1993; 1999; Osada 1991; Peterson 2010; 2015; Saxena 2015). We will therefore not look for signs of a larger 'South Asian linguistic area' here but will instead point out what appear to be contact-induced phenomena where these are suggested by the data.

In addition to identifying likely contact-induced areal patterns, we also hope to determine the societal conditions which led to the patterns we observe in the data. For example, recent works in sociolinguistic typology (e.g. Trudgill 2011) show that certain linguistic structures are more likely to emerge from one type of contact situation than from another. Simplifying somewhat, the argumentation in Trudgill (2011) which is relevant for our analysis can be summarised under the two following types:

- when a large percentage of speakers of a particular language are adult learners, this often leads to phonological and morphological simplifications in that language;

- in contrast, long-term societal bilingualism, especially in cases where speakers learn their second language during childhood, often leads to complexity.

"Simplification" involves the following three processes (Trudgill 2011, 20-2):

- the regularisation of irregularities, e.g. in English *cows* as the plural of *cow*, instead of earlier *kine*;
- an increase in lexical and morphological transparency, e.g. *twice* and *went* are less transparent than *two times* and *did go*, so replacing the former by the latter represents an increase in transparency;
- loss of redundancy, of which there are two types: a. syntagmatic redundancy or the repetition of grammatical information, e.g. grammatical agreement on adjectives; b. paradigmatic redundancy or the morphological expression of grammatical categories, such as number, case, tense, aspect, voice, mood, person, and gender.

"Complexification" is essentially the opposite of simplification and involves the following processes (Trudgill 2011, 62):

- irregularisation;
- increase in opacity (less transparency);
- increase in syntagmatic redundancy;
- addition of morphological categories.

As noted above, "complexification" can arise from long-term, stable language contact in which both languages are learned predominantly by children, as opposed to adult learners. This primarily concerns the addition of morphological categories in such contact situations, where new categories are copied from one or more neighbouring languages into another language, but which do not replace other categories but rather are then found in addition to these (Trudgill 2011, 27).

Many of these tendencies have been confirmed in quantitative studies (e.g. Bentz, Winter 2013; Sinnemäki 2009; Sinnemäki, Di Garbo 2018 among others), and the underlying assumptions of Trudgill (2011) have also been used to try to unravel prehistorical settlement patterns in South Asia (e.g. Peterson 2022). The present study represents a further step in this direction.

3 The Sample and the Database

In this section we discuss the choice of languages in our sample (§ 3.1) as well as the features in our database (§ 3.2).

3.1 The Sample

The major difference between our study and Miestamo (2005), the most exhaustive typological study of negation we know of, is with respect to the sample on which it is based. Miestamo (2005) aims to be a representative and areally and genealogically balanced database of negation in human languages. As such, that sample has been compiled taking genealogical and areal biases into account. In contrast, our primary aim is to describe negation in as many languages from as many regions and language families in mainland South Asia as possible with the secondary goal of identifying signs of language contact in the data. For practical reasons, we have not yet been able to include Trans-Himalayan and Tai-Kadai languages in our database, but we hope to add languages of these two families soon. Our sample is therefore of an entirely different nature than Miestamo's and is basically one of convenience, essentially using any grammars for any of these languages which were detailed enough for us to get the necessary information on negation for the respective language, although every attempt has been made to include as many grammars from all branches of all families as possible. The present study provides an overview of the database in its current form.

For each language, the database currently contains only one variety. For languages which have a well described standard variety, such as Hindi or Kannada, it is this variety which we have documented. For others, such as Kharia, Northwestern Kolami, Gta? etc., it is the variety described in the grammar which we used. We hope to add further (dialectal) varieties at a later date.

Unfortunately, negation is not dealt with in equal detail in the grammars we consulted, so that not all of our questions could be answered definitively for all languages. In order to maintain a consistent level of representativeness for all of the languages contained in our sample, we therefore excluded all languages from our database for this study for which we did not have sufficient data; our lower limit for inclusion in the present study was set at at least 66% of the features in the database (see § 3.2). With presently 25 features in the database to be described, this means that data for at least 17 features (= 68%) was required for the inclusion of the respective language in our sample. This narrowed the database down to 39 languages. These 39 languages and their respective genealogical information are given in Appendix A. Their approximate locations, which have been taken from Glottolog (Hammarström et al. 2021) and mapped with the help of *lingtypology* (Moroz 2017), are given in Appendix B.

Despite not being completely balanced, the study nevertheless includes languages from all three major families other than Trans-Himalayan and Tai-Kadai, i.e. Indo-European, with 15 Indo-Aryan languages and one Iranian language (Balochi, spoken in Pakistan), 11 Dravidian languages, 9 Munda languages, and the two isolates Nihali and Kusunda. It also includes languages from all major branches of the three major language families. Furthermore, with 39 languages, the sample contains data for ca. 11.4% of the 341 languages of South Asia (i.e. India, Bangladesh, Bhutan, Nepal and Pakistan) without the "Sino-Tibetan" and "Kra-Dai" languages of South Asia as listed in the *Ethnologue* (Eberhard, Simons, Fennig 2021) and thus provides a good overview of the various negative-marking strategies found in the region.

3.2 The Database

In his studies of negation, Miestamo (2005; 2013a; 2013b) deals with negative marking and its relation to affirmative constructions from a typological perspective, and we largely follow him in the present study. We therefore begin with a brief introduction to the central concepts relevant to negation and the distinctions which Miestamo makes and in which we follow him, while also discussing the differences between our study and his with respect to the database. Miestamo defines a "standard negation" or "SN" construction as follows:

> A SN construction is a construction whose function is to modify a verbal declarative main clause expressing a proposition *p* in such a way that the modified clause expresses the proposition with the opposite truth value to *p*, i.e. ~*p*, or the proposition used as the closest equivalent to ~*p* in case the clause expressing ~*p* cannot be formed in the language, and that is (one of) the productive and general means the language has for performing this function. (Miestamo 2005, 42)

We follow Miestamo's definition of standard negation in the present study but expand the object of our investigation to include negative imperatives and other negative non-indicative categories as well as suppletive negative copular verbs to check these for potential areal clusters. We also include a discussion of the so-called 'zero negation' in Dravidian, in § 3.3, as it appears to be unique in the languages of the world (e.g. Miestamo 2010; Pilot-Raichoor 2011) and as such should not be lacking in a discussion of negation in South Asian languages.[3]

3 However, as zero-negation presently only occurs in Kannada in our sample, it is not yet included in our database but can be added at a later date as more languages are incorporated into the database.

The following discussion illustrates the individual distinctions made in our database with languages from our sample. In doing so, it also illustrates the types of negative constructions found in our data. While not all distinctions are illustrated here, all major negation types are illustrated, as well as some minor but common variations of these different types. It should therefore be sufficient to give the reader a general impression of the different negative constructions found in the subcontinent south of the Himalayas.

The primary distinction with respect to negative marking and its relation to affirmative marking is what Miestamo refers to as **symmetric** vs **asymmetric** structures. Symmetric structures are those which show no structural differences between the affirmative and the negative constructions other than the addition of the negative marker(s) in negation. A simple illustration of this is given in example (2) from Sadri (Indo-Aryan), where the only difference between the affirmative (2a) and the negative (2b) is the absence vs presence of the negative marker *ni*.

Sadri (Indo-Aryan: Jharkhand, Chhattisgarh, Odisha)

(2) a. *bujh-on=a*
understand-PRS.1SG=NAR
'I understand'

b. ni *bujh-on=a*
NEG understand-PRS.1SG=NAR
'I don't understand'

However, in asymmetric constructions other differences are also found. This can be seen in example (3) from Konkani. In the affirmative (3a) the finite verb is marked by the future-tense marker *-tøl*, to which the PNG marker *-õ* '1SG.M' attaches. In contrast, in the negative (3b) the main verb is a participle (i.e. non-finite) and marked as masculine singular (=*c-ɔ* 'FUT.PART-M.SG');[4] this form is then followed by the negative copula in the present tense, marked for 1st person, singular. In other words, the presence of *nã̃* 'I am not' in the negative construction is not the only difference between the two forms, as the marker of future tense is different in both, and the finite status of the main verb is also different in the affirmative and negative.

4 The form *-ɔ* marks only masculine, singular; the 1st person, singular is marked by nasalisation.

Konkani (Indo-Aryan: Goa, Maharashtra, Karnataka, Kerala)

(3) a. *rig-tɵl-ɔ̃*
enter-FUT-1SG.M
'I will enter'

b. *rig=c-ɔ* *nã*
enter=FUT.PTCP-M.SG NEG.COP.PRS.1SG
'I will not enter'

Different types of asymmetry are possible, such as **constructional asymmetry**, as in example (3) from Konkani, where the respective affirmative and negative forms are different, but the individual categories of the paradigm as a whole are the same in both the affirmative and negative, i.e. there is a positive and a negative form for the future tense in Konkani. There can also be **paradigmatic asymmetry**, e.g. in Kannada in example (4), where a distinction made between the future (4a) and the present (4b) tenses in the affirmative is lost in negation (4c). In short, the affirmative and negative paradigms are different with respect to the temporal distinctions they make, in addition to the asymmetric construction.

Kannada (Dravidian: Karnataka)

(4) Present affirmative

a. *nānu* *māḍ-utt-ēne*
1SG do-PRS-1SG
'I do'

Future affirmative

b. *nānu* *māḍ-uv-enu*
1SG do-NPST-1SG
'I will do'

Present / future negative

c. *nānu* *māḍ-uv-ud=illa*
1SG do-NPST-NMLZ=NEG.COP
'I do not / will not do'

In the present study we are primarily interested in constructional symmetries/asymmetries and will not generally refer further to paradigmatic asymmetries, with the exception of the following type, which is directly related to the forms themselves: e.g. in the South Munda language Gutob, TAM markers have different values in the affirmative and negative paradigms, as shown in **[tab. 1]**. In other words, in this kind of paradigmatic asymmetry the value of the individual TAM markers differs with respect to polarity. Miestamo refers to this kind of system as "paradigmatic displacement" (2005, 55).

Table 1 Negation in Gutob (Munda: Odisha) (Voß forthcoming)

	Affirmative		**Negative**	
	middle	**active**	**middle**	**active**
FUT	*-loŋ*	*tu*	*-a*	*Ø*
PST	*-gV*	*-oʔ*	*-to*	
HAB	*-to*		-	
IMP	*-a*	*Ø*	*-gV*	*-oʔ*
OPT	*-eʔ*		*-eʔ*	

For example, in the future affirmative in Gutob in **[tab. 1]** we find the markers *-loŋ* in the middle voice and *tu* in the active, whereas the corresponding tense markers in the negative are *-a* and zero (*Ø*), respectively. These last two markers are also found in the affirmative paradigm, however as markers of the affirmative imperative, not the future. Furthermore, the affirmative past middle marker *-gV* and active-voice marker *-oʔ* are also found in the negative paradigm, where they however mark the negative imperative, not the past tense etc.

Miestamo (2005) makes further distinctions in his study, such as the different types of asymmetric categories with respect to the finite status of the main or auxiliary verb (Type A/Fin), or types of grammatical categories involved in the asymmetry (Type A/Cat), such as TAM, evidentiality, voice, person and number etc. As ours is a preliminary typological study of negative constructions in South Asian languages and we are primarily interested in general patterns involving symmetry vs asymmetry, these subcategories will only be referred to in passing where relevant, and finer distinctions such as these will not be dealt with here in any systematic fashion in the database. We hope to add these at a later date.

Another basic distinction, made in both Miestamo (2005) and Bhatia (1995) and which we also make here, is the type of negative marking in a particular construction and its position with respect to the main verb. For example, if the negative marker is an affix/clitic, a feature which we have borrowed from the GramBank consortium (feature GB107 in our database),[5] we would like to know whether it is a prefix/proclitic or a suffix/enclitic.[6] Furthermore, if the negative mark-

5 https://glottobank.org/#grambank.

6 We consciously chose not to differentiate between affixes and clitics in our database, as the criteria for differentiating between these two categories, if a distinction is made at all in the respective studies, are not always clearly stated, and in many cases different authors working on the same language come to different conclusions with respect to their status.

Further difficulties surfaced with respect to whether or not a negating element was a 'particle' or a bound form (see further below in the main text). Often authors were somewhat inconsistent in their treatment of these units as one type or another, so that we had

er is a suffix/enclitic, we note in the database whether it is word-final or if it is followed by markers of other categories, such as PNG. Indicative negation in Nepali is, for example, generally expressed as a suffix, following the verbal root and TAM markers but preceding (or fusing with) person/number markers. It is never word-final, except when it fuses with person/number marking (1st person singular) or with zero person/number markers (3rd person, singular), hence we consider it to be a non-final negative suffix. This is shown in **[tab. 2]**, adapted from Matthews (1998, 94) for the past tense of the verb *gar-* 'do'.

Table 2 Past-tense negation in Nepali (Indo-Aryan: Nepal, Sikkim, Bhutan) (adapted from Matthews 1998, 94)

	Affirmative	**Negative**
1SG	*gar-ẽ*	*gar-i-nã̃*
2SG	*gar-i-s*	*gar-i-na-s*
3SG.NF	*gar-y-o*	*gar-e-na-Ø*
3SG.F	*gar-ĩ*[*]	*gar-i-na-Ø*
1PL	*gar-y-ə̃w*	*gar-e-n-ẽw*
2PL	*gar-y-əw*	*gar-e-n-əw*
3PL.NF	*gar-e*	*gar-e-na-n*
3PL.F	*gar-i-n*	*gar-i-na-n*

[*] Written with a long <ĩ>, however vowel length is not phonemic in Nepali and Matthews (1998, 3-4) writes that there is no difference in pronunciation between <i> and <ĩ>.

In addition to the fact that person/number markers differ to some extent between the affirmative and negative forms, which is not of concern here at the moment (but see further below), the negative suffix *-na* in Nepali fuses with person/number agreement in **[tab. 2]** in the 1st person singular but not elsewhere. Furthermore, where it does not fuse with person/number marking, it is clear that person/number marking follows the negative suffix *-na*. We therefore take *-na* to be a non-final suffix in Nepali in standard negation.

As noted above, we also include non-indicative negative marking in our database, although this is not standard negation, as we wished to differentiate between those languages with only one type of negative marker and those with various markers based on mood. For example, the injunctive in Nepali is negated through the word-initial prefix *na-* (Matthews 1998, 197), not a suffix as in the indicative (cf. e.g. *ma gar-ũ̃* 'may I do?' vs *ma na-gar-ũ̃* 'may I not do?'). In other words,

to pick one of the alternatives, and in a few further cases we disagreed with an author's decision. Here we took a variety of factors into account, including the mobility of this unit in the sentence, including in 'poetic' or other special language (e.g. did it necessarily appear before or after the verb?), whether it could receive independent stress etc.

here we have two different strategies for marking negation based on differences in mood, one in the indicative (suffixed *-na*), one in the injunctive (prefixed *na-*), both of which are encoded in the database.

Other types of negative marking include suppletion and two marking features taken from the GramBank consortium, namely inflecting words such as negative copulas (GB298, 299) and non-inflecting words, so-called 'particles'.[7] Particles can also differ from language to language. One type, the simplest of all types in our database, is found e.g. in Maithili, where negation is always marked by the particle *nəi* (in formal and written styles, *nəhi*), which is usually positioned before the verb.[8] There are no alternative forms based on TAM, no suppletive negative copulas, and no asymmetrical constructions. Consider the examples in (5) and (6), from Yadav (1996, 305-6).

Maithili (Indo-Aryan: Bihar, Nepal)

(5)	*chora*	*nəi*	*sut-əit*	*əich*
	boy	NEG	sleep-IMPF	AUX.PRS.3NH
	'The boy does not sleep'.			

(6)	*nəi*	*j-o!*
	NEG	go-IMP.2NH
	'Don't go!'	

In other languages, the morphosyntax of negative particles can be somewhat more complex, even ignoring here differences in negative markers with respect to mood. For example, in the South Munda language Kharia, indicative negation is marked by the particle *um*, which generally appears directly before the predicate. In this case, the enclitic subject index in all persons except the 2nd person singular, non-hon-

7 We deviate here somewhat from GramBank with respect to the definition of "inflecting words", which we consider to be all words that can either be used by themselves as predicates, with finite verbs, or which e.g. can be used as light verbs to form acceptable predicates in a language requiring predicates to have a verbal element, i.e. a copula. This is independent of whether or not these units are marked for person, number, TAM etc.

We also differ in our analysis in some cases from Miestamo (2005) with respect to whether an element is a negative auxiliary or a negative particle. For example, Miestamo (2005, 78-9) considers Kannada *illa* 'am/is/are not' to be a suffix (see also Miestamo 2005, 141 in this respect), however this is more an artefact of the writing system than an indication of the status of this unit as a suffix. *illa* is in fact the negative copula and enclitic in this position. Since we consider a 'finite form' in this study to be a word which can either be used as a main predicate in its own right or which functions as a light verb to make non-verbal predicates acceptable as main predicates, such as *illa* in Kannada, we view this form as finite.

8 Although it can take other positions for stylistic purposes, such as in poetry (Yadav 1996, 387-8).

orific, obligatorily 'moves' away from its position following the predicate (7a) and attaches to the negative particle (7b), from Peterson (2011, 335). With the 2nd person singular, non-honorific, however, this index may optionally attach either to the predicate or to the negative particle, as in (8), from Malhotra (1982, 285).

Kharia (South Munda: Jharkhand, Chhattisgarh, Odisha)

(7) a. *ter[=e]=iɲ*
give=ACT.IRR=1SG
'I will give'.

b. *um=iɲ* *ter=e*
NEG=1SG give=ACT.IRR
'I will not give'.

(8) a. *ubhroŋ* *um=em* *ɖe=na*
these.days NEG=2SG come=MID.IRR
'These days you do not come'.

b. *ubhroŋ* *um* *ɖe=na=m.*
these.days NEG come=MID.IRR=2SG

This type of variable marking with respect to person and number is found in our corpus only in Munda languages such as Kharia (South Munda), Santali, Mundari and Ho (North Munda), and only in one region, namely Jharkhand, Chhattisgarh and Odisha, hence we did not encode this 'movement' in the database. If required, this can easily be added to the database at a later date.

Negation is also marked periphrastically in many languages, generally with a non-finite form of the main verb and a finite auxiliary. Examples of these are given in (9) from Konkani. While for the most part these periphrastic formations represent asymmetric constructions which differ from the affirmative forms in more than one way, the past tense in Konkani in (9a) is symmetric, as the only difference between affirmative and negative is the presence of the negative copula in negation. In contrast, the future, present perfect and present tense are all asymmetric constructions, as the form of the main verb is different in the affirmative from that in the negative, in addition to the negative auxiliary (9b-9d).[9] Only in (9a) are both parts of the negative predicate 'finite', whereas in all other negative forms only the copula is finite while the main verb is non-finite.

9 There is a further asymmetry in the present perfect with respect to gender, which is expressed in the affirmative but not in the negative. Otherwise, gender is expressed either in both the affirmative and negative forms (past tense, future tense) or in neither of these (present tense), so that there is no asymmetry in these other categories with respect to gender. This was not noted specifically in the database.

Konkani (Indo-Aryan: Maharashtra, Goa, Karnataka, Kerala)

(9) a. Past tense – main verb is finite in the negative

rig-l-ɔ̃	*rig-l-ɔ̃*	*nã*
enter-PST-1SG.M	enter-PST-1SG.M	NEG.COP.PRS.1SG
'I (M) entered'	'I (M) did not enter'	

b. Future tense (= (3) above) – main verb is a participle in the negative

rig-təl-ɔ̃	*rig=c-ɔ*	*nã*
enter-FUT-1SG.M	enter=FUT.PTCP-M.SG	NEG.COP.PRS.1SG
'I (M) will enter'	'I (M) will not enter'	

c. Present perfect – main verb is an infinitive in the negative

rig-lã	*rig-ũk*	*nã*
enter-PERF.1SG.M	enter-INF	NEG.COP.PRS.1SG
'I (M) have entered'	'I have not entered'	

d. Present tense – the main verb consists only of the stem in the negative*

rig-tã	*rig=nã*
enter-IPFV.1SG	enter=NEG.COP.PRS.1SG
'I enter'	I don't enter'

* Although these two elements are written together as one word in the negative present tense, the verb stem can stand on its own as a separate word in some environments (including but not restricted to the imperative). We therefore consider the negated present tense to consist of the stem and the enclitic negative copula.

With respect to mood, we also noted for each language whether different negative strategies were found based on any TAM categories, not just mood. For example, Bengali shows an asymmetry in the indicative in the present and past perfect: The Bengali indicative normally shows symmetry between affirmative and negative paradigms, the only difference being the verb-final enclitic *=na* in the negative, as in (10a) vs (10b). However, the present and past perfect are asymmetric; here the same marker that is used to mark the present tense in the affirmative and negative combines with a different negative marker, *=ni*, to negate the present and past perfect, as shown in (11a-c). The two perfect categories thus show constructional and paradigmatic asymmetry and, like the Gutob data (see **[tab. 1]**), are an example of paradigmatic displacement.

(10) a. *kor-i*
do-PRS.1
'I/we do'

b. *kor-i=**na***
do-PRS.1=NEG
'I/we do not do'

(11) a. *kor-e-chi*
do-LNK-PRS.PERF.1
'I/we have done'

b. *kor-e-chilam*
do-LNK-PST.PERF.1
'I/we had done'

c. *kor-i=**ni***
do-"PRS".1=NEG.PERF
'I/we have/had not done'

As a last entry in the database, we noted whether asymmetric negative constructions were found at all in a language, in order to catch any possible types of asymmetry which may be found in a particular language but have not been treated systematically in the database. This is the case for example with the past tense in Nepali shown in **[tab. 2]** above, where PNG markers differ to some extent between the affirmative and negative forms. The past tense is indicated by one of the allomorphs *-e* / *-i* / *-y*, followed by the negative marker *-na* and PNG marking, which differs for some forms, such as *gar-y-o* [do-PST-3SG.NF] 'he did' vs *gar-e-na-Ø* [do-PST-NEG-3SG.NF] 'he did not do' or the corresponding plurals *gar-e* [do-PST.3PL.NF] 'they did' vs *gar-e-na-n* [do-PST-NEG-3PL] 'they did not do' (cf. once again **[tab. 2]** above).

Similar to Miestamo (2005, 58-9) we ignored minor phonological differences between affirmative and negative forms which were not connected to an identifiable function. For example, in Nepali the copula *ho* 'is' has the negated form *hoi-na* 'is not', not the expected form **ho-na*. However, this 'suffix' *-i* cannot be assigned any function, at least not from a synchronic perspective. As we are clearly not dealing here with suppletion, and as this *-i* has no identifiable function, this difference was not documented in the database.

We also did not document asymmetries in our database that are not related to the verb phrase, such as variations in case marking between the affirmative and negative. While no such examples came to our attention, we made no systematic attempt to document such features.

Summing up, we documented the following features with respect to negation:

- whether negation can be marked by a particle, inflecting word (e.g. negative copula as an auxiliary) or a clitic/affix, as well as the position of this last type. Also, if this unit is a suffix/enclitic, whether this marker is word-final (e.g. Bengali) or non-word-final (e.g. Nepali);
- whether copular or other verbs can be marked as negative through suppletion;
- whether negation can be marked by an inflecting word together with a finite predicate, a participle, an infinitive or another type of (non-finite) verb form and whether this negative construction is asymmetric;
- whether there are any different negation strategies based on TAM categories and if so, which and whether these are cases of asymmetric negation;
- whether TAM markers with the same form have different TAM values in affirmative and negative categories, and finally
- whether there is any asymmetric negation in the language, in order to locate possible asymmetries not included above.

The individual features documented are given in Appendix C.

3.3 'Zero Negation' in Dravidian

Various South Dravidian languages such as Toda, Kannada and Tamil show a special negative form which appears to be unique crosslinguistically (Miestamo 2010; Pilot-Raichoor 2011). What makes this Dravidian construction crosslinguistically unique is that it consists only of the stem and PNG marking, with no further marking, including no overt negative marking. This is in contrast to other finite verb forms, where a TAM marker intervenes between the stem and PNG marking. This is shown for Literary Tamil and Old Kannada in **[tab. 3]** for the Tamil verb *paṭi-* 'learn' and the Old Kannada verb *nōḍ-* 'see'. Thus, the negative is quite literally 'zero marked'; **[tab. 4]** illustrates this for Modern Kannada for the verb *māḍ-* 'do'.[10]

Table 3 The zero negative in comparison with affirmative finite forms in Literary Tamil and Old Kannada (from Pilot-Raichoor 2011, 269)

	Literary Tamil			Old Kannada		
	Root	**Tense**	**Person**	**Root**	**Tense**	**Person**
Past	*paṭi*	*-tt-*	*ēṉ, āy* etc.	*nōḍ*	*-id-*	*eṃ, ai* etc.
Future	*paṭi*	*-pp-*	*ēṉ, āy* etc.	*nōḍ*	*-uv-*	*eṃ, ai* etc.
Negative	*paṭi*	*-Ø-*	*ēṉ, āy* etc.	*nōḍ*	*-Ø-*	*eṃ, ai* etc.

Table 4 The zero negative in Modern Kannada (from Zydenbos 2020, 209)

Singular		**Plural**	
1SG	*māḍ-enu*	1SG	*māḍ-evu*
2SG	*māḍ-i*	2SG	*māḍ-iri*
3SG.M	*māḍ-anu*	3SG.HUM	*māḍ-aru*
3SG.F	*māḍ-aḷu*		
3SG.INAN	*māḍ-adu*	3SG.INAN	*māḍ-avu*

While the origins of this construction were openly debated by specialists in Dravidian linguistics in the 19th century, this discussion appears to have more-or-less ended soon thereafter, reappearing only briefly in Bloch (1935) and Master (1946) before once again disappearing from academic discourse. It was not until Pederson (1993) and Pilot (1997) that the topic was once again revived, with both authors com-

10 Sridhar (1990, 227-8) assumes an *-e/-a* negative marker in Kannada, appearing between the stem and PNG marking. However, as Pilot-Raichoor (2011, 276-7) shows, this interpretation is incorrect, as this *-e/-a* is part of the PNG marking. In fact, the PNG markers found in the zero negation construction in Modern Kannada are the same as those found in the future tense (compare e.g. the forms found in the table in Zydenbos 2020, 65 for the future tense with those of the zero negative in Zydenbos 2020, 209).

ing to quite different conclusions with respect to its origin. With the appearance of Miestamo's (2005) monograph on negation, the topic has now become part of the larger typological discussion, and has since been dealt with in at least two further studies, Miestamo (2010) and Pilot-Raichoor (2011).

Despite its unique status among the world's languages, the zero negative is excluded from Miestamo's study (2005, 121), as it is not standard negation, due to its somewhat marginal status in these languages. For example, Zydenbos (2020, 209) writes that the forms of the zero negative in Modern Kannada "are **absolute** negations, negating the occurrence of an action or process categorically, without reference to a specific point in time". This is thus not the negation of a present, past, or future (etc.) action or state, but more of a categorical statement of the type "I have never done such a thing, I am not doing it now, and I will never do it" (Zydenbos 2020, 209; emphasis in the original).

While we include a brief discussion of zero-negation here in order to present as many different types of negative constructions in South Asia as possible, it is only found in our sample in Kannada. Zero negation is therefore presently not documented in the database.

4 Results

To visualise the data, we used SplitsTree4 (version 4.15.1) (Huson, Bryant 2006) to construct a NeighborNet network[11] and an unrooted UPGMA tree for the sake of comparison.[12] These are shown in **[figs 1-2]**. These figures are not offered as proof of any clusters in the region but are merely intended to help visualise the data with respect to negation in these languages and to serve as a starting point for further discussion, as these algorithms show a number of clusters – in fact, almost the exact same clusters in both figures – suggesting that it will be worthwhile to take a closer look manually at the underlying simi-

11 NeighborNet (Bryant, Moulton 2004) is often used in contact linguistics to portray the effects of language contact. In these networks, the length of branches corresponds directly to the degree of divergence or 'distance' between individual languages. Instead of trying to find an optimal tree-like format to portray similarities and differences between languages, NeighborNet suggests alternative trees to portray the possible paths which may be taken between two points when there are conflicting signals in the data, as is commonly the case with language contact, but also with language isolates or languages which otherwise lack close relatives, or with data scarcity. Cf. Holman et al. 2011.

12 UPGMA (Unweighted Pair Group Method with Arithmetic Mean), attributed to Sokal, Michener 1958 (cf. Wikipedia, "UPGMA", `https://en.wikipedia.org/wiki/UPGMA#cite_note-`). This clustering algorithm is a distance-based means of portraying similarities/differences between languages which assumes a constant rate of change for all languages.

larities of these clusters.[13] Therefore, in the following discussion we take a closer look at the clusters in the two figures and the typological features which motivate them.

The following clusters list the member languages of the individual clusters which are found in both figures. Here cluster-internal differences, such as the somewhat different position of Malto, Kurukh and Gujarati on the left-hand side in both figures, will not be commented upon further, as we are only interested here in the general groupings and their features. The geographical distributions of these four clusters are illustrated in Appendix D. The respective cluster numbers are indicated in the figures.

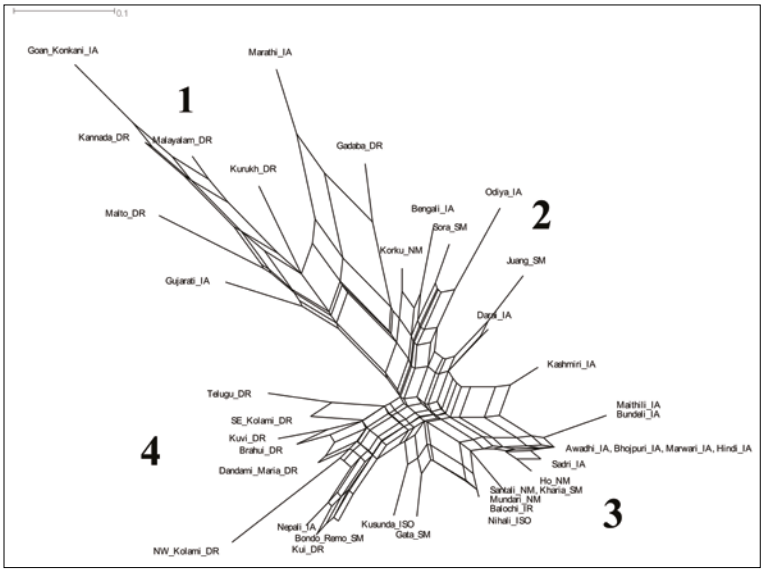


Figure 1 A NeighborNet representation of negation in South Asian languages (25 features in 39 languages)

Cluster 1 – This cluster is the most conspicuous in both figures. It consists of various Dravidian languages (Gadaba, Kannada, Malayalam, Kurukh and Malto), although not all (e.g. Telugu, Southeastern Kolami, Kuvi, Kui, and Dandami Maria are not included), and three Indo-Aryan languages, namely Goan Konkani, Marathi and Gujarati, all three of which are spoken in western India.

13 As Borin et al. (2021, 228) so aptly formulate it: "We see the function of the computational tools […] primarily as 'filters' helping the linguist to separate small amounts of wheat from large volumes of chaff, not by identifying the wheat directly, but by identifying those parts of the data where it is likely to hide and be found by manual inspection".

Cluster 2 – This rather heterogeneous cluster consists of the languages Odiya, Bengali and Darai (Indo-Aryan), the South Munda languages Juang and Sora, and the North Munda language Korku, spoken in central India.
Cluster 3 – To this very large cluster belong the Indo-Aryan languages Awadhi, Bhojpuri, Sadri, Kashmiri, Marwari, Hindi, Maithili, and Bundeli, spoken in western, central and eastern North India; the Munda languages Ho, Mundari, Santali and Kharia, spoken in Jharkhand in eastern central India; Gtaʔ spoken considerably further to the south, along the border with Andhra Pradesh; the Iranian language Balochi, spoken in Pakistan; and the isolates Kusunda (central Nepal) and Nihali (western central India).
Cluster 4 – In this cluster we find the Central Dravidian language Southeastern Kolami, the South Central Dravidian languages Telugu, Dandami Maria, Kuvi and Kui, and the South Munda language Bondo/Remo, all spoken in southern eastern/eastern central India; the North(west) Dravidian language Brahui, spoken in Pakistan, and the Indo-Aryan language Nepali.

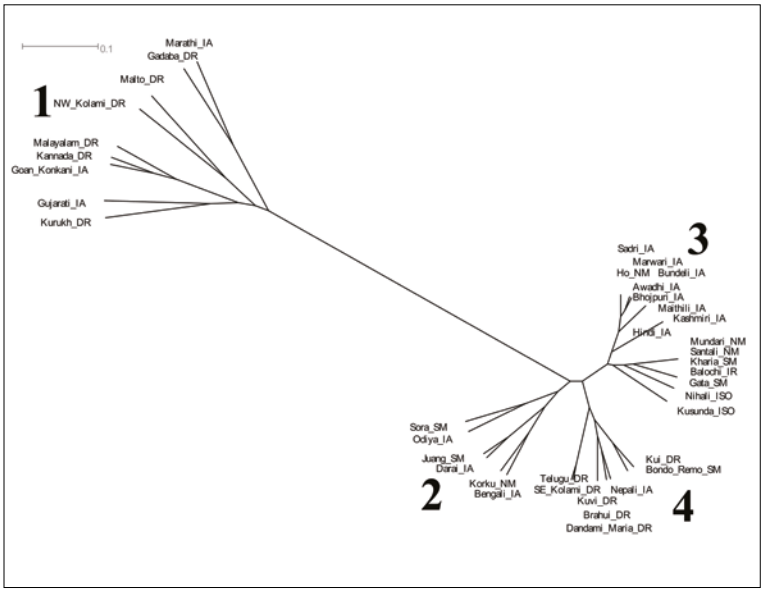


Figure 2 A UPGMA representation of negation in South Asian languages (25 features in 39 languages)

Of all 39 languages it is only the Central Dravidian language Northwestern Kolami which is in different clusters in the two figures: In Cluster 4 in **[fig. 1]** and Cluster 1 in **[fig. 2]**. For ease of presentation, it will be discussed together with Cluster 4 in § 5, where its commonalities with Cluster 1 will also be highlighted.

With respect to the different negative markers, of the 39 languages in our sample 22 languages make use of affixes in negation in at least one category, 9 languages make use of a negative auxiliary verb in at least one category, and 23 of the languages in our corpus make use of a negative particle in at least one category. Affixes and negative particles are thus very evenly distributed in our corpus (22 and 23 languages, respectively), and both are more than twice as common as negative auxiliaries. These figures total more than the 39 languages in our sample as 13 languages combine different types of markers to some extent, e.g. Darai, which uses both the prefixal type as well as the negative particle type in different categories. Of these languages which combine different types, 11 use two different negative strategies while a further two – Marathi and Korku – use all three strategies.

25 languages make use of only one of these three strategies (i.e. affix, auxiliary or particle) in negation:[14] nine languages use only affixes, although some of these languages do make use of different affix types, such as Nepali, which has both prefixes and non-word-final suffixes, 13 languages use only negative particles, and three make exclusive use of negative auxiliaries.

As will be discussed below, the distribution of the languages in our sample with respect to these three types is not entirely random. The most obvious example are the three languages which negate only with negative auxiliaries, namely Malayalam and Kannada (both South Dravidian) and Konkani (Indo-Aryan), of which many speakers are bilingual with Kannada. Also, 10 of the 13 languages which make exclusive use of a negative particle are spoken in a more-or-less contiguous area from Rajasthan (Marwari) via central North India (Hindi) to Bihar and Jharkhand (several Indo-Aryan and Munda languages), with the other three far to the north (Kashmiri), southwest (Nihali) or west (Balochi). Similarly, with three exceptions, namely Nepali, Kusunda and Brahui, the other six languages which negate exclusively through affixes are all found in central and eastern India. Clearly, genealogical tendencies and areal pressure both play a role in the distribution of these features.

The significance of the data which is visualised in **[figs 1-2]**, and above all the features behind these clusters, are discussed in detail in § 5, where we show which areal patterns are most likely due to language contact, and suggest, where possible, what type of language contact in the past has led to the observed results.

14 If we include Odiya here, for which we could not be sure that it only has one category, then there are 26.

5 Analysis

In the following we discuss each of the individual clusters with respect to the predominant negating strategies documented in the database and what information this provides us with respect to historical language contact.

5.1 Cluster 1

What is most notable about this cluster is that it consists of the three western Indo-Aryan languages Konkani, Marathi and Gujarati, and just to the south of these the South Dravidian languages Kannada and Malayalam. While Gadaba, Kurukh and Malto are also Dravidian languages, Gadaba is spoken in eastern Andhra Pradesh and Kurukh and Malto are spoken much further north and east, primarily in western and northeastern Jharkhand, respectively. We therefore begin here with the western Indo-Aryan and South Dravidian languages in this cluster.

The most notable aspect of this cluster in **[fig. 1]** is the exposed position of Konkani, Kannada and Malayalam. The reason for this likely lies in the fact that these languages make exclusive use of negative auxiliaries (GB298) generally deriving from a suppletive negative copula (SA075, SA076). Negation here thus consists of finite (SA078) or non-finite (SA079, SA080, SA081) forms of the lexical predicate, of which most are asymmetric constructions (SA079a, SA080a, SA081a, SA086). Also, there are different negating strategies in all three languages for TAM categories (SA083, SA084), and again generally asymmetric constructions (SA083a, SA084a). This is especially true of Konkani and Kannada.

Consider the data in **[tabs 5-6]**, which illustrate the affirmative and negative categories in the indicative and the imperative in both of these languages. The form of the lexical predicate in negation (i.e. infinitive, participle, finite form) is given in bold print directly above the corresponding negative verb form in both tables.

Although there are other Indo-Aryan languages with negative copulas, it is much less common elsewhere in Indo-Aryan to use these as a major negative strategy than in Konkani, and to a much lesser extent in Marathi (see further below), and Konkani is one of only three languages to make exclusive use of negative auxiliaries in negation – the other two being, crucially, Kannada and Malayalam.[15] Otherwise,

15 It is is also found in some dialects of Sadri (first author's own data), but not in the standard dialect, from which the Sadri data for this study were taken. Note that Miranda (2003, 760) gives a short list of examples of Kannada influence on Konkani, one of which is negation, although very brief and rather vague: "Non-finite forms of the verb are used in the various tense-aspect forms of negative sentences".

negative particles and affixes are generally used in Indo-Aryan. It is therefore clear that Konkani has developed the negative patterns illustrated in **[tab. 5]** through contact with Kannada.

Table 5 Affirmative and negative strategies in Goan Konkani (based on Almeida 2004, 98-9 and examples throughout that book)[16]

	Affirmative form	**Negative form**
		Simple finite verb plus negative copula
Simple Past	*rig-l-ɔ̃* [enter-PST-1SG.M]	*rig-l-ɔ̃ nã*
		Stem plus negative copula
Present	*rig-tã* [enter-IPFV.1SG]	*rig=nã*
Past imperfective	*rig-ta-l-ɔ̃* [enter-IPFV-PST-1SG.M]	*rig=naslɔ̃*
		Future participle (=*cɔ*) plus negative copula
Future	*rig-təl-ɔ̃* [enter-FUT-1SG.M]	*rig=cɔ nã*
		Infinitive 2 (-*ũk*) plus negative copula
Present perfect	*rig-lã* [enter-PERF.1SG.M]	*rig-ũk nã*
Past Perfect	*rig-ləl-ɔ̃* [enter-PST.PERF-1SG.M]	*rig-ũk naslɔ̃*
		Infinitive 1 (-*ũ*) plus specialised form of negative copula
Imperative	*rig*	*rig-ũ naka*

Table 6 Affirmative and negative strategies in Standard Kannada (adapted from Zydenbos 2020, 149-50, 160, 179-82, 184-9) for *māḍ(u)* 'do'

	Affirmative form	**Negative form**
		Verbal noun -*uvud(u)*[i] plus negative locative copula *illa*
Present	*māḍ-utt-ēne* [do-PRS-1SG]	*māḍ-uvud=illa*
Future	*māḍ-uv-enu* [do-FUT-1SG]	
		Present participle -*utt* plus negative locative copula
Present continuous	*māḍ-utt=iddēne* [do-PRS-PRS.COP.1SG]	*māḍ-utt=illa*
		Infinitive in -*al* plus negative locative copula
Simple past	*māḍ-id-enu* [do-PST-1SG]	*māḍ-al=illa*
		Sequential converb[ii] plus negative locative copula
Present perfect	*māḍ-i=ddēne* [do-CVB-PRS.COP.1SG]	*māḍ-i=lla* (**-i-i* > *-i*)
		Infinitive in -*a* plus *bēḍa* 'is not needed/wanted'
Imperative	*māḍu*	*māḍ-a=bēḍa*[iii]

i This form consists of the non-past tense marker *-uv* and the nominaliser *-ad(u)/-ud(u)*.

ii Referred to in Zydenbos (2020) as the "gerund"

iii *bēḍa* is written together with the preceding infinitive, however since it can also stand alone, we consider it here to be enclitic.

16 The present tense is indicated through a lack of overt tense marking following the imperfective marker *-ta*, to which nasalisation (denoting the 1st person singular) then directly attaches. *nã* and *naslɔ̃* in **[tab. 5]** are the forms of the 1st person, singular (masculine) of the negative copula in the present and past tenses, respectively.

Although we argue that Konkani has developed these complex negative strategies through contact with Kannada, even a brief glance at the data in **[tabs 5-6]** shows that the Konkani constructions are not simply direct borrowings from Kannada. To begin with, all negative constructions in Konkani are based on Indo-Aryan morphs, not morphs borrowed from Kannada. Instead, what has been borrowed here is the general pattern of almost entirely asymmetric negative constructions which make use of a negative auxiliary generally deriving from the negative copula. 'Borrowing' of this type, as opposed e.g. to that of simple lexical items, is only possible with speakers who are fluent in both languages. This speaks for a prolonged period of stable bilingualism between Konkani and Kannada, which is also grounded in Kannada's and Konkani's historical relationship (e.g. Miranda 2003, 760).

Furthermore, despite all similarities, there is no exact fit between the individual categories in both languages, which again implies that the respective speakers will have been fluent in both languages and will have been able to 'borrow' structures in a way so as to maintain the TAM distinctions which both languages otherwise show. In other words, while the overarching pattern which was copied into Konkani was one of predominantly asymmetric negation with a specialised negative auxiliary, this occurred in Konkani in a way which was in synch with the overall system of that language and not just a copy of the Kannada structures.

For example, Kannada shows a paradigmatic asymmetry in which the present and future distinction found in the affirmative is lacking in the negative, whereas Konkani shows no such TAM paradigmatic asymmetries, and both the present and the future in Konkani are negated through constructions, neither of which is found in that form in Kannada. Also, while the infinitive followed by an auxiliary is the negative strategy for the perfect in Konkani, it negates the past tense in Kannada.

In fact, the Konkani system is morphologically even more complex than the Kannada system which served as a model for its negative patterns, further showing how the new structures were integrated into the existing grammatical structures of Konkani. With respect to the affirmative categories, Konkani distinguishes person, number and gender in all persons in most TAM categories; in Kannada these are restricted in the affirmative to the 3rd persons. Consider the data for the Konkani and Kannada affirmative future in **[tabs 7-8]**.[17]

17 The alternative forms in Kannada are not related to gender distinctions but are free or regional variants.

Table 7 Affirmative future in Konkani (*kər-* 'do', from Almeida 2004, 77)

	Singular			Plural		
	m	f	n	m	f	n
1	*kər-təl-ɔ̃*	*kər-təl-ĩ*	*kər-təl-ɛ̃*	*kər-təl-ɛ*	*kər-təl-yo*	*kər-təl-ĩ*
2	*kər-təl-ɔ*	*kər-təl-i*	*kər-təl-ɛ̃*	*kər-təl-ɛ*	*kər-təl-yo*	*kər-təl-ĩ*
3	*kər-təl-ɔ*	*kər-təl-i*	*kər-təl-ɛ̃*	*kər-təl-ɛ*	*kər-təl-yo*	*kər-təl-ĩ*

Table 8 Affirmative future in Kannada (*māḍ(u)* 'do', from Zydenbos 2020, 66)

Person	Gender	Singular	Gender	Plural
1		*māḍ-uv-enu / māḍ-uv-e*		*māḍ-uv-evu*
2		*māḍ-uv-e / māḍ-uv-i*		*māḍ-uv-iri*
3	M	*māḍ-uv-anu / māḍ-uv-a*	HUM	*māḍ-uv-are*
	F	*māḍ-uv-aḷu*		
	NHUM	*māḍ-uv-adu / māḍ-uv-udu*	NHUM	*māḍ-uv-uvu / māḍ-uv-avu*

The differences with respect to morphological complexity in the negative are even greater. In Kannada, the entire affirmative paradigm is negated by the invariable form *māḍ-uvud=illa* (cf. **[tab. 6]** above), consisting of the non-present verbal noun *māḍ-uvud(u)* and the invariable negative copula *illa*. By contrast, in Konkani all PNG distinctions are retained for all persons in the negative (except in the present perfect), which e.g. in the case of the future consists of the future participle in *=c*, marked for gender and number (cf. **[tab. 9]**). PNG marking is then marked on the negative auxiliary which follows the participle. The forms of the negative auxiliary are given in **[tab. 10]**. Thus *kər=c-ɔ nã* 'I (M) will not go' etc.

Table 9 The gender/number forms of the future participle in Konkani

Singular			Plural		
M	F	N	M	F	N
kər=c-ɔ	*kər=c-i*	*kər=c-ɛ̃*	*kər=c-ɛ*	*kər=c-yo*	*kər=c-ĩ*

Table 10 The present-tense negative auxiliary in Konkani (Almeida 2004, 98)

Person	Singular	Plural
1	*nã*	*nant*
2	*na*	*nant*
3	*na*	*nant*

Thus, in Konkani no TAM distinctions are lost in the negative of the type found with the Kannada present/future-distinction in **[tab. 6]**, and

all affirmative and negative predicates are marked for the same PNG categories, with the exception of the present perfect, where gender distinctions are lost in the negative (cf. again the discussion of example (3) in § 3 above).

Nowhere else in Indo-Aryan is the copying of a general negating strategy from one language family into another as pervasive as it is along the Konkani-Kannada border, suggesting that this area of contact has been shaped by centuries of highly stable bilingual contact, with the Indo-Aryan-speaking regions slowly but surely progressing southwards and Dravidian-speaking areas gradually receding before them. Even today we find large numbers of Konkani speakers in Karnataka, the state whose official language is Kannada, with syntactic borrowings from Kannada in the Konkani of this region (cf. e.g. Nadkarni 1975). Although Konkani speakers here constitute a minority in most areas, it is nevertheless noteworthy that most Konkani speakers live in Karnataka, and it is only the state of Goa where Konkani speakers predominate (Almeida 1989, 5-7). This type of situation between Konkani and Kannada has thus likely existed for several centuries or perhaps even millennia, although slowly progressing southwards.

Thus, in our view, only a prolonged period of intense bilingual language contact between Konkani and Kannada can account for the development of this type of complex negation in Konkani (cf. Trudgill 2011),[18] although it is not yet possible to say whether large numbers of speakers of both languages learned the other language or whether only one of these two groups was bilingual. We know that the Indo-Aryan-speaking area has been steadily progressing southwards along the west coast since Vedic times (cf. e.g. the discussion of Maharashtrian place names in Southworth 2005, 288-321), however this could be due to extensive bilingualism by Indo-Aryan L1 speakers, by Dravidian L1 speakers, or by both groups. Future research is required here.

The main differences between Marathi and the Konkani-Kannada pattern are that Marathi has all three different types of negative markers, i.e. prefix, particle and also a negative auxiliary (GB107, GB298, GB299; SA071, SA072, SA074, SA075, SA076), whereas Konkani and Kannada only have negative auxiliaries. Marathi also has periphrastic negative constructions (but fewer than Konkani), some of which are asymmetric (SA079, SA080, SA080a), as well as different negative strategies based on TAM, which are asymmetric (SA082, SA083, SA083a, SA084, SA084a, SA086). The situation in Gujarati is schematically similar to that in Marathi, although with some differences.

18 Further evidence for this type of contact scenario is cited in Peterson 2022, e.g. correspondences in the imperative paradigm in both Konkani and Kannada, with identical PNG marking for the 1st person singular and the 3rd person singular and plural, which is otherwise at the very least uncommon in South Asia. Here as well, the morphs in Konkani have not been directly borrowed from Kannada.

On the other hand, Malto, Kurukh and Gadaba are quite unlike the Konkani-Kannada type with respect to periphrastic negative constructions. The first two share with Cluster 1 the presence of a negative auxiliary (in addition to a negative affix) as well as different negative strategies for some TAM categories. Similar comments hold for Gadaba, which however does not have a negative auxiliary. It does however also make use of non-final and word-final suffixes in negation, like the languages in Clusters 2 and 4 (see below), with Gadaba bordering on the southern edge of Cluster 2 and on the northern edge of Cluster 4.

However, schematically Gadaba shares a number of characteristics with Cluster 1 languages, such as the use of different negative strategies for tense-aspect and mood, both of which are asymmetrical. It also makes use of a negated copula with the infinitive to negate the past tense, like Kannada **[tab. 6]** or the perfect of Konkani **[tab. 5]**, however this is not a suppletive form, as it is in those two languages. Its status within this cluster is thus somewhat unclear.

In contrast, the similarities of both Malto and Kurukh with the other members of this cluster are likely coincidental. While they do share many features with other Cluster 1 languages such as negative auxiliary verbs (GB298), suppletive negative copular verbs (SA076), a construction with a negative auxiliary and a participle in an asymmetric construction (SA079, SA079a), and different negative markers based on tense-aspect (SA083, SA083a) and mood (SA084), at least at present we have no reason to assume that this is due to a family bias with the South Dravidian languages Kannada and Malayalam or with the Central Dravidian language Gadaba, nor to areal pressure, as the nearest Indo-Aryan language, Marathi, is spoken at a considerable distance from these two.

5.2 Cluster 2

Cluster 2 is quite heterogeneous with respect to the geographical location of languages. Some, such as Odiya and Bengali, are direct neighbours and very closely related, hence the similarities between these two languages are to be expected. These languages both have negative word-final suffixes, with a particle also found in Bengali (GB107, SA072, SA073). Both also have suppletive negative copulas (SA076) and different negative strategies for certain TA categories (SA082, SA083), which is asymmetric in Bengali (SA083a).

Sora and Juang are quite different with respect to negative marking. In both languages this marker can be a prefix (GB107, SA071), but in Sora it can also be expressed through a suffix, both in word-final and non-word-final position (SA072, 073, 074), similar to Gadaba in Cluster 1 above, whose status in that group is unclear. In Sora but not in Juang we also find suppletive negative copulas (SA076), while in both we find some periphrastic constructions, including asymmetric ones,

with slight differences between the two languages (SA082-086). Juang is also one of three languages in our database which show paradigmatic displacement (SA085), the other two being Gutob, discussed in **[tab. 1]** above, and the present and past perfect in Bengali (examples 10-11).

Korku and Darai do not really fit into this cluster with respect to the form of the negative marker. In Korku it can take the form of a suffix, an auxiliary or a particle, whereas in Darai it can be a prefix or a particle. What they share with the other languages appears to be mostly the presence of different negative marking strategies with respect to TAM categories, at least some of which are asymmetric, although this also holds for many languages from other clusters as well.

In sum, although the negative strategies in Sora and Juang may to some extent have been affected by language contact with Odiya, this presumed influence would appear to be quite weak at best. On the other hand, areal influence can be entirely ruled out with respect to Korku and Darai on geographical grounds. We therefore do not view membership in this cluster as due to areal influence or family bias, with the obvious exception of Odiya and Bengali, but most likely as coincidental similarities among these languages. Also, as noted above for Cluster 1, Gadaba shares with most members of this cluster the fact that it has suffixal negative markers, although here as well similarities to this cluster are rather weak and are stronger with Cluster 4 (below).

5.3 Cluster 3

Cluster 3 for the most part consists of Indo-Aryan and Munda languages spoken in a more-or-less contiguous area stretching from Rajasthan through Uttar Pradesh to Bihar and Jharkhand. In this cluster we also find the South Munda language Gtaʔ, spoken along the border between Odisha and Andhra Pradesh, Kashmiri, and the Iranian language Balochi. What all languages in this cluster other than Gtaʔ have in common is that they possess a non-inflecting negative particle. In addition to this, Kashmiri also has a negative suffix. In contrast, Gtaʔ makes exclusive use of a negative prefix.

This is the only cluster in our present database where we find languages where a negative particle is the only negative-marking strategy (all except Kashmiri and Gtaʔ). Furthermore, apart from Kashmiri, all Indo-Aryan languages of this cluster belong to the so-called ‘Hindi Belt’. Two of these, Bundeli and Maithili, are also the two languages with the simplest negative strategy found in our database, with a negative particle and no further positive values in the database, including no negative copulas, no asymmetric constructions and no different strategies based on TAM.

All other languages in this cluster have different negative strategies based on mood (SA082, SA084); these non-indicative strategies

are asymmetric in all Munda languages other than Ho and in Balochi and Nihali (the latter of which also has an asymmetric negative construction based on tense/aspect, SA083, SA083a), but symmetric in all Indo-Aryan languages with a different non-indicative negative marker (SA084a). Finally, the Munda languages in this cluster, the neighbouring Indo-Aryan language Sadri, and Balochi all have suppletive negative copulas (SA076).

Despite having the same negative-marking strategies as Munda languages, Balochi's similarity to these languages is clearly coincidental, as it is spoken far to the west in Pakistan. It is also genealogically too distant from Indo-Aryan to be due to family bias. Kashmiri, although Indo-Aryan, also belongs to a different group than the 'Hindi Belt' languages and therefore also likely represents an independent retention of this earlier negational strategy (see below).

The isolates Nihali and Kusunda are probably only found in this cluster due to chance similarities in their negative strategies. They are not related genealogically to either Munda or Indo-Aryan, and areal pressure can most likely be ruled out for both.

Despite its relative geographical proximity to the eastern languages of this cluster, the South Munda language Gtaʔ is quite different from the other languages in this cluster in that it does not have a negative particle, the main defining structural characteristic of this cluster, although it does have a suppletive negative copula and different negative strategies based on mood. As especially this last feature is very common, Gtaʔ's inclusion in this cluster is therefore almost certainly due to chance similarities and not to areal or genealogical pressure.

Family Bias, Areal Pressure, or a Bit of Both?

As discussed in Peterson 2022, the eastern part of the 'Hindi Belt' region consists of Indo-Aryan languages which display considerable simplifications in comparison with western Indo-Aryan languages. Peterson argues that these simplifications resulted when large numbers of Indo-Aryan speakers entered eastern India, where their languages quickly became the *lingua franca* of the region. As argued there, this will have resulted in large numbers of speakers – in many regions perhaps a considerable majority of the speakers – being adult learners of Indo-Aryan, which gave rise to a dramatic amount of morphological simplification in eastern Indo-Aryan. It is interesting to note that the Munda languages of this contact area are also found in this cluster. This suggests that contact may be a factor behind the existence of this cluster.

Nevertheless, this is primarily a case of family bias, as the Indo-Aryan languages of this cluster have retained the features from older stages of these languages, going back to OIA, with few negative par-

ticles, differing with respect to mood, and no structural asymmetries. Thus, while this 'simple' negative pattern may be expected in a situation where a large percentage of speakers are adult L2 speakers, family bias alone will suffice to explain the 'Hindi Belt' members of this cluster, especially since many of these are spoken further to the west, where the massive simplifications noted in Peterson (forthcoming) for eastern Indo-Aryan did not take place. Thus, this 'simple' negative strategy is compatible with the predominant negational pattern of this cluster, but is not likely its primary motivating factor.

What remains to be accounted for is the status of the Munda members of this cluster. As Jenny, Weber, Weymuth (2015, 107) note, it is extremely difficult to posit any negative-marking strategy for Austro-Asiatic, as the negative constructions in that family are so diverse. A bias only for the Munda group is, however, equally difficult as it is only in the Munda languages of this group that negation is marked exclusively by means of a particle. While Juang has a negative particle, it also negates through prefixes. Gta? on the other hand negates only through prefixes while Sora negates with both types of affixes. Korku negates with suffixes, but it also has an auxiliary negative verb and a negative particle. As the Munda languages in our sample are found in three of the four clusters determined by both algorithms, this suggests that Munda languages in general cluster with their linguistic neighbours, regardless of genealogical relationships.[19] We therefore assume here that negative-marking strategies in these Munda languages arose through contact with Indo-Aryan.

In sum, the predominant negative-marking strategy in the Indo-Aryan languages of this cluster is due to family bias, while the similar negative-marking strategies of the Munda languages in this cluster is likely due to contact with the eastern 'Hindi Belt' languages.

5.4 Cluster 4

Cluster 4 largely consists of Central and South Central Dravidian languages, but also the South Munda language Bondo-Remo, spoken along the Odisha-Andhra Pradesh border, the northwestern Dravidian language Brahui, and Nepali. In addition, in **[fig. 1]** Northwestern Kolami also belongs to this cluster, although it is in Cluster 1 in **[fig. 2]**.

The members of this cluster all share various structural features with respect to negation. First, all mark negation through an affix (GB107), including Northwestern Kolami, which in all languages except Bondo-Remo is a non-final suffix (SA072, SA073). The data in Bon-

19 Cf. also Borin et al. (2021) with respect to the clustering of Munda languages in general.

do-Remo were unfortunately not explicit enough so that we have no entries for features SA072 or SA073 for this language. Furthermore, all languages show some form of asymmetric negative construction, which in five out of eight languages in this cluster (and also in Northwestern Kolami) is due to asymmetric negative constructions based on mood distinctions. Nepali shows a further asymmetry in the past tense where the PNG suffixes in negation differ to some extent from those in the affirmative.

With the obvious exceptions of Nepali and Brahui, this cluster appears to represent an older (South) Central Dravidian negative-marking strategy which has survived in these languages up to the present. The only other language in this cluster in both **[figs 1-2]** is the South Munda language Bondo-Remo. While this language may well belong to this cluster due to areal pressure from the neighbouring Dravidian languages, this is not entirely clear, as we presently have no data for three of the critical features of this cluster.

It is especially noteworthy that North Dravidian Brahui clusters in this group with (South) Central Dravidian languages, which are spoken at a great distance from the Brahui-speaking region, but does not cluster with Balochi, which virtually surrounds the Brahui-speaking area and which most Brahui also speak.[20] Whether this similarity is due to family bias or to chance definitely warrants further study.

With respect to the ambiguous status of Northwestern Kolami in **[fig. 1]** (Cluster 4) and **[fig. 2]** (Cluster 1), it is worth noting that this language is located in the border region of Marathi and the Dravidian languages of central India and shows features common to both Clusters 1 and 4. Like all other members of Cluster 4 for which we have the respective data, Northwestern Kolami has non-final negative suffixes, shows some form of asymmetry and has distinctive mood-based negative marking. Like most Cluster 1 languages, however, Northwestern Kolami also negates with an inflecting word which derives from the copula, has suppletive negative copular forms, and again, distinctive negative strategies based on mood. Like some other languages of Cluster 1 it also makes use of both prefixes and suffixes in negation. Hence its different status in **[figs 1-2]**.[21]

Thus, while a family bias is likely behind the membership of many Dravidian languages in this cluster, we also see some likely signs of areal pressure at the fringes of this area, with Northwestern Kolami oscillating between this cluster and Cluster 1.

20 We are grateful to an anonymous reviewer for calling this to our attention.

21 While this strongly suggests a high degree of long-term bilingualism between Northwestern Kolami and Marathi, the *Ethnologue* (`https://www.ethnologue.com/language/kfb`) claims that Northwestern Kolami speakers have limited proficiency in Marathi (cf. Eberhard, Simons, Fennig 2021). Further work is required.

Summarising our results of the present section, each of the respective clusters has a 'core group' of languages which share, for the most part, a preferred type or preferred types of negative marking as well as the presence or lack of periphrastic negative constructions and/or constructional asymmetries.

The languages of Cluster 1 most clearly illustrate that negative structures in one language or group of languages, here Konkani (and to a much lesser extent Marathi) have been motivated by structures in a neighbouring language, in this case Kannada, which has very similar structures to those in Malayalam. As the Konkani structures are certainly innovations, while both South Dravidian languages involved show similar structures, we assume that long-term community bilingualism lies behind the imitation of South Dravidian structures in Konkani with native Indo-Aryan morphology. This is in line with the arguments in Peterson 2022, who finds signs of long-term bilingualism for these languages with respect to other features, especially of the nominal system.

In Cluster 3 the so-called 'Hindi Belt' languages (and a few other Indo-European languages) have retained an older system of negation with a small number of negative particles, based on mood, and no constructional asymmetries, i.e. this is a clear example of family bias. However, the Munda languages of this cluster are also quite similar with respect to negating strategies. As Munda languages in general tend to cluster with their geographical neighbours and not with other Munda languages further afield, and as it is only the Munda languages of this group which negate exclusively with a negative particle, we assume that these Munda languages have developed this marking pattern though contact with Indo-Aryan.

Signs of areal pressure are also found with languages which are at the fringes of their respective areas, e.g. Northwestern Kolami, which shares features of Clusters 1 and 4, between which it is located, as well as perhaps Gadaba from Cluster 1, which shares some features with Clusters 2 and 4. But both **[figs 1-2]** include in each cluster languages which are genealogically and geographically quite far removed from the other languages of their respective clusters, showing that the same negative strategies can arise and/or be preserved independently of their linguistic neighbours, despite all genealogical and areal pressure.

6 Summary and Outlook

In this study we present a first typology of negative-marking strategies in South Asian languages, based on a database of 25 structural features from 39 languages belonging to Indo-European (Indo-Aryan and Iranian), Dravidian, and Munda families, as well as the two isolates Nihali and Kusunda. The features documented for each language are large-

ly a subset of those in Miestamo (2005), although we deviate here occasionally from that study as our goals differ somewhat from Miestamo's. The features we have documented in our database include the form of the negative marker, the relation of the negative construction to the corresponding affirmative construction, e.g. whether the negative construction is symmetric or asymmetric, as well as whether there are alternative negative constructions based on TAM or other categories.

In a second step, we make use of two different algorithms to visualise the patterns in the data in a first attempt to determine which languages most likely cluster together and why. Here we discuss the relevant features of the languages in each of the clusters suggested by the two algorithms as well as their genealogical and geographical distributions to determine whether the clustering is due to family bias, areal pressure, both, or merely due to chance similarity.

The data includes examples for all four of the scenarios just mentioned; e.g. the Dravidian languages of Cluster 4 are likely an example of family bias, as this cluster has a clear regional focus in Andhra Pradesh, Telangana, and southern Chhattisgarh and Odisha, while other languages in this cluster are certainly due to chance similarities, such as Nepali.

Cluster 1, on the other hand, provides the strongest example of contact-induced negative marking in our sample. Here, the traditional Indo-Aryan negative marking system, with a small number of negative particles based on mood distinctions and no asymmetries, has been entirely remodelled in Konkani along the lines of the negative-marking strategies found in its Dravidian neighbour Kannada. This is a strong indication that this is due to a situation of long-term, stable bilingualism between Konkani and Kannada which has resulted in the copying of complex negative paradigms from Kannada into Konkani (cf. also Peterson 2022).

Other clusters, such as Cluster 3, involve a combination of both tendencies. The languages of this cluster, most of which belong to the 'Hindi Belt', generally make exclusive use of negative particles to express negation, a clear case of family bias. However, this marking pattern includes not only the 'Hindi Belt' languages but also the neighbouring North and South Munda languages of Jharkhand, Chhattisgarh and Odisha, suggesting that areal pressure is the motivating factor behind the inclusion of these latter languages in this cluster. In fact, as Munda languages are found in three of the four clusters identified by both algorithms, this suggests that perhaps all Munda languages have been heavily influenced by their neighbours with respect to negative marking and do not show any family bias (cf. also Jenny, Weber, Weymuth 2015), in line with the findings in Borin et al. (2021) for this family with respect to other features.

However, not all languages fit neatly into one of these categories with respect to negative marking. To begin with, we find zero-nega-

tive marking in Kannada (and elsewhere in South Dravidian), a highly archaic – and crosslinguistically unique – form of negative marking. We also find paradigmatic displacement to differing degrees in three eastern languages, most notably in Gutob (South Munda), but also in Juang (South Munda) and Bengali (Indo-Aryan), although at present this restriction to eastern India appears to be coincidental. To these we can also add the use of negative particles in Nihali; despite its membership in Cluster 3, this appears to be a chance similarity, as Nihali is an isolate and is geographically quite distant from most other languages of this Cluster, so that neither family bias nor areal pressure seems likely at present.

There is still much to discover with respect to negation in South Asian languages and the present study can only be seen as a first step towards an exhaustive typology of these languages in this respect. With only 39 languages, and currently still without any languages of the Trans-Himalayan and Tai-Kadai families or the island languages such as Sinhala, Dhivehi etc., our database is still quite small. There are thus still likely many types of negative-marking strategies which we have not yet found. In addition, as more languages are included in the sample, we anticipate that further genealogical and areal tendencies will also become clearer.

Nevertheless, despite its size our database has already highlighted numerous examples of both genealogical and areal tendencies, as well as a number of 'linguistic loners' with respect to negation. Since both new languages and new features can easily be added to the database, this provides a solid base for future work on all aspects of negation for the languages of this region.

Acknowledgements

The authors wish to thank Søren Wichmann for his valuable comments on an earlier version of this article, Govind Mopkar for some last-minute questions on Konkani, as well as many useful suggestions from two anonymous reviewers. It of course goes without saying that we alone are responsible for all errors and misconceptions which the article may still contain.

Abbreviations

1, 2, 3	person
ACT	active
AUX	auxiliary
COP	copula
CVB	(sequential) converb
F	feminine
FUT	future
HAB	habitual
HUM	human
IMP	imperative
IMPF	imperfective
INAN	inanimate
IND	indicative
INF	infinitive
IPFV	imperfective
IRR	irrealis
LNK	linker
M	masculine
MID	middle
N	neuter
NAR	narrative
NEG	negative
NF	non-feminine
NH	non-honorific
NMLZR	nominaliser
NHUM	non-human
NPST	non-past
OPT	optative
PERF	perfect
PNG	person/number/gender
PRS	present
PST	past
PTCP	participle
Q	interrogative
SG	singular
SUBJ	subjunctive
TAM	tense/aspect/mood

Bibliography

General Works

Abbi, A. (1997). "Languages in Contact in Jharkhand". Abbi, A. (ed.), *Languages of tribal and Indigenous Peoples of India. The Ethnic Space*. Delhi: Motilal Banarsidass, 131-48.

Almeida, M.S.J.(1989). *A Description of Konkani*. Panaji; Goa: Thomas Stephens Konknni Kendr.

Bentz, C.; Winter, B. (2013). "Languages with More Second Language Learners Tend to Lose Nominal Case". *Language Dynamics and Change*, 3(1), 1-27. https://doi.org/10.1163/22105832-13030105.

Bhatia, T.K. (1995). *Negation in South Asian Languages*. Patiala: Indian Institute of Language Studies.

Bloch, J. (1934). *L'Indo-aryen. Du Veda aux temps modernes*. Paris: Librairie d'Amérique et d'Orient.

Bloch, J. (1935). "La forme négative du verbe dravidien". *Bulletin de la Société de Linguistique*, 36, 155-62.

Borin, L. et al. (2021). "A Bird's-eye View on South Asian Languages through LSI. Areal or Genetic Relationships?". *Journal of South Asian Languages and Linguistics*, 7(2), 203-37. https://doi.org/10.1515/jsall-2021-2034.

Bryant, D.; Moulton, V. (2004). "Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks". *Molecular Biology and Evolution*, 21(2), 255-65.

Campbell, L. (2017). "Why Is It So Hard to Define a Linguistic Area". Hickey, R. (ed.), *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press, 19-39. https://doi.org/10.1017/9781107279872.003.

Cardona, G.; Jain, D. (eds) (2003). *The Indo-Aryan Languages*. London; New York: Routledge.

Dryer, M.S.; Haspelmath, M. (eds) (2013). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info.

Eberhard, D.M.; Simons, G.F.; Fennig, C.D. (eds) (2021). *Ethnologue. Languages of the World*. 24th ed. Dallas (TX): SIL International. http://www.ethnologue.com.

Ebert, K.H. (1993). "Kiranti Subordination in the South Asian Areal Context". Ebert, K.H. (ed.), *Studies in clause linkage: Papers from the First Köln-Zürich Workshop*. Zurich: Universität Zürich, 83-110.

Ebert, K.H. (1999). "Nonfinite Verbs in Kiranti Languages – An Areal Perspective". Yogendra, P.Y.; Warren, W.G. (eds), *Topics in Nepalese Linguistics*. Kathmandu: Royal Nepal Academy, 371-400.

Ebert, K. (2006). "South Asia as a Linguistic Area". Brown, K. (ed.), *Encyclopedia of Languages and Linguistics*. 2nd ed. Oxford: Elsevier, 557-64.

Emeneau, M.B (1956). "India as a Linguistic Area". *Language*, 32(1), 3-16.

Hammarström, H. et al. (2021). *Glottolog 4.4*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.4761960.

Holman, E.W. et al. (2011). "Correlates of Reticulation in Linguistic Phylogenies". *Language Dynamics and Change*, 1(2), 205-40. https://doi.org/10.1163/221058212x648072.

Huson, D.H.; Bryant, D. (2006). "Application of Phylogenetic Networks in Evolutionary Studies". *Molecular Biology and Evolution*, 23(2), 254-67. https://doi.org/10.1093/molbev/msj030.

Ivani, J.K.; Paudyal, N.; Peterson, J. (2021). "A House Divided? Evidence for the East-West Indo-Aryan Divide and Its Significance for the Study of Northern South Asia". *Journal of South Asian Languages and Linguistics*, 7(2), 235-74. https://doi.org/10.1515/jsall-2021-2029.

Jenny, M.; Weber, T.; Weymuth, R. (2015). "The Austroasiatic Languages. A Typological Overview". Jenny, M.; Sidwell, P. (eds), *The Handbook of Austroasiatic Languages*, vol. 1. Leiden; Boston: Brill, 13-143. Grammars and Language Sketches of the World's Languages. Mainland and Insular South East Asia.

Masica, C.P. (1976). *Defining a Linguistic Area. South Asia*. Chicago; London: University of Chicago Press.

Master, A. (1946). "The Zero Negative in Dravidian". *Transactions of the Philological Society*, 45(1), 137-55. https://doi.org/10.1111/j.1467-968X.1946.tb00226.x.

Miestamo, M. (2005). *Standard Negation. The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. Berlin; New York: Mouton de Gruyter.

Miestamo, M. (2010). "Negatives without Negators". Wohlgemuth, J.; Cysouw, M. (eds), *Rethinking Universals*. New York: De Gruyter Mouton. Empirical Approaches to Language Typology EALT 45. https://doi.org/10.1515/9783110220933.169.

Miestamo, M. (2013a). "Symmetric and Asymmetric Standard Negation". Dryer, Haspelmath 2013. http://wals.info/chapter/113.

Miestamo, M. (2013b). "Subtypes of Asymmetric Standard Negation". Dryer, Haspelmath 2013. http://wals.info/chapter/114.

Miranda, R.V. (2003). "Konkani". Cardona, Jain 2003, 729-65.

Moroz, G. (2017). *Lingtypology. Easy Mapping for Linguistic Typology*. https://CRAN.R-project.org/package=lingtypology.

Nadkarni, M.V. (1975). "Bilingualism and Syntactic Change in Konkani". *Language*, 51, 672-83.

Nichols, J. (1992). *Linguistic Diversity in Space and Time*. Chicago: Chicago University Press.

Nichols, J. (1997). "Modeling Ancient Population Structures and Movement in Linguistics". *Annual Review of Anthropology*, 26(1), 359-84. https://doi.org/10.1146/annurev.anthro.26.1.359.

Osada, T. (1991). "Linguistic Convergence in the Chotanagpur Area". Mullick, S.B. (ed.), *Cultural Chotanagpur: Unity in Diversity*. New Delhi: Uppal Publishing House, 99-119.

Pederson, E. (1993). "Zero Negation in South Dravidian". Dobrin, L.M.; Nichols, L.; Rodriguez, R.M. (eds), *CLS 27*. Part 2, *The Parasession on Negation*. Chicago: Chicago Linguistic Society, 233-45.

Peterson, J. (2010). "Language Contact in Jharkhand. Linguistic Convergence Between Munda and Indo-aryan in Eastern Central India". *Himalayan Linguistics*, 9(2), 56-86.

Peterson, J. (2011). *A Grammar of Kharia. A South Munda Language*. Leiden; Boston: Brill. Brill's Studies in South and Southwest Asian Languages 1.

Peterson, J. (2015). "Introduction – Advances in the Study of Munda languages". *Journal of South Asian Languages and Linguistics*, 2(2), 149-62.

Peterson, J. (2017). "Fitting the Pieces Together. Towards a Linguistic Prehistory of Eastern-Central South Asia (and Beyond)". *Journal of South Asian Languages and Linguistics*, 4(2), 211-57.

Peterson, J. (2022). "A Sociolinguistic-Typological Approach to the Linguistic Prehistory of South Asia. Two Case Studies". *Language Dynamics and Change*, 1-50. https://doi.org/10.1163/22105832-bja10018.

Pilot, Ch. (1997). "Le 'zéro négatif' dans les conjugaison dravidiennes", in "Les langages d'Asie du Sud", monogr. no., *Faits de langues*, 10, 77-102.

Pilot-Raichoor, Ch. (2011). "The Dravidian Zero Negative. Diachronic Context of Its Morphogenesis and Conceptualisation". Wohlgemuth, J.; Cysouw, M. (eds), *Rara and Rarissima. Documenting the Fringes of Linguistic Diversity*. Berlin; New York: De Gruyter Mouton. Empirical Approaches to Language Typology (EALT) 46. https://doi.org/10.1515/9783110228557.267.

Saxena, A. (ed.) (2015). "Micro-Linguistic Areas in South Asia", special issue, *Journal of South Asian Languages and Linguistics*, 2(1).

Sinnemäki, K. (2009). "Complexity on Core Argument Marking and Population Size". Sampson, G.; Gil, D.; Trudgill, P. (eds), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press. Studies in the Evolution of Language.

Sinnemäki, K.; Di Garbo, F. (2018). "Language Structures May Adapt to the Sociolinguistic Environment, but It Matters What and How You Count. A Typological Study of Verbal and Nominal Complexity". *Frontiers in Psychology*, 9, 1-22. https://doi.org/10.3389/fpsyg.2018.01141.

Sokal, R.R.; Michener, C.D. (1958). "A Statistical Method for Evaluating Systematic Relationships". *University of Kansas Science Bulletin*, 38, 1409-38.

Southworth, F.C. (2005). *Linguistic Archaeology of South Asia*. London; New York: Routledge.

Sridhar, S.N. (1990). *Kannada*. London: Routledge. Descriptive Grammars.

Trudgill, P. (2011). *Sociolinguistic Typology. Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

Whitney, W.D. (1889). *A Sanskrit Grammar, Including Both the Classical Language, and the Older Dialects, of Veda and Brahmana*. Leipzig; London: Breitkopf & Härtel; Trübner. Bibliothek indogermanischer Grammatiken 2.

Voß, J. (forthcoming). "Negation in Gutob". Veselinova, L.; Miestamo, M. (eds), *Negation in the Languages of the World: Typologically and Functionally Based Descriptions*. Berlin: Language Science Press.

Consulted grammatical works (excluding those cited in section I of this bibliography)

Almeida, M.S.J. (2004). *Koṃkaṇī mulāvo kors* कोंकणी मुळावो कोर्स *Konkani Basic Course*. Language Consultant: Maya Kharangate. 2nd revised ed. Alto Porvorim; Goa: Thomas Stephens Konknni Kendr.

Anderson, G.D.S. (ed.) (2008). *The Munda Languages*. London; New York: Routledge. Routledge Language Family Series.

Asher, R.E.; Kumari, T.C. (1997). *Malayalam*. London; New York: Routledge. Descriptive Grammars.

Deeney, J.J. (s.d.). *Ho Grammar and Vocabulary*. Chaibasa: Xavier Ho Publications, St. Xavier's High School.

Dhongde, R.V.; Wali, K. (2009). *Marathi*. Amsterdam: Benjamins. London Oriental and African Language Library 13.

Elfenbein, J. (1998). "Brahui". Steever, S.B. (ed.), *The Dravidian Languages*. London; New York: Routledge. Routledge Language Family Descriptions, 388-414.

Fernandez, F. (1967). *A Grammatical Sketch of Remo. A Munda Language* [PhD dissertation]. Chapel Hill (NC): University of North Carolina.

Ghosh, A. (2008). "Santali". Anderson 2008, 11-98.

Griffiths, A. (2008). "Gutob". Anderson 2008, 633-81.

Hoffmann, J. (1903). *Mundari Grammar*. Calcutta: Bengal Secretariat Press.

Israel, M. (1979). *A Grammar of the Kuvi Language (with Texts and Vocabulary)*. Vanchiyoor; Trivandrum: Dravidian Linguistics Association.

Kobayashi, M. (2012). *Texts and Grammar of Malto*. Vizianagaram: Kotoba Systems. Text and Analysis of Indian Languages Series 1.

Kobayashi, M.; Tirkey. B. (2017). *The Kurux Language. Grammar, Texts and Lexicon*. Leiden; Boston: Brill. Brill's Studies in South and Southwest Asian Languages 8.

Krishnamurti, B.; Gwynn, J.P.L. (1985). *A Grammar of Modern Telugu*. Delhi; Bombay; Calcutta; Madras: Oxford University Press.

Malhotra, V. (1982). *The Structure of Kharia. A Study of Linguistic Typology and Language Change* [PhD dissertation]. New Delhi: Jawaharlal Nehru University.

Matson, D.M. (1964). *A Grammatical Sketch of Juang* [PhD Dissertation]. Madison (WI): University of Wisconsin.

Matthews, D. (1998). *A Course in Nepali*. London; New York: Routledge.

Neukom, L. (2001). *Santali*. München: Lincom Europa. Languages of the World, Materials 323.

Neukom, L.; Patnaik, M. (2003). *A Grammar of Oriya*. Zurich: Universität Zürich.

Osada, T. (1992). *A Reference Grammar of Mundari*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Osada, T. (2008). "Mundari". Anderson 2008, 99-164.

Patnaik, M. (2008). "Juang". Anderson 2008, 508-556.

Radice, W. (1994). *Bengali. A Complete Course for Beginners*. London; Lincolnwood (IL): Hodder & Stoughton. Teach Yourself Books.

Raeside, I.M.P.; Nemade, B.V. (1991). *Marathi Reading Course*. London: School of Oriental and African Studies, University of London.

Schiffman, H.F. (1983). *A Reference Grammar of Spoken Kannada*. Seattle (WA): University of Washington Press.

Schulze, F.V.P. (1911). *A Grammar of the Kuvi Language, with Copious Examples*. Madras: Graves, Cookson & Co.

Shukla, S. (1981). *Bhojpuri Grammar*. Washington, D.C.: Georgetown University Press.
Snell, R. (1991). *The Hindi Classical Tradition. A Braj Bhāṣā Reader*. London: University of London School of Oriental and African Studies. SOAS South Asian Texts 2.
Thompson, H.-R. (2010). *Bengali*. Amsterdam; Philadelphia: Benjamins. London Oriental and African Language Library 18.
Tiwari, U.N. (1960). *The Origin and Development of Bhojpuri*. Calcutta: The Asiatic Society.
Verma, M.K. (2003). "Bhojpuri". Cardona, Jain 2003, 515-37.
Winfield, W.W. (1928). *A Grammar of the Kui Language*. Calcutta: The Asiatic Society of Bengal. Bibliotheca Indica 245.
Yadav, R. (1996). *A Reference Grammar of Maithili*. Berlin: Mouton de Gruyter. Trends in Linguistics, Documentation 11.
Zydenbos, R.J. (2020). *A Manual of Modern Kannada*. CrossAsia-eBooks. Heidelberg University Library. https://doi.org/10.11588/xabooks.736.

Appendix A

Languages in the sample (for the literature consulted, see Part II of the references)

Glottocode	Language name
Indo-Aryan	
awad1243	Awadhi
beng1280	Bengali
bhoj1244	Bhojpuri
bund1253	Bundeli
dara1250	Darai
goan1235	Goan Konkani
guja1252	Gujarati
hind1269	Hindi
kash1277	Kashmiri
mait1250	Maithili
mara1378	Marathi
marw1260	Marwari
nepa1254	Nepali
oriy1255	Odiya
sadr1248	Sadri
Dravidian	
brah1256	Brahui
dand1238	Dandami Maria
pott1240	Gadaba
nort2699	Northwest Kolami
sout1549	Southeast Kolami
kuii1252	Kui
kuru1302	Kurukh
kuvi1243	Kuvi
nucl1305	Kannada
mala1464	Malayalam
saur1249	Malto
telu1262	Telugu
Munda	
bond1245	Bondo / Remo
gata1239	Gtaʔ
hooo1248	Ho
juan1238	Juang
khar1287	Kharia
kork1243	Korku
mund1320	Mundari

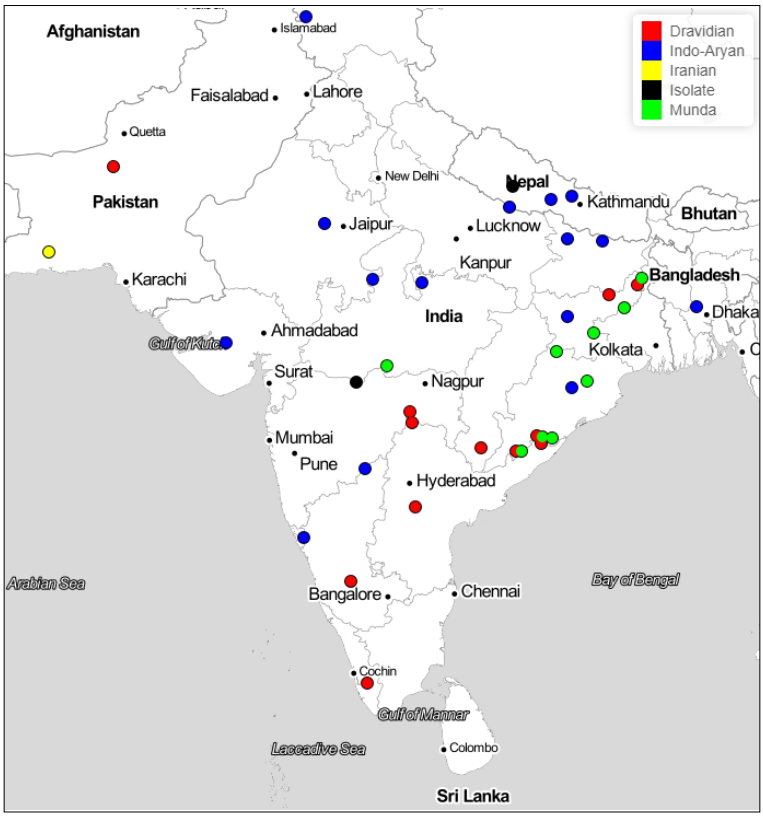
sant1410	Santali
sora1254	Sora

Iranian	
sout2642	Balochi
Isolates	
niha1238	Nihali
kusu1250	Kusunda

Appendix B

Map of languages, mapped with the help of *lingtypology* (Moroz 2017)

Languages in the sample



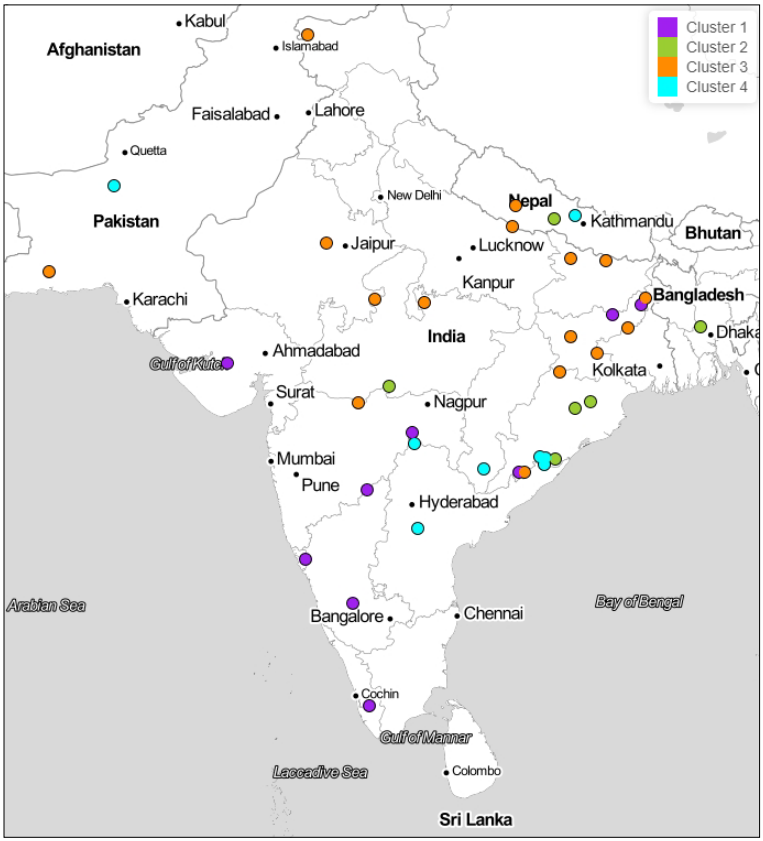
Appendix C

Features documented in the database

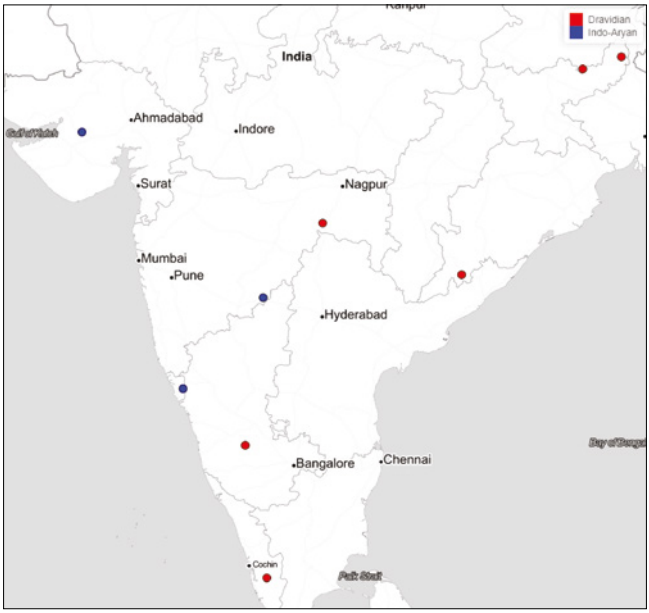
GB107	Can standard negation be marked by an affix, clitic or modification of the verb?
GB298	Can standard negation be marked by an inflecting word (“auxiliary verb”)?
GB299	Can standard negation be marked by a non-inflecting word (“auxiliary particle”)?
SA071	Can standard negation be marked by a prefix/proclitic?
SA072	Can standard negation be marked by a suffix/enclitic?
SA073	Can standard negation be marked by a word-final suffix/enclitic?
SA074	Can standard negation be marked by a non-final suffix?
SA075	Can standard negation be marked by an inflecting word homophonous with or deriving from the copula?
SA076	Can copula verbs be negated though suppletion?
SA077	Can standard negation be marked through suppletion with non-copular verbs?
SA078	Can standard negation be marked by an inflecting word together with a finite predicate?
SA078a	Is this an asymmetric negation strategy?
SA079	Can standard negation be marked by an inflecting word together with a participle?
SA079a	Is this an asymmetric negation strategy?
SA080	Can standard negation be marked by an inflecting word together with an infinitive?
SA080a	Is this an asymmetric negation strategy?
SA081	Can standard negation be marked by an inflecting word together with a type of verb form other than those in SA078-SA080?
SA081a	Is this an asymmetric negation strategy?
SA082	Are there different negation strategies based on any TAM categories?
SA083	Are there different negation strategies based on tense/aspect categories?
SA083a	Is this an asymmetric negation strategy?
SA084	Are there different negation strategies based on mood categories?
SA084a	Is this an asymmetric negation strategy?
SA085	Are there markers for TAM which have the same form but different values in standard negation than in non-negation?
SA086	Is there any asymmetric negation in this language?

Appendix D

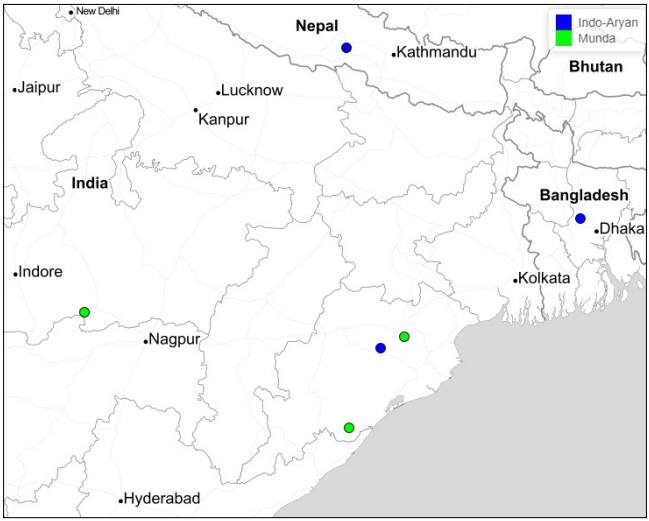
Language clusters suggested by the two algorithms, mapped with the help of *lingtypology* (Moroz 2017)



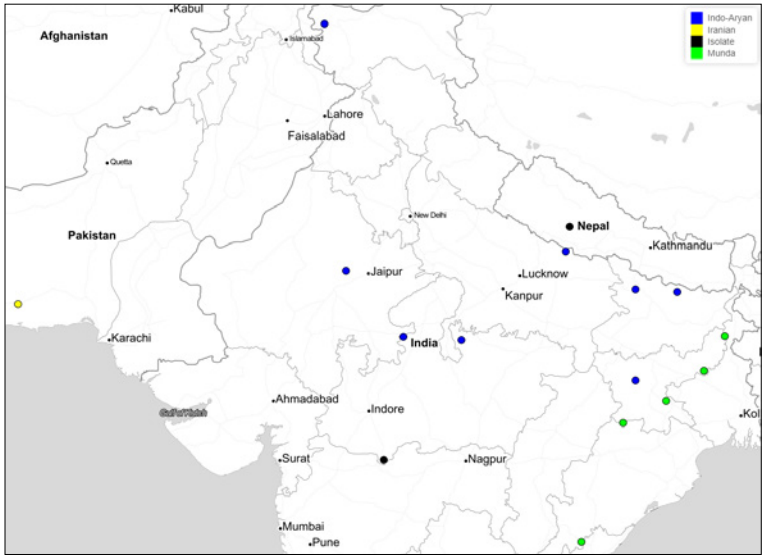
Cluster 1



Cluster 2



Cluster 3



Cluster 4