e-ISSN 2785-5953

Bhāṣā

Journal of South Asian Linguistics, Philology and Grammatical Traditions

Vol. 1 – Num. 1
April 2022

Edizioni Ca'Foscari

e-ISSN 2785-5953

Bhasha

Journal of South Asian Linguistics,
Philology and Grammatical Traditions

Editor-in-chief
Andrea Drocco

Edizioni Ca' Foscari - Venice University Press
Fondazione Università Ca' Foscari
Dorsoduro 3246, 30123 Venezia
URL https://edizionicafoscari.unive.it/it/edizioni/riviste/bhasha/

Bhasha
Journal of South Asian Linguistics, Philology and Grammatical Traditions

Semestral journal

Head office Ca' Foscari University of Venice | Department of Asian and North African Studies | Dorsoduro 3462, 30123 Venice, Italy | bhasha_journal@unive.it

Publisher Edizioni Ca' Foscari | Fondazione Università Ca' Foscari | Dorsoduro 3246, 30123 Venice, Italy | ecf@unive.it

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della rivista. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.
Scientific certification of the works published by Edizioni Ca' Foscari: all essays published in this issue have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Advisory Board of the journal. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

Bhasha
Vol. 1 – Num. 1 – April 2022
e-ISSN 2785-5953

Table of Contents

URL https://edizionicafoscari.unive.it/it/edizioni/riviste/bhasha/2021/1/
DOI http://doi.org/10.30687/bhasha/8409-3769/2021/01

e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

Welcome Note

E. Annamalai
Director (retired), Central Institute of Indian Languages, Mysore
Professor of Tamil (retired), The University of Chicago, Chicago

The languages of South Asia are studied by scholars from various disciplinary points of view and with various disciplinary tools. These languages have a long history and have troves of written texts of many centuries, whose study contributes to historical linguistics and philological studies. The major languages of South Asia have a long history of grammatical descriptions and philosophical investigations from the ancient times. The richness of the South Asian languages in all the above aspects is unparalleled in any region of the world.

The current journals devoted to promote scholarship on South Asian languages, however, are fragmented, being focused on particular specialisations or periods. The new journal *Bhāṣā. Journal of South Asian Linguistics, Philology and Grammatical Traditions* is an appreciated alternate in this scenario. It is particularly welcome for the scholars of languages working in South Asia who generally combine the knowledge of modern grammatical theories and traditional grammars, and of language descriptions and language histories. This new journal provides them with an avenue to break the existing boundaries between linguistics, philology and grammatical traditions.

I welcome such scholars to contribute their research work to the Journal and to build bridges between various aspects of their language of study through their publication.

e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

The Range of *Bhāṣā*

Hans Henrich Hock
Professor Emeritus of Linguistics and Sanskrit at the University of Illinois, Urbana-Champaign, USA

It is a great pleasure to welcome the first issue of a new journal dedicated to the study of South Asian languages in the broadest sense possible.

During the 20th century the earlier combination of textual studies (philology), investigation of indigenous grammatical traditions, and what may be called 'pure linguistics' has increasingly become severed. True, practicing historical-comparative linguists continue to do a fair amount of philological work on the texts from which they draw their data, and many of them go beyond, in terms of detailed textual studies. Some also still study the ancient grammarians, either in their own right or in terms of what insights they may offer for general linguistic studies. But academic institutions and journals tend to separate linguistics from philology, and the study of ancient grammatical traditions tends to fall between the chairs.

It is gratifying, therefore, that *Bhāṣā* भाषा aims to bring philology, linguistics, and the study of Indian grammatical traditions together, under one roof. The fact that it does so in an open-access format adds further reasons for appreciation and gratitude.

In fact, the contributions to this first issue already manifest much of the intended breadth of the Journal, ranging from 'purely' linguistic investigations on South Asian negation and predicative possessives in Hindi, to a study of sarcasm in Kashmiri, to the ideology underlying the Sanskrit revival movement and speakers' declarations in the Indian Census, to – last, but by no means least – a 'purely' philological investigation on the Pāli *Milindapañha* and its Chinese counterparts.

It is to be hoped that future issues continue in this broad direction and further increase the breadth by addressing issues in all of the languages and language families of South Asia, whether linguistic, philological, or focused on the different grammatical traditions.

The Editor-in-Chief, Professor Andrea Drocco and his home institution, the Department of Asian and North African Studies of Ca' Foscari University of Venice, deserve our deep gratitude for having embanked on this ambitious voyage.

वाक्शास्त्रे शब्दशास्त्रे च भाषाशास्त्रे तथैव च ।
पत्रिकेयं प्रकाशित्री जीवेत शरद: शतम् ॥

e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

Welcome *Bhāṣā*

Silvia Luraghi
Università degli Studi di Pavia, Italia

It is always exciting to see a new outlet for linguistic publications opening up opportunities for scholarly exchange, especially one which is explicitly engaged in the mission of bringing together linguistic, philological and grammatical studies. Cross-disciplinary research is of paramount importance in a field such as the one that *Bhāṣā* aims to cover, the languages of South Asia, in which linguistic, philological and grammatical studies can all count on a long and rich tradition, with a visible, and regrettable, lack of communication that has remained quite steady over time. The result of mutual ignorance is often a duplication of efforts by scholars that work in neighbouring fields, and could profit from the advancements of colleagues working on the same topic but within different traditions and with different means of investigation. Fostering communication, on the other hand, widens the opportunity to build on results that have already been achieved, resulting in a growth that ultimately promotes research irrespective of the separate tradition to which individual scholars belong. Building bridges is not an easy task when it is confronted with well established habits, which sometimes bring philologists to be suspicious of linguists or the other way around. In this respect, the first issue of *Bhāṣā* is up to its commitment, as it comprises a wide range of studies covering both philological and linguistic research, as witnessed by the papers by Bryan De Notariis on the reconstruction of the archetype of a Buddhist manuscript on the one hand, and by Lucrezia Carnesale concerning Hindi possessive constructions on the other hand. In addition, Patrick McCartney addresses a sociolinguistic issue connected with the speakers attitudes as witnessed by the 2001 and the 2011 censuses. Finally, two other es-

sentially linguistic contribution, one by John Peterson and Lennart Chevallier and the other by Peter Hook and Omkar Koul, tackle their topics from the point of view of typological comparison (negation in South Asian languages) and of areal linguistics (sarcasm in Kashmiri as an areal feature), covering a large spectrum of languages from different language families that reflects the linguistic wealth of the area. For these reasons, professor Andrea Drocco, the founder and editor of *Bhāṣā*, deserves the gratitude of all scholars interested in various ways in the languages of South Asia.

e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

Editorial

Andrea Drocco
Università Ca' Foscari Venezia, Italia

It is widely recognised that Sanskrit and more generally the study of the languages of South Asia played an extremely important role in the development of historical and comparative linguistics during the nineteenth century. Besides, it has been witnessed the same regarding the study of Indian grammatical tradition, especially Sanskrit grammatical tradition, which sees Pāṇini as its greatest exponent and representative. However, until the end of the first decade of twenty-first century, no single journal in Western countries has been devoted to the linguistics of these languages. This is a relevant fact, considering the importance of old Indo-Aryan and Dravidian languages for the study of diachronic linguistics, grammatical tradition and/or textual criticism – thanks also to the multitude of texts representing the various phases of linguistic evolution of these languages – and considering the high linguistic diversity of South Asia and, consequently, its importance for the study of multilingualism, language contact and language policy and planning. With the exception of *Indian Linguistics*, the quarterly journal (produced annually) of the Linguistic Society of India, only recently, in fact, specific academic journals started being published whose focus is the linguistic analysis of South Asian languages. This is the case of the *Journal of South Asian Linguistics* which is an online and open access publication edited by Sameerud ud Dowla Khan and Emily Manetta (initially by Miriam Butt and Rajesh Bhatt) that, from 2008 onward, publishes original research articles and book reviews. John Benjamins Publishing Company (Amsterdam and Philadelphia) publishes the *Journal of South Asian Languages and Linguistics* the first volume of which came out in 2014. The focus of the journal is on descriptive, functional and typological investigations, but descriptive studies are also en-

couraged to the extent that they present analyses of lesser-known languages, based on original fieldwork. The editor-in-chief is Leonid Kulikov (initially Anju Saxena).

It is important to note that the scope of these journals is not a single group/family of South Asian languages, but, on the contrary, all languages used, today and in the past, in the region and, moreover, the South Asian linguistic landscape as a whole. Certainly, this is the result of a changing approach in studying these languages. As a matter of fact, according to this relatively new approach, in order to have a correct view not only on the characteristics of the contemporary languages of this region, but also of what their evolution must have been, even recently, it is necessary to consider language contact as an essential factor of language change in the Indian sub-continent. As Hock (2016, 1) has pointed out, this change of perspective has been a consequence of the fact that this region started to be studied, from the point of view of linguistics, as a linguistic area. Emeneau's papers (1980) on this topic were fundamental in this respect (especially Emeneau 1956), followed in importance by Masica's study of 1976. Thence, over the past five-six decades there has been a growing interest in the study of South Asian languages from different perspectives and adopting various methodological approaches. More recently, this interest has also expanded to the field of endangered and unwritten languages to the extent that it has now become an established field of research for various academic projects around the World (cf., for example, the *Himalayan Languages Project*, as well as the various studies, grammars and/or documentation works published in the context of the *Hauns Rausing Endangered Languages Project* and *Living Tongues Institute for Endangered Languages*).

Notwithstanding this ferment in the discipline with the increasingly presence of a new generation of established scholars, in some countries the impact of linguistic studies on South Asian languages has not been as relevant as it has been in other countries. In this regard, Italy is a case in point. As a matter of fact, after the pioneering work of some eminent scholars dealing with Middle Indo-Aryan or New-Indo-Aryan languages (the emblematic example is represented by Luigi Pio Tessitori, in particular by his well-known *Grammar of the Old Western Râjasthânî With Special Reference to…* 1914-16), the few Italian linguists who studied some of the various languages of South-Asia have devoted their attention to the analysis of Vedic, especially from an Indo-European perspective, or Sanskrit grammatical tradition. Unfortunately, with rare and recent exceptions, most of these few studies are written in Italian language, thus difficult to access for foreign scholars.

It is precisely in this context that *Bhāṣā. Journal of South Asian Linguistics, Philology and Grammatical Traditions* has been launched. Starting from an idea of Antonio Rigopoulos and mine and patronised

by the Department of Asian and North African Studies of Ca' Foscari University of Venice, this new peer-reviewed, international research journal on the languages of the Indian subcontinent welcomes submissions adopting evidence-based approaches to all areas of linguistics related to South Asian literary (classical and modern/contemporary), spoken and/or endangered languages. The Journal is published online by Edizioni Ca' Foscari, which was created in 2011 with the aim of encouraging dissemination of research results within the University and from here to both the national and international scientific community. For this reason, all publications are made available online with free and open access in order to bolster and encourage the free sharing of knowledge.

The main purpose of the Journal is twofold.

On the one hand, to collect papers devoted to general synchronic linguistic analysis (including sociolinguistic analysis) of a particular language (even as testified in a specific text) and even those incorporating a comparative analysis with other languages. According to this aim, the papers submitted can have a theoretical and/or a descriptive approach, in both respects adopting a typological and functional perspective.

On the other hand, one of the major aims of *Bhāṣā* is to better understand the evolution of the various languages employed in South Asia today and in the past. The term 'evolution' is here understood from the point of view of linguistic history – according to a pure diachronic linguistic perspective – as well as from the point of view of the history of these languages, concerning thus the dynamics existing between a specific language and the culture and socio-political context of the society where this language is spoken. For this reason, the Journal also includes in its scope the analysis of the history of reading and reception studies in South Asia and, accordingly, articles focusing on textual details and criticism and on the history of manuscript traditions and circulation will also be considered. In fact, it is our firm belief that only through the study of the relationship occurring between the languages of 'texts', with their marked bias toward the educated classes, and their variants (the languages of non-[canonised] texts) that it is possible to understand the sociolinguistic relationship between a more standard, established language and a plethora of sub-standard 'languages'. Above all, we are wondering how the studies of philology and historical sociolinguistics can come to the aid of historical linguistics analysis. Therefore, the Journal is not solely devoted to the 'pure' linguistic study of South Asian languages with a synchronic or a diachronic approach. Indeed, one of the main purposes of this Journal is to transcend the old dichotomy of synchrony and diachrony and to combine philology and linguistics through some of the papers that will be submitted. Last but not the least, particular emphasis is also placed on the study of

grammatical traditions that have thrived in the South Asian regions.

A considerable part of *Bhāṣā. Journal of South Asian Linguistics, Philology and Grammatical Traditions* will be devoted to reviews of new books or specific important papers related in some way to the aims of the Journal. The Journal predominantly publishes articles in English but will occasionally also publish in Italian, French and German.

In this first issue, *Bhāṣā* offers to the academic reader five innovative papers that show the results of some of its research themes. We look forward to future article's submissions mostly dealing with research themes related to the topics of this new Journal, but also, and especially, focusing on the whole of South Asian languages and on the Indian grammatical tradition – i.e. Sanskrit, Prakrit, Pali, Tamil, etc. – as well as on the grammatical tradition of Modern Indian languages.

The paper by Peterson and Chevallier offers a very detailed (even if preliminary according to the authors) typology of negation based on a database of 25 structural features for 39 languages from three language families and two language isolates providing thus a first analysis of the distribution of these different negative-marking strategies throughout the subcontinent. In some cases of language contact, the data also allows us to determine with some certainty the type of contact situation that has led to the negative-marking patterns documented.

The paper by Hook and Koul explores a particular formulaic construction specifically dedicated to the expression of sarcasm in many Indo-Aryan languages. Even if the focus of the two authors is on a particular language, i.e. Kashmiri, their aim is to show that this particular construction is attested non only in the majority of Indo-Aryan languages, but also in some but not all of the major Dravidian languages. As a consequence, they suggest that this specific construction is another clear example of 'trait' of South Asia as a linguistic area.

The article by Patrick McCartney deals with 2001 and 2011 *Census of India* data in which L1-L3 (first to third language) Sanskrit tokens were returned during census enumeration. The main goal of the paper, that is part of the *Imagining Sanskrtland* project, is on locating and documenting how, where, and why the most important literary Indo-Aryan language, Sanskrit, is spoken in the twenty-first century. In particular, a theo-political discussion of Sanskrit's imaginative power for faith-based development is provided, in order to discuss how 'Sanskrit-speaking' villages signify an ambition toward cultural renaissance.

The article by De Notariis is conceived as an introduction to questions concerning the relationship between various versions of a Buddhist text known in its Pāli variant as *Milindapañha*, and in its Chinese versions as *Nàxiān bǐqiū jīng* (那先比丘經; T 1670 versions A and

e-ISSN 2785-5953

B). With respect to the latter, particular attention to the Western reception and the problem related to the reconstruction of a possible archetype, adopting the guidelines provided by Gérard Fussman, are provided.

Carnesale's paper deals with the semantic/pragmatic-syntactic interplay of Hindi predicative possessive constructions, especially taking into account the concept of linguistic iconicity. Therefore, the aim of the paper is to show that each possessive construction in the largest new Indo-Aryan language of modern South Asia is customised to encode specific semantic properties.

I wish to thank the members of the Editorial Board, in particular E. Annamalai, Hans Henrich Hock and Silvia Luraghi for their welcome notes, the reviewers of this first issue, the entire staff of Edizioni Ca' Foscari, and the Department of Asian and North African Studies of Ca' Foscari University of Venice.

Bibliography

Emeneau, M.B. (1956). "India as a Linguistic Area". *Language*, 32, 3-16. Reprinted in Emeneau 1980, 105-25.

Emeneau, M.B. (1980). *Language and Linguistic Area*. Essays selected by A.S. Dil. Stanford: Stanford University Press.

Hock, H.H. (2016). "Introduction". Hock, H.H.; Bashir, E. (eds), *The Languages and Linguistics of South Asia: A Comprehensive Guide*. Berlin, Boston: De Gruyter Mouton, 1-7.

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

Tessitori, L.P. (1914-16). "Notes on the Grammar of the Old Western Râjasthânî with Special Reference to Apabhraṃśa and to Gujarâtî and Mârwâṛî". *Indian Antiquary*, XLIII-XLIV, 1-106.

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 *-ɔ̃* '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 *nɑ̃* '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-ī*[*]	*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 <ī>, however vowel length is not phonemic in Nepali and Matthews (1998, 3-4) writes that there is no difference in pronunciation between <i> and <ī>.

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.

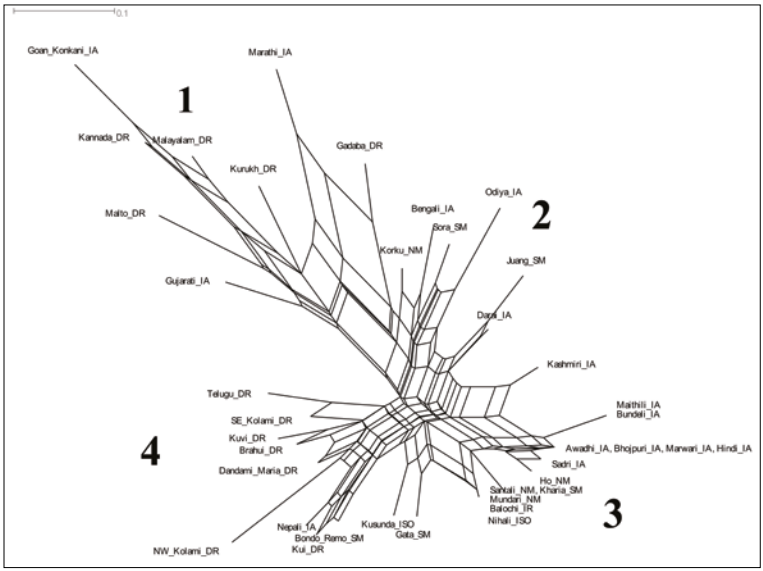


Figure 1 A NeighborNet representation of negation in South Asian languages (25 features in 39 languages)

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.

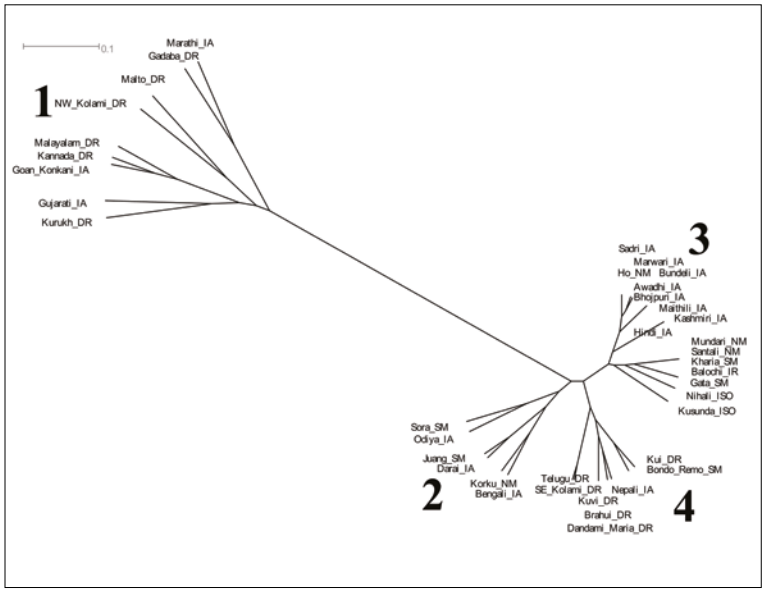


Figure 2 A UPGMA representation of negation in South Asian languages (25 features in 39 languages)

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

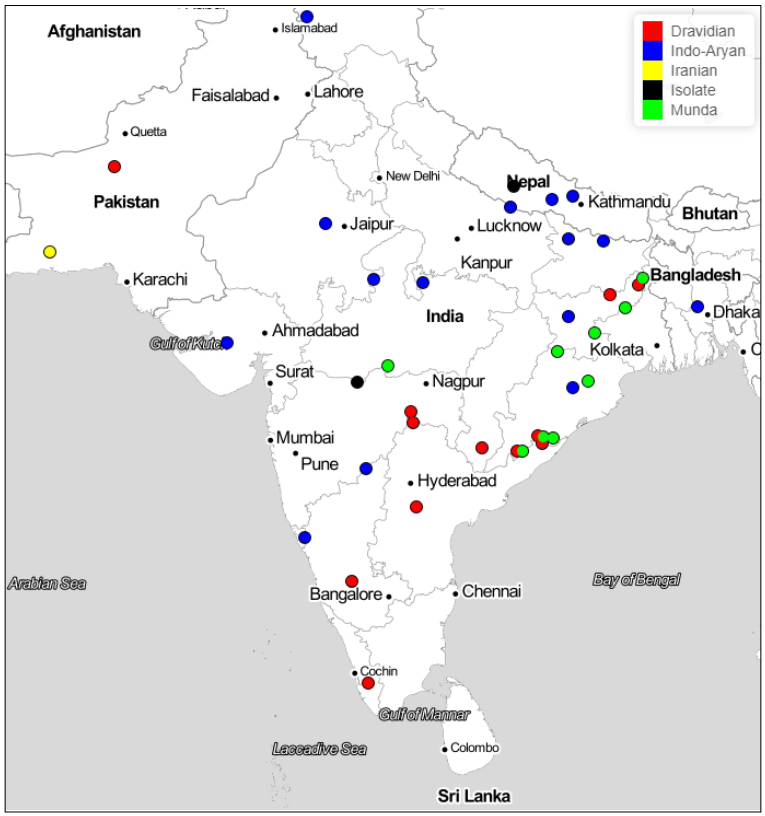
sant1410	Santali
sora1254	Sora

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



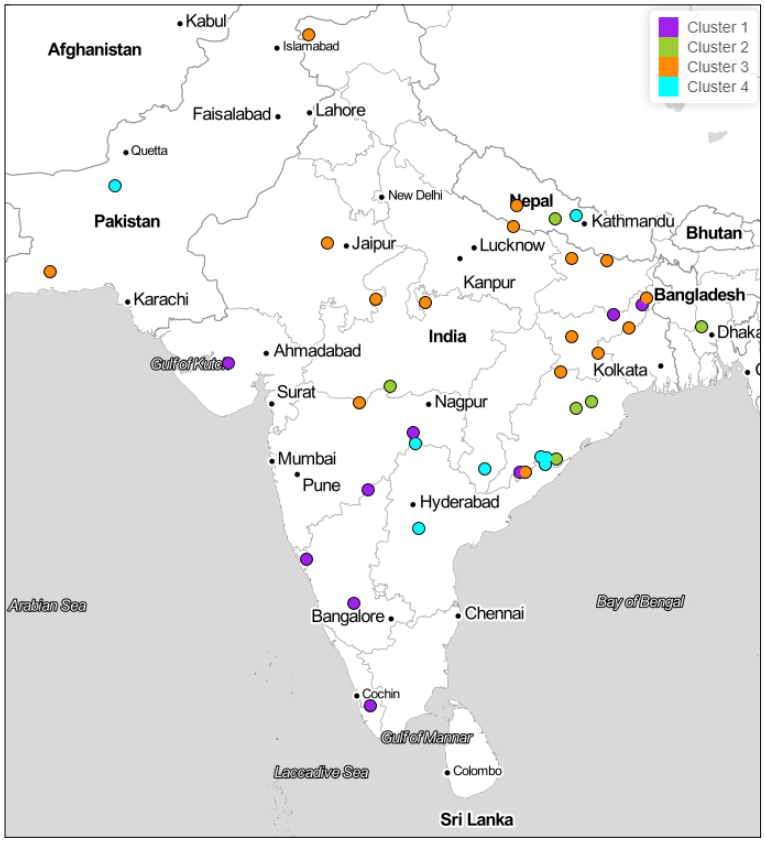
Appendix C

Features documented in the database

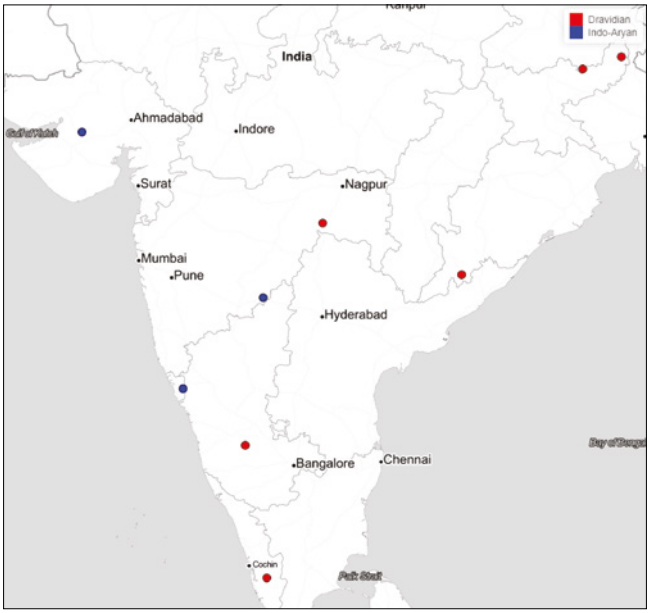
GB107	Can standard negation be marked by an affix, clitic or modification of the verb?
GB298	Can standard negation be marked by an inflecting word (“auxiliary verb”)?
GB299	Can standard negation be marked by a non-inflecting word (“auxiliary particle”)?
SA071	Can standard negation be marked by a prefix/proclitic?
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

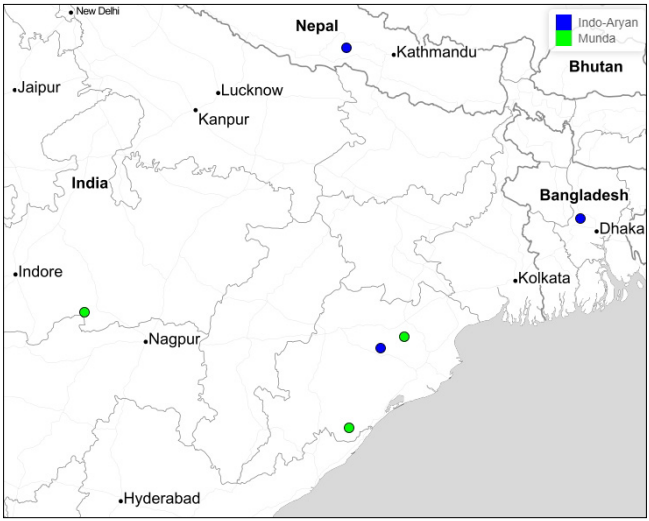
Language clusters suggested by the two algorithms, mapped with the help of *lingtypology* (Moroz 2017)



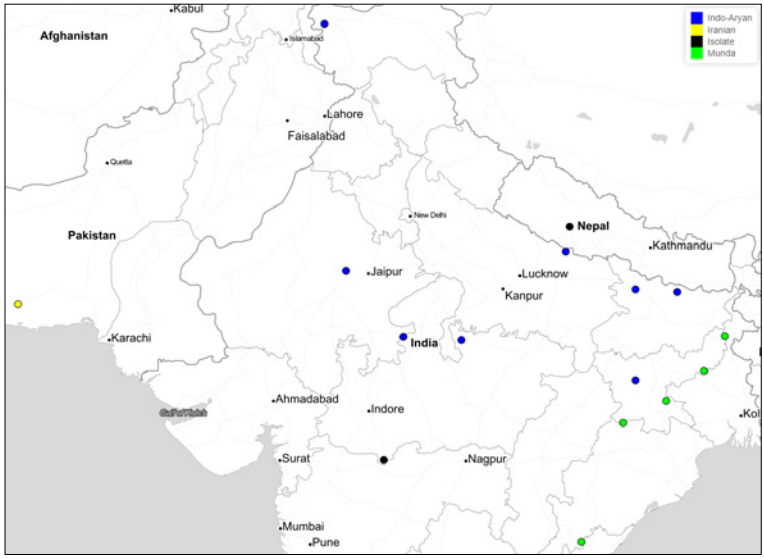
Cluster 1



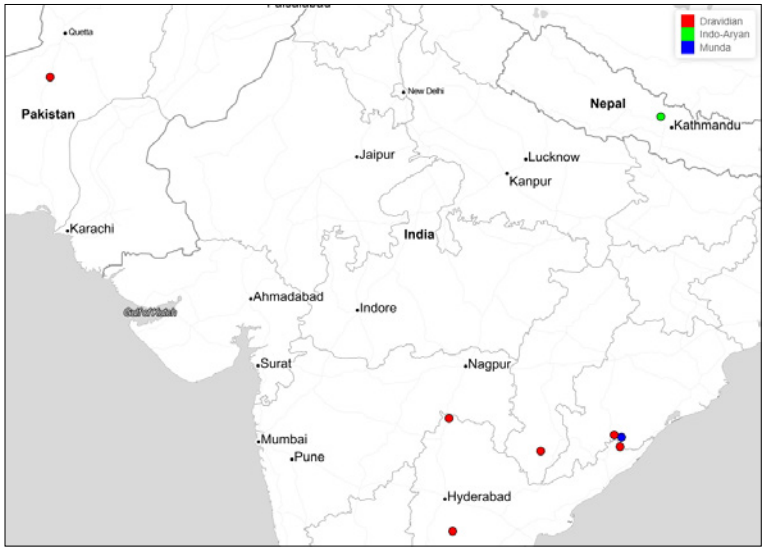
Cluster 2



Cluster 3



Cluster 4



e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

A Dedicated Sarcasm Construction in Kashmiri as a Feature of the South Asian Linguistic Area

Peter Edwin Hook
University of Michigan, USA

Omar N. Koul †
Indian Institute of Language Studies, Karnataka, India

Abstract This paper has two parts. In the first part (§§ 2-3) we cover some of the resources that the Kashmiri sarcast draws on. In the second part (§§ 4-6) we explore a particular formulaic construction that we maintain is specifically dedicated to the expression of sarcasm, not only in Kashmiri but in most languages of the Indo-Aryan family – with the possible exception of Bangla and its neighbours. Since this construction occurs in some but not all the major Dravidian languages, it may be regarded as another feature or 'trait' of the South Asian Linguistic Area first adumbrated by Emeneau (1956) and further extended by Masica (1976).

Keywords Insincere speech. Intent. Cue. Dedicated construction. Indo-Aryan. Dravidian. Linguistic area.

Edizioni Ca'Foscari

Peer review

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

Citation Hook, P.E.; Koul, O.N. (2022). "A Dedicated Sarcasm Construction in Kashmiri as a Feature of the South Asian Linguistic Area". *Bhasha*, 1(1), 63-76.

DOI 10.30687/bhasha/8409-3769/2022/01/006

1 Introduction[1]

Sarcasm, the deliberate attempt to point out, question, or ridicule attitudes and beliefs by using words or gestures in ways that run counter to their normal meaning, is probably a universal of human society. Like all ironic discourse (itself a subcategory of insincere speech) the power of sarcasm depends on the listener's being (or becoming) aware that the speaker does not mean what is being said. Often the gap between what is said and what the participants in a conversation all know to be true is sufficient to tip off listeners. Perhaps because this pragmatic kind of sarcasm occurs so commonly, it does not usually fall within the range of phenomena that a grammarian might feel obliged to account for. For instance, while the intention of (1) in the circumstances of its utterance may indeed be sarcastic, there may or may not be any audible indicator of sarcastic intent: The hearer may have to depend on the mismatch between the literal meaning of the message and what he knows to be reality (IM = implied / intended meaning):[2]

(1)	*šābāš*	*čhu-y*	*čĕĕnyis*	*gāṭijār-as*
	bravo	is-2SG.DT	your.DAT	wisdom-DAT
	'Congratulations on your wisdom'.			
	(IM: 'You're not so smart as you think!')			

But there are other kinds of sarcasm in which the speaker must indicate his intent through some behavioural or linguistic cue: a lexical item (2), a special intonational contour (3), over-articulation (4), pauses (5), inappropriate formality (6), overcareful framing (7), hy-

1 The first part of this paper is modelled on a paper co-authored by Hook and Kusum Jain entitled "How to be Sarcastic in Hindi-Urdu", drafted in India during the summer of 1997 and published in 2002 in a felicitation volume for George Cardona. There is a running comparison of the modes of sarcasm in Kashmiri and Hindi-Urdu from fn. 7 onward and a comparison of dedicated sarcasm constructions in Indo-Aryan and Dravidian languages in §§ 5 and 6.

2 The transcription used for Kashmiri in this paper is the one worked out by Kenneth Hill and Sajad Mir in Prof. Hill's course in linguistics field methods taught at the University of Michigan in Fall, 1984. Based on a system often found in the linguistics literature on contemporary Indo-Aryan languages, it was designed to minimise the use of diacritics and special symbols. The letter {e} represents a mid (either front or central) vowel while {i} represents a high (either front or central) vowel. Fronting is determined by the presence of {y} or other palatal consonants. Palatalisation is uniformly indicated with the letter {y} (except that {j}, {čh}, {č}, and {š} are inherently palatalised); {ʦ} is a dental affricate; and {ṭ}, {ṭh}, and {ḍ} are retroflex stops. In the transcription of data from Hindi-Urdu the letter {e} always represents a mid front (never a central) vowel while {i} always represents a high front (never a central) vowel. See the list of abbreviations.

perbole (8), understatement (9), mimicry (10), inversion (11), a wink, a deadpan expression etc:[3]

(2)	**Sure** he'll return your books!	(IM: 'He won't return your books')
(3)	**Great!** (with a prenuclear drop in pitch)	(IM: 'Terrible!')
(4)	**Pretty** bad! (post-tonic geminate: ['prItti])[4]	(IM: 'Extremely bad!')
(5)	**What… a… find**!	(IM: 'This is trash!')
(6)	**William Winkler**, you clean up your room!	(IM: 'Billy, I'll withdraw my normal affection, if you don't clean up your room')
(7)	What **seems** to be the problem?	(IM: 'You're not sick. You just think you are! Don't waste my time!')
(8)	He's **a genius**!	(IM: 'He's far below average')
(9)	**I wouldn't bet on it.**[5]	(IM: 'Absolutely not!')
(10)	Did **W**ick**y** sp**w**ain his **wittle w**ist?	(IM: 'Ricky is a childish complainer')
(11)	**Don't let pregnancy spoil your drug**	(The usual message is reversed with the intention of persuading drug addicts to get themselves sterilised)

Many of these phenomena are fair game for the grammarian. An account of any linguistic cue (intonational, morphological, lexical, or syntactic) that distinguishes sarcastic from non-sarcastic utterances in a regular or predictable way falls within the jurisdiction of the grammarian. However, in our discussion of Kashmiri we will be particularly interested in lexical and constructional cues that are themselves sites of sarcasm.

Mechanisms of sarcasm can be classified by target into two groups: A. Certain expressions invoke and attack beliefs of the per-

3 Haiman (1990) attempts an exhaustive taxonomy of linguistic cues to sarcastic intent. Jagannathan (1981, 338) discusses the sarcastic use among Hindi-speakers of appellations like *guru* and *xalīfā*. Cf. Taing 1984 for irony in Kashmiri literature.

4 The apical stop in sarcastic *pretty* shows an affective gemination that blocks the flapped articulation normally expected for post-tonic intervocalic /t/ (thanks to Alexis Manaster-Ramer and Bill Darden for this observation). A similar gemination with displacement to the left of the tonic (using extra high pitch) is audible in the sarcastic articulation of *okay*: ['okke].

5 Understatement is typical of 'dry' sarcasm. A more complex example: "On this date in 1492 Christopher Columbus signed a contract with the Spanish Crown to sail the ocean blue… in search of Asia. He did not find it" (National Public Radio's Morning Edition, 17 April 1998). This instance shows several cues of sarcastic intent: 1. Partial quoting of the children's rhyme "In 1492 Columbus sailed the ocean blue"; 2. the extra pause between 'blue' and the phrase 'in search of Asia'; 3. the understated 'He did not find it'; 4. a prolongation – with rising pitch – of the nucleus of 'find'. The sting of this sarcasm is dilute: its targets – Columbus and his royal backers – now all safely dead – are gently twitted for their geographic illusions

son to whom they are addressed (i.e. the hearer).[6] B. Others impugn the presumed beliefs of some participant in the situation denoted by the utterance. When the speaker himself or herself is the target, the sarcastic intent of the utterance is usually directed to his or her previous beliefs and the sting is lightened to what may be characterised as the ironic expression of regret.

2 Hearer-Oriented Sarcasm

Among the cues Kashmiri speakers use to signal that beliefs of the hearer are under attack is the use of the simple past tense (aka preterite)[7] to denote actions that the hearer knows very well have yet to occur:

(12) *temy* *dyity-iy* *ti* *ʦye* *reṭy-th-as!*
he.ERG gave-M.PL-2SG.DT and you.ERG took.M.PL.2SG.ER-3SG.DT
'He gave (them) to you and you got (them) from him!' ('them' refers to money)
(IM: 'He will not give you (the money) and so of course you will not get it from him!')

This use of the past tense for future actions provides a cue to sarcastic intent that can be reinforced with *hay* 'indeed' or its abbreviated affixal form *-ay*:

6 Jagannathan (1981, 337) draws a further distinction between 'sharp' sarcasm (*tīkhā vyangy*) in which the addressee is supposed to recognise the speaker's intent and 'subtle' sarcasm (*sūkṣm vyangy*) in which only hearers other than the addressee are supposed to realise that the addressee is a target.

7 This sarcastic use of the past tense for future action has its Hindi-Urdu counterpart in the auxiliary use of V *čuk-* 'have already V-ed' which derives historically from the main verb *čuk-* 'be finished, used up'. As an indicator of sarcasm, auxiliary *čuk-* corresponds to English cues of sarcasm like 'sure!' or 'you bet!'. Its simple past (i.e. preterite) tense as in (a) [an illustration from Dasa et al.'s 1965-75] and in (b) most typically negates an act in the future:

(a) *tum* *ab* *ā* *čuke!* *(arthāt* *'tum* *ab* *nahĩ* *ā-oge')*
you now come already that.is you now not come-2PL.FUT.M.PL
'You've already come!' (Intended meaning: 'You will not come now.')

(b) *us.ke* *pās* *khā.ne* *ke.liye* *paise* *nahĩ* *tumhāre* *paise* *vāpas* *de* *čukā!*
him near eating for money not your money back give has.already
'He doesn't have enough for food! I'm sure he's gonna return your money!'
IM: *bhūl* *jāo:* *vo* *tumhẽ* *paise* *nahĩ* *dene* *kā* *(hai)!*
forget GO he you.DAT money not giving of is
'Forget it! He's not about to give you your money!'

e-ISSN 2785-5953

(13) */ su hay // s-oy/ āv! tsi gatsh-akh khwaš!*
/ he indeed // he-indeed / came you go-FUT.2SG happy
'He came indeed! You may rejoice!'(IM: 'He's not about to come so don't get all happy!')

(14) *b-ey gō-s tsi čhu.kh khōts.ān*
I-indeed went.M.SG-NM.1SG you are fearing
'Sure I went. You are worrying?'(IM: 'Don't worry! I'm not gonna go!')

Indeed, the particle *hay* or its affixal counterparts[8] are sufficient in themselves to mark sarcastic intent:

(15) *ʦ-ey yi-kh tang kōryen manz*
you-indeed come-FUT.2SG tight girls.DAT.PL among
'You will be bored in the company of girls!' (IM: 'Sure you'll be bored with all those girls around!')

Another common cue of sarcastic intent is use of the invariant (oblique singular) form of *baḍ-* 'big' as an adverb of quantity: *baḍi* 'a lot; very' (sarcastically 'sure; you bet!'):[9]

8 Note the sandhi: *su* + *hay* => *soy* 'he indeed'; *tsi* + *hay* => *tsey* 'you indeed'; *bi* + *hay* => *bey* 'I indeed'.

9 The Hindi-Urdu parallel to this is the use of the adverb of quantity *baṛ-* 'a lot', sometimes in its masculine singular default form *baṛā* (a) (Kusum Jain from Hook, Jain 2002, 364):

(a) *vo baṛ-ā d-egī pārṭī!*
she big-DEF give-FUT.F.SG party(F.SG) [DEF = default = M.SG]
'Sure she's gonna give a party!'

and sometimes as an absolutive concordant adverb agreeing in gender and number with an intransitive subject (b) or a (transitive) object (c):

(b) *tum baṛ-e ā-oge madad kar-ne! jhūṭhe vāyde karte ho!*
you big-M.PL come-2.FUT help do-INF false promises make are
'Sure you'll come to help us! You don't keep your word.' (Kusum Jain from Hook, Jain 2002, 364)

(c) *vo baṛ-ī d-egā pārṭī!*
he big-F.SG give-3SG.FUT.M.SG party(F.SG)
'Sure he's gonna give a party!' (Kusum Jain from Hook, Jain 2002, 364)

(16) *swa* *čha* *ālitsy;* *swa* *kar-yi* *baḍi* *vudyūg!*
she is lazy she do-3SG.FUT a.lot hurry
'She is very lazy; sure she'll be quick!'

That the adverb *baḍi* in these examples functions as an adverb which takes the entire utterance in its scope may explain its unusual ability to occur in root clauses together with another constituent to the left of the finite verbal element without forcing the other constituent to change its position:

(17) *ʦ-ey* *baḍi* *yi-kh* *kẽẽsyi* *bakār!*
you-indeed a.lot come-FUT.2SG someone.DAT assistance
'Sure you will be of assistance to someone!'

In its power to condition V-3 word order sarcastic *baḍi* should perhaps be grouped together with other sentence-operators such as the set of Wh-words. Or the particle *hay* / *-ay* / *-y* may reset the V-2 count.[10] Compare (18) with (19) and contrast them both with (20) where *baḍi* is not being used sarcastically:

(18) *ʦ-ey* *baḍi* *khwaš* *gaʦh-akh!*
you-indeed a.lot happy go-FUT.2SG
'Sure you will be happy!'

(19) *ʦi* *kūt* *khwaš* *gaʦh-akh!*
you.NOM how.much happy go-FUT.2SG
'How happy you will be!'

(20) *ʦi* *gaʦh-akh* *baḍi* *khwaš!*
you go-FUT.2SG a.lot happy
'You will be very happy!'

10 See Hook, Koul forthcoming for examples and discussion of the role that the emphatic particle -(*a*)*y* 'indeed' may have in conditioning V-3 word order in Kashmiri root clauses.

3 Neutral Sarcasm

There is sarcasm of a second kind in which the speaker is not concerned with attacking or ridiculing the beliefs in particular of the hearer. Rather it is the event or situation itself that is being held up to some implicit standard that the speaker assumes his listeners subscribe to (or believes they ought to subscribe to it). The target is the subject of the clause but the focus is on the action or event predicated of the subject rather than on his or her (imputed) beliefs or attitudes. A cue to sarcasm of this sort is the use of the exclamatory particle *-ā*(*h*) suffixed to the first constituent of the clause:

(21) *šakil-**ā*** *čha-s!*
beauty-**ā** is-3SG.DT
'What a beauty he / she is!'

(22) *čāy-**ā*** *čēvyi-kh!*
tea-**ā** served-3PL.ER
'The tea they served!'

If the hearer happens to be the subject then he or she becomes the target:

(23) *hyemith-**ā*** *keri-th!*
courage-**ā** did-2SG.ER
'What courage you showed!'

(24) *poz-**ā(h)*** *čhukh* *van-ān!*
truth-**ā** are.2SG tell-ing
'Right!'

Why are rhetorical questions recruited as a mode for delivering sarcasm? Perhaps because they have the same form as real questions, they may provide the sarcast with desired cover: it is harder for the hearer to stop the conversation to make an explicit complaint or accusation if the speaker's words (and intentions) are ambiguous. It is, after all, the desire of the speaker to deliver a psychological blow without assuming all the risks of making an explicitly hostile remark that is the fundamental motive for using sarcasm.

4 Hearer-Oriented versus Subject-Oriented Sarcasm

The cues we have examined so far are used to indicate to the hearer that the speaker does not subscribe to the hearer's views (25). In (26a-b) we may observe another kind of sarcasm, one which targets not the hearer of the utterance but rather its 'subject' (usually an agent-subject or experiencer-subject). Sarcasm of this subject-oriented kind shows up when the speaker assumes that the hearer agrees. Contrast (25) with (26a-b):

(25) *su* *hāv-yi* *baḍi* *panun* *bajar*
he show-FUT3sg a.lot self's greatness
'Sure he'll show his greatness!' [IM: 'He's not so great (as you think he is)!']

(26a) *baḍi* *āv* *bajar* *hāv-an.vōl*
a.lot came.M.SG greatness show-er.M.SG
'Here comes the big cheese!' [IM: 'He is not so great (as he thinks he is)!']

(26b) *baḍi* *āyi* *bajar* *hāv-an.vājinyi*
a.lot came.F.PL greatness show-ers.F.PL
'Here come the big cheeses!'
[IM: 'They are not so great (as they think they are)!']

In (25) the speaker is using sarcasm to undermine or ridicule the hearer's expressed or implied position that someone has great potential. Whereas in (26a) and (26b) the speaker is not attacking the views of the hearer but rather ridiculing the self-indulgent behaviour or the pretensions of the subject.

If the hearer happens to be the subject then it follows that the sarcast is attempting to degrade her or him:

(27) *tswapi* *kar!* *baḍi* *āy-akh* *tatyi* *pyeṭhi* *sād* *ben-yith*
silence make a.lot came.F.SG-2SG.NM there from saint become-GER
'Shut up! You've come from there so pure and holy!' [IM: 'You're not so holy as you think!']

5 Is there a Construction Dedicated to Subject-Oriented Sarcasm?

The construction in (26ab) and (27), widespread in languages spoken on the northern and western sides of India, displays a specific pattern, definable as in (28). One may speak of a 'dedicated' construction:

(28) (subj) + *aggrandising element* (or *Wh-*) + *finite form of {come}* + mocked action, state, attitude[11]

11 On hearing some of these examples Colin Masica (pers. comm.) objected that a sarcastic construction "Here comes / a // the / big X" exists in English (and presumably in all languages). Of course, in the right situation and with the right intonation any utterance can be interpreted as sarcasm. Construction (28) as illustrated in (26)-(27) and (29)-(40) differs from speech overlain with sarcastic intonation in being special-

(29) ***khuba ā-ek-i če*** *sevā gar-na!* [Nepali]
a.lot came-PP-F.SG is.F.SG service do-INF (Netra Paudiyal, pers. comm.)
'She's come to help!' [IM: 'She's too late to help.' or 'She's not one to help others.']

(30) ***baṛ-i ā-ī*** *čāvi dend-ār-i…* [Garhwali]
big-F.SG came-F.SG key giving-NOM.AG-F.SG
'So here she is, the big key-giver!' (Ghildiyal 1981, 49)

(31) *vo* ***baṛ-ī ā-ī*** *paropakār kar.ne-vāl-ī!* [Hindi-Urdu]
big- F.SG came-F.SG help do-er-F.SG
'Here she comes, the Good Samaritan!' (Hook, Jain 2002, 365)

(32) ***baḍḍ-ī āy-ī*** *ũč-ī bātã̃ kar.ṇa-āḷ-ī* [Bagri / Haryanavi]
big-F.SG came.F.SG high-F.PL things do-er-F.SG
'Here she is, the big talker (IM: the pompous ass!)' (Lakhan Gusain, personal communication)

(33) *vaḍ-ī* ***ā-ī*** *paṛī-likhī… paṛī-likhī hai to*
big-F.SG came.F.SG read-F.SG-written-F.SG read-F.SG-written-F.SG are then
is-kā mīning batā
this-GEN meaning tell
'Here she is, the highly educated one. So, if you're so educated, tell me the meaning of this!' (sharechat.com/video/MxQ8q4W?referrer=url) [Panjabi]

(34) *vaḍ-o* ***āyo*** *mohabbat kar-ṇa-vār-o*
big-M.SG came.M.SG love do-INF-ER-M.SG
'Here he is, the great lover!' [Sindhi]
(sindhiadabiboard.org/Catalogue/mehran/Book25/Book_page7.html)

(35) *jā jā pāgal gayo če,* ***moṭ-o*** ***āvy-o*** *papi māg.vā-vāḷ-o*
go go crazy gone are big-M.SG came-M.SG kiss ask-er-M.SG
koi juv.e to šũ kah.e
someone see then what say
'Get away. You crazy? The big kiss-demander! What would someone say if they saw?' (gadyasarjan.wordpress.com/2012/12/26) [Gujarati]

ised for [or dedicated to] the delivery of sarcastic intent. That is, for speakers to utter any of these examples without sarcastic intent is unlikely maybe even impossible.

(36) *hā koṇ ālā ṭikojīrāv āmhā-lā sāṅg-ṇār-ā* [Puneri Marathi]
he who came-Msg nosy.parker us-DAT tell-PRESPRT-Msg
'Who is this kibitzer-shmitzer to tell us (what to do)?' (https://www.maayboli.com/node/26768)

(37) *lay āl-ā sallā de.ṇār-ā* [Wardha / Nagpuri Marathi]
much came-M.SG advice giver-M.SG
'Here comes the great advice giver!'(P. Mashram and R. Mhaiskar, via Sonal Kulkarni-Joshi)

(38) *vhəll-ɛ̃ ayl-ã səllɔ div-pa-k* [Goan Konkani]
great-F.SG came-F.SG advice give-INF-DAT
'She's a great one to give us advice!'(N.F. Gaonkar and G. Mopkar, via Sonal Kulkarni-Joshi)

(39) *pedda* ***vatts-āḍu*** *bādhyata gala pauruḍ.i-lāgā* [Telugu]
big come-PST.M.SG responsibility with citizen-like
'… as if he were a responsible citizen!'*
* Validity of Telugu data, glossing, and analysis confirmed by K.V. Subbarao, Peri Bhaskararao, and Shalinee Gusain.

(40) *avaḷ* ***perīyya*** ***vant(-uṭṭ)[12]-ā(ḷ)*** *eṉak-ku camaiyal collit-tar-a*
she bigly came-LET-F.SG me-DAT cooking teach-GIVE-INF
taṉak-k.ē tōcai kūṭa vārk-ka teriyātu
self-DAT dosa even pour-INF know-NEG
'She's a good one to teach me how to cook. She can't even manage to make a dosa herself!' (This Tamil example is from Kanaka Jagannathan, via Bharati Jagannathan, personal communication)[13]

12 The morpheme *-uṭṭ-* -LET- is the colloquial abbreviation of *viṭṭ-*, the past tense form of vector (*vi*)*ṭu* {LET GO, RELEASE}. See Annamalai 2021, 308 ff. for detailed description of (*vi*)*ṭu* and other Tamil vectors.

13 The Tamil example in (40) patterns identically to most of the Indic examples in (26a-b)-(38). However, not every Tamil speaker accepts *perīyya* in (40), preferring instead the adverb *perusā*. (Umarani Pappuswami, pers. comm.):

(a) ***per-usā vant(-uṭṭ)-ā(ḷ)*** *eṉak-ku camaiyal collit-tar-a! taṉak-kē tōcai kūṭa vārk-ka teriyātu*

6 Is the Sarcasm Construction Displayed in (28) a Feature of the South Asian Linguistic Area?

In all of the South Asian languages surveyed expression of sarcasm may use either a concordant form of a mocking aggrandising adjective *baṛ*- 'big' (Hindi, Garhwali) / *baḍḍ*- (Bagri, Haryanavi) / *vaḍ*- (Panjabi, Sindhi) / *moṭ*(*h*)- (Gujarati, Marathi) / *vhəll*- (Konkani); an invariant adjective *bhāri* 'heavy' (Bengali) / *mahā* 'great' (Kannada) / *pedda* 'big' (Telugu); an adverb *perusā* (Tamil) / *baḍi* (Kashmiri) / *khuba* (Nepali) 'bigly, a lot' / *lay* (Nagpuri) 'very'; or an interrogative pronoun *koṇ* (Marathi) 'who'. Contrastingly, the full pattern displayed in (28) is not found in Bengali (Probal Dasgupta, pers. comm.), while Kannada has what seems like an inverted form[14] of it (S.N. Sridhar, pers. comm.). The distribution of the templatic pattern shown in (28) across most of Indo-Aryan in a contiguous block taken together with its presence in Dravidian Telugu and Tamil accords with its being regarded as an incomplete or fragmentary feature of South Asia as a linguistic area (Masica 1976),[15] perhaps one yet to reach its full extent.

It remains to be seen if specialised constructions or explicit markers of sarcastic intent parallel to Kashmiri's *baḍi yi*- V-*an*-*vōl*- (26a) are found in languages spoken outside South Asia and if so whether they are commonly used in the languages that have them.[16]

14 Notice that in (a) with respect to the template in (28) finiteness and non-finiteness of forms have switched places:

(a)	*avaḷu*	*ba-nd.u*	*(mahā)*	*nan-gē*	*hēḷi-koḍ-tā-ḷe*	[Kannada]
	she	come-PPART	great	me-DAT.EMPH	tell-GIVE-NON.PST.F.3sg	

'She presumes to teach me – (the big know-it-all)'. (S.N. Sridhar, personal communication)

This difference brings the Kannada construction closer to normal South Asian SOV word order. Thus, rather than constructionally, the cue to sarcastic intent must be indicated intonationally and/or by the presence of *mahā*.

15 See Emeneau's definition of a linguistic area: "an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of at least one of the families" (1956, 16 fn 28).

16 Does the construction in (28) share features with other kinds of insincere speech? One reason for thinking it may is that in all these languages the finite form of {come} is not in its normal clause-final position. In all of them some material follows the verb: an infinitive, a participle, or a noun phrase expressing the actions or attitudes on the reality or legitimacy of which the sarcast casts doubt. Comparable to this displacement from the canonical clause-final position of the finite verb is the 'move-left' phenomenon observed in Hindi-Urdu expressions of irony, especially those involving inceptives (Hook 2011):

(a)	***lage***	*aurõ*	*kī.tarah*	*tum*	*bhī*	*čāplūsī kar-ne*
	begun.2PL.M	others	like	you	too	flattery do-INF

'There you go, just like the others, trying to flatter me'. (Premchand [1936] 1960, 51)

Abbreviations

1	first person
2	second person
3	third person
ABL	ablative
AC	accusative pronominal suffix
AG	agentive
DAT	dative
DEF	default or invariant form
DT	dative pronominal suffix
EMP	emphatic particle
ER	ergative pronominal suffix
ERG	ergative
F	feminine
FUT	future
GEN	genitive
GER	gerund
IM	implied / intended meaning
INF	infinitive
M	masculine
NEG	negative
NM	nominative pronominal suffix
NOM	nominative
NON.PST	non-past
PL	plural
PPART	past participle
PRESPRT	present participle
PST	past tense
SG	singular

Bibliography

Annamalai, E. (2021). "The Matrix of Verb-Verb Sequences in Tamil". Kageyama, T.; Hook, P.E.; Pardeshi, P. (eds), *Verb-Verb Complexes in Asian Languages*. Oxford: Oxford University Press, 300-26. https://doi.org/10.1093/oso/9780198759508.003.0012.

Dasa, S. et al. (1965-75). *Hindī šabd sāgar* (The Ocean of Hindi words = Hindi Dictionary). Kāšī: Nāgarī pračārinī sabhā.

The leftward position of *lag-* 'begin', *ā-* 'come', and perhaps other predicates governing non-finite complements is recurrent in Hindi-Urdu dialogue and in them functions as a marker of a set of related attitudes.

Emeneau, M.B. (1956). "India as a Linguistic Area". *Language*, 32, 3-16.
Ghildiyal, D. (1981). *Gāri (Pebbles?)*. New Delhi.
Haiman, J. (1990). "Sarcasm as Theatre". *Cognitive Linguistics*, 1, 186-209.
Hook, P.E. (2011). "Semantics and Pragmatics of Non-Canonical Word Order in South Asian Languages. <verb-left> of *Lag-* 'Begin' as an Attitude-Marker in Hindi-Urdu". *Lingua Posnaniensis*, 53(2), 25-34. https://doi.org/10.2478/v10122-011-0010-9.
Hook, P.E.; Jain, K. (2002). "How to Be Sarcastic in Hindi-Urdu". Deshpande, M.M.; Hook, P.E: (eds), *Indian Linguistic Studies. Festschrift in Honor of George Cardona*. Delhi: Motilal Banarsidass, 358-73.
Hook, P.E.; Koul, O.N. (forthcoming). "Exceptions to V-2 in Kashmiri". In a commemorative volume for V.K. Kaul.
Jagannathan, V.R. (1981). *Prayog aur prayog. Handbook of Modern Hindi Usage.* New Delhi: Oxford.
Masica, C.P. (1976). *Defining a Linguistic Area. South Asia*. Chicago: The University of Chicago Press.
Premchand [1936] (1960). *Godān* (*The Gift of a Cow*). Allahabad: Saraswati Press.
Taing, M.Y. (1984). "Makbūlinyis kalāmas manz tanz ti myizah" (Irony and satire in Maqbul's writings). *Šeerāza*, 21(5), 21-30.

e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

‘Sanskrit-Speaking’ Villages, Faith-Based Development and the Indian Census

Patrick S.D. McCartney
Nanzan University Anthropological Institute, Nagoya

Abstract Over three sections, the 2001 and 2011 Indian censuses are scrutinised to locate, down to the sub-district administrative and village levels, where L1-L3 (first to third language) Sanskrit tokens were returned during census enumeration. First, there is a theo-political discussion of Sanskrit’s imaginative power for faith-based development. This includes a discussion on how ‘Sanskrit-speaking’ villages signify an ambition toward cultural renaissance. Next, Sanskrit’s national-level enumeration is discussed. Finally, closer scrutiny is paid to the top four states (Maharashtra, Bihar, Madhya Pradesh, and Uttar Pradesh). On average, more Sanskrit tokens were returned by men than women; 92% of L2-Sanskrit tokens are linked to L1-Hindi; most L1-L3-Sanskrit tokens cluster with Hindi, English, and/or the State Official Language; most Sanskrit tokens are Urban, as opposed to Rural; and most tokens are found across the Hindi Belt of north India.

Keywords ‘Sanskrit-Speaking’ villages. Hindu nationalism. Linguistic utopia. Social imaginary. Mother tongue.

Edizioni Ca’Foscari

Peer review

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

Citation McCartney, P.S.D. (2022). “‘Sanskrit-Speaking’ Villages, Faith-Based Development and the Indian Census”. *Bhasha*, 1(1), 77-110.

DOI 10.30687/bhasha/8409-3769/2022/01/007

1 Introduction

bhāṣā saṃskṛtāpabhraṃśaḥ bhāṣāpabhraṃśastu vibhāṣā |
sātattaddeśa eva gahvaravāsināṃca prakṛtavāsināṃca ||
Abhinavabharati 17.49

The corruption of Sanskrit is *bhāṣā*, and the corruption of *bhāṣā* is *vibhāṣā*. It is the language of the same countries of forest dwellers and of rustic people. (Kavi 1934, 376)

This paper is part of the *Imagining Sanskritland* project, which focuses on locating and documenting how, where, and why the Middle Indo-Aryan language, Sanskrit, is spoken in the twenty-first century. This builds on the previous work of Nakamura (1973), Hock and Pandaripande (1976), Hock (1991), Aralikatti (1989; 1991), Aklujkar (1996), Hastings (2004; 2008), and Deshpande (2011; unpublished). Generally, this project expands beyond linguistic analysis of sentence structure to document the aspirations, ideologies, and moral horizons inherent in identifying as a speaker, as well as documenting second-language acquisition through a focus on imperfect learning, substrate interference, and bilingualism.

The project first focused on code-switching between Hindi and Sanskrit and the transubstantiation of symbolic capital in a residential Sanskrit college/yoga ashram in Gujarat, India (McCartney 2011; 2014a; 2017a; 2018a). Another focus was the two-week intensive speaking course in New Delhi (McCartney 2014b). The focus pivoted to include a relatively famous 'Sanskrit-speaking' village in Madhya Pradesh, India (McCartney 2015; 2016a; 2016b; 2016c; 2017b; 2017c; 2017d; 2018b), which includes discussion of language revival and hybridisation (McCartney, Zuckermann 2019). This combines with analysing the theo-politics of Sanskrit's imaginative consumption within the transglobal wellness industry and the topic of 'yoga fundamentalism', which maps out the distanced and banal ways that consumption of yoga lifestyles can potentiate tacit and unwitting support of Hindu supremacism (McCartney 2017e; 2017f; 2017g; 2019a; 2019b; 2020). More recently, the project has pivoted to cover matters related to Sanskrit's soft power potential within the context of faith-based, sustainable development and competitive diplomacy. This involves environmental impact assessments of yoga lifestyle brands and the cultivation of nostalgic moods predicated by Neo-Romanticism, mystical holism, and dark green religion (McCartney 2021a; 2021b; 2021c).

Informed by Deumert's (2009) and Posel and Zeller's (2015) demographic analyses of census data related to language shift and bilingualism in South Africa, this project pivots to look for Sanskrit

I would like to thank Andrey Klebanov for helping with particular issues in developing this paper.

('speakers') in the Indian census data. An initial result, upon which this present paper directly builds, is McCartney (unpublished). The policy of India's census enumeration states that if the number of speakers of any language drops below 10,000 then it will no longer be reported as a separate language (Goswami 2012).[1] If Sanskrit were to dip below the threshold, this would have unintended consequences for its soft power deployment. Therefore, the politics of census enumeration for the purposes of state building are relevant. Simply looking for Sanskrit speakers is something of a fool's errand. Sanskrit is a post-vernacular language in a perpetual state of acquisition. The media only reports the perceived success of Sanskrit's revival (Indian Yug 2020).

In short, India's 2011 census results clearly demonstrate that Sanskrit 'speakers' are overwhelmingly found in urban areas spread across the 'Hindi heartland',[2] which is only a part of India's complex linguistic ecology and "linguistic area" (Emeneau 1956). What this paper does is use the Indian government's census data to geographically locate where *people who identify as speakers of Sanskrit were at the time of census enumeration*. For now, this is as good as it gets, as the data presented below are not capable of verifying the fluency of people who claim to speak Sanskrit. Discussion of several issues relating to this are found across the project's publications mentioned above. The outcomes from this present paper include future research being more strategic.

Even though some consider Sanskrit to be the "language of the rural masses" (Deopujari 2009) and the "language of future India" (Mohan 2020), it is also thought that "Sanskrit is the forgotten language of urban India" (Indian Eagle 2020) and that "NASA believes San-

1 This is one reason why the RSS (Rashṭriya Svayamsevak Sangh, 'National Volunteer Corps') "wants citizens to voluntarily register Sanskrit as their second language in the census. The RSS feels that if people register the language, the final census data would reflect higher literacy of Sanskrit, which will force the government to take measures to preserve the language" (Tare 2010).

2 Compare the notion of the "Vedic God" by Hebden (2011). The geographic focus of this paper is the 'Hindi Heartland' or 'Hindi Belt', which covers most of the plains of north India, where Sanskrit's close relatives – Hindi and its related languages – are spoken. This is pertinent because, as is discussed below, this is the linguistic area where the census results indicate most of the 'Sanskrit speakers' live. As is evident, below, the link between people who identify as Sanskrit speakers with both Hindi and English is immense. By way of example, the state of Bihar's Sanskrit areas is discussed below. This paper expands upon Jha's (2017) explanation of how studies of language politics in north India tend to focus on the Hindi-Urdu debate. This debate builds on a centuries-old development of language order in premodern India (Ollet 2017), taking on communalist narratives. This culminated in the nineteenth century around which of these mutually intelligible languages – Hindu and Urdu, which derive from Hindustani, but use different scripts, respectively, Devanāgarī देवनागरी and Nasta'līq نستعلیق – should become the national language. Currently, India does not have a national language. Instead, it has two official languages, Hindi and English (Dasgupta 1995).

skrit is a scientific language for programming" (TNN 2019). Moreover, Sanskrit is thought to be a "gift of India for [the] entire humanity" (India Education Diary Bureau 2020) and that the "Sanskrit effect" is caused by "chanting Sanskrit", which "increases brain cognitive areas" (Sanskriti 2018). The benefits of chanting are predicted to extend beyond humans, as "Cows will talk in Tamil and Sanskrit" (Patherpanchali 2018). Back down on the ground, it is difficult to locate Sanskrit speakers because the available information are unreliable factoids mentioned on the internet,[3] copied and pasted from

3 Yet, these lofty ambitions have humble origins amongst the mythical villages, whose inhabitants are meant to be grateful that Sanskrit's perceived civilising power will finally reach them, even if this ideological benevolence is soaked in a neo-colonial Sanskritisation impetus that is made explicit in ways, such as "*Saṃskṛtaṃ sarveṣāṃ kṛte… sarvadā*; Sanskrit for everyone… forever" (Amaravāṇi 2020). Strength, it seems, is not found in linguistic diversity. Amara-vāṇī (immortal-language) is ultimately a part of Samskrita Bharati, which itself is the linguistic node of the more prominent Hindu nationalist parent organisation, the Rashṭriya Svayamsevak Sangh (RSS). It is through its international branch, the Hindu Svayamsevak Sangh (HSS), that Samskrita Bharati operates at an international level. It mostly services the Indian diaspora through cultural and linguistic events, which can be an opportunity for the collapsing of the 'big' versus 'little' tradition binary (Vertovec 1994), such as a potential muddling of Sanskrit as it goes through its interlanguage stage of acquisition. However, the expansion beyond the imagined borders, especially through extending into cyberspace, requires a recalibration of relations, especially to the *puṇya bhūmī* ('sacred land') of the *Hindutvavādins* imagination, within the context of transnational development and multiple modernities (Jaffrelot 2017). After all, "Sanskrit is a gift of India for entire humanity". At least, that is what India's HRD Minister, Ramesh Pokhriyal, asserted just after the *Central Sanskrit Universities Bill, 2020* was passed by India's upper house of parliament to upgrade three Deemed Sanskrit Universities to Central University status (India Education Diary Bureau 2020). Amaravāṇī promotes Sanskrit through songs. One example is the song *Viśva-bhāṣā Saṃskṛtam* ('The Universal-Language is Sanskrit'). Information on the song's page, on Amaravāṇī's website, also claims, that "There are many villages in India where the entire population speaks solely and fluently in Saṃskṛtam!" (Amaravāṇi 2020). Such truth claims are a curious thing. I am reminded of one verse from a seventh century Ayurvedic text, which discusses poor vision resulting from false perception: "*Dūrāntikasthaṃ rūpañca viparyāsena manyate | doṣe maṇḍalasaṃsthāne maṇḍalānīva paśyati* || AS.Utt.15.4 ||" "Due to false perception (*viparyāsa*), a patient perceives a thing located far away, as close by, and things located close by, as far away" (Aṣṭhāṅgasaṅgraha, Uttarasthānam, 15.4 [Vāgbhaṭa 2020]). In a topsy-turvy way, *viparyāsa* refers to the act of imagining something to be real and true, when it is false. The term *niścaya* can mean both correct perception and enquiry. We find in the Vaiśṇava tradition encouragement to cultivate *niścaya* (*Śrīmad Bhāgavatam* 3.26.30), which is recommended for both soteriological aspirations and mundane matters. This concept is similar to that of Nāgārjuna (c. 150 CE-c. 250 CE), who discussed, in the *Mahāprajñāpāramitāśāstra*, the concept of *pratītyasamutpāda*. This refers to the "basic principle of thought that no two contradictory judgements can hold good in regard to the same thing in the same respect" (Ramanan 1987, 167). In other words, 'Sanskrit-speaking' villages either are true and do exist or they do not. There is very little available evidence, be it documentary, direct, real, circumstantial or testimonial. Furthermore, what percentage of a village's population and to what degree of fluency and ordinal ranking of usage, considering frequency of code-switching, domains and topics, might be the minimum requirements to satisfy claims of any village being one that is 'Sanskrit-speaking'?

other websites lacking accurate details. Nonetheless, these factoids are enough for people to believe that 'Sanskrit-speaking' villages exist.[4] People rely on this meme to provide emotional reinforcement for deeply held religious beliefs hoping it will re-enchant the world to provide profound meaning in one's life. 'Sanskrit-speaking' villages endow the world with a perceived sense of divinity, meaning, and significance. As an empty signifier, the village is remote enough to remain perceived as an infallible closed circuit. Similarly, Olshansky, Peaslee and Landrum (2020) provide insight into the cognitive defence mechanisms of flat earthers, which include motivated reasoning to dilute cognitive dissonance and maintain cognitive consistency. Having visited several of these so-called 'Sanskrit-speaking' villages, I became increasingly frustrated, as most of these villages contain hardly anyone who can hold a casual conversation across general domains and topics or utter increasingly complex sentence patterns, and include more complicated use of tense, aspect, or mood. Nonetheless, throughout the last decade, my primary question has been: "Where are the Sanskrit speakers?".[5]

Jhirī is a village in Madhya Pradesh that I have paid more attention to (McCartney 2015; 2016a; 2016b; 2016c; 2017b; 2017c; 2017d; 2018b) where apparently everyone speaks Sanskrit (Samskrit101 2009). Oblivious to issues of linguistic human rights (Skutnabb-kangas 2012), Ghosh (2008) celebrates how the residents "hardly speak the local dialect, Malvi, any longer. Ten years have been enough for the sanskritization of life here". A potent claim that "even those who don't know the technicalities of the language still speak fluent Sanskrit" (Hindutvainus 2011) is demonstrably false. For instance, while conducting fieldwork in Jhiri, one of the residents repeatedly claimed that everyone in village fluently spoke Sanskrit. Beaming with pride, he often emphatically asserted, "*Bhoḥ, asmin grāme iva sarve janāḥ saṃskṛtaṃ vadituṃ śakyante khalu* [Sir, in this village everyone can speak Sanskrit!]". However, this was easily disproven by speaking to

4 My term, *laukika-saṃskṛta-bhāṣā-aviparyāsa-ābhāsa-samanveṣaṇam*, refers to the attempt to look everywhere (*samanveṣaṇa*) for the semblance (*ābhāsa*) and distorted perception (*aviparyāsa*) of vernacular (*laukika*) variants, or regional dialects, of Sanskrit (*saṃskṛta*), as a spoken language. This includes its mention in yoga-related outlets (Dizon 2016; Bedewi 2020). By first identifying the 'pseudo-perception' (*pratyakṣa-ābhāsa*) found in discourse (*vyavahāra*), the process moves to falsifying (*dūṣayati*) fallacies (*hetvābhāṣa*) related to the enduring claims about 'Sanskrit-speaking' villages, which are legion.

5 In logic, *ābhāsa* takes the meaning of 'erroneous though plausible argument'. These rumours act as a buffer warding off existential anxiety. At least it is potentially comforting to know that a 'real' and 'true' India still exists and that "India's own Jurassic Park" is found in rural Madhya Pradesh, in a village that "is a lost world that has been recreated carefully and painstakingly, but lives a precarious existence, cut off from the compelling realities of the world outside" (Ghosh 2008). The *village* holds an ambiguously utopian relation to future India (Nandy 2001; Mohan 2012).

anyone within arm's reach, who would require prompts from others as they could not reply to simple sentences, such as "*Bhavataḥ nāma kiṃ* [What is your name?]" or "*Mitram, kutra gacchasi* [Friend, where do you go?]", let alone pass sentence repetition tests. Yet, these rumours become factoids and bloom into unassailable facts in support of "core Indian values" (Prabhu 2014). One potent example is the use of spoken Sanskrit to sell a motorbike, which uses a 'Sanskrit-speaking' village as the backdrop (Sharma 2009).

One issue with validating the returned Sanskrit *tokens* in the census data is that they are self-reported. Another issue is that census enumerators are bound by law not to question the maximum three languages given. Therefore, if someone believes they speak Sanskrit (or any other language) and identify as a speaker of Sanskrit then it becomes a demographic fact.[6] Simply put, one token refers to one individual instance of someone claiming to be either a L1 ('mother tongue'), L2 (second-language), or L3 (third language) speaker.[7] These epistemic methodological issues have been discussed at every census (Office of the Registrar General 2020).

All of the data are publicly available in downloadable spreadsheets (Office of the Registrar General 2020). The 2011 data come from the relevant C-16 ('mother tongue')[8] and C-17 (bi-/tri-lingualism) tables, which only became available in late 2018. Multiple versions of these

6 India, like many nations, has used different and equally ambiguous terminology to capture the primary language(s) used by their citizens. There are actually three language 'situations' that can be captured by censuses: (a) the language first used by the respondent; (b) the language most commonly used by the respondent at the time of the census; and (c) the knowledge of particular official language(s) by the respondent (Arel 2004). The Indian census does not appear to achieve this tertiary system, simply because it does not ask the necessary questions. Is this due to either pragmatic expediency, ideology, or both? Building upon Foucault 2007, Duchêne, Humbert, Coray 2018 explain the consequences of reducing real world complexities through statistics to quantifiable categories. These can become tools of governance for national solutions. This, after all, is probably the point. Yet, if there is an ideological component, it then becomes more challenging, particularly when added to the epistemic relativism of self-reporting, to gather meaningful data to implement productive policy.

7 This present study does not have the capacity to verify whether these tokens translate into real-world pragmatic abilities to communicate in Sanskrit.

8 A 'mother tongue', is defined in the Indian Census guidelines (3.1) as, a "the language spoken in childhood by the person's mother to the person. If the mother died in infancy, the language mainly spoken in the person's home in childhood will be the mother tongue. In the case of infants and deaf mutes, the language usually spoken by the mother should be recorded. In case of doubt, the language mainly spoken in the household may be recorded" (Office of the Registrar General 2020). Section 3.2.a-d stipulates that an 'enumerator' is "Bound to record the language as returned by the person as her/his mother tongue"; and that they must record the "mother tongue in full, whatever is the name of the language returned by the respondent and do not use abbreviations". Collectors are not expected to determine if a "language returned by a person is a dialect of another language"; and, if there is any "relationship between mother tongue and religion" (Office of the Registrar General 2020). This gets complicated

tables exist, as each state has a separate C-16/-17 table. Having collected and sorted millions of data points, each table was filtered using excel data analytic functions. The subsequent aggregate token amounts were then cross-analysed over several levels of administration. These levels include the national, state, district, sub-district levels, as well as rural and urban zones and sex ratios.

A key finding demonstrates high levels of affinity between the triune of Sanskrit (S), Hindi (H), and English (E), which is also nuanced by each state's official language (SOL). This means that people who identify as L1-L3 speakers of Sanskrit are overwhelmingly clustered within a specific set of L1-L3 languages, which more often than not includes Hindi and English.[9] Therefore, the statistically relevant identity of the typical 'Sanskrit speaker' is an educated, middle-class urbanite who lives in north India.[10]

Based on 2011's C-16 table, the total nation-wide number of L1-Sanskrit Persons amounts to 24,821 (tokens); **[fig. 1]** shows the totals for all states and union territories.

when Hindu nationalist groups urge people to list Sanskrit as their L1 in the lead up to each census (Tare 2010).

9 This has much to do with India's language planning policy, which prefers a Sanskritised Hindi as the official language (see Articles 343 and 351 of India's constitution). For example, L1-S_L2-H_L3-E would mean an individual's languages are L1-Sanskrit, L2-Hindi, and L3-English. The following formula S=(L1α_L2β_L3γ)(SOL) is an attempt to articulate the ways in which Sanskrit can be found within complex linguistic ecologies and the influence of the state language in modifying the triadic SHE. For instance, across the language area of Maharashtra, the state language, Marathi, competes, as it were, for space on, like different taste buds, on the tongues of the state's citizens.

10 In the post-Independence era, studies related to contemporaneous spoken Sanskrit were initiated first by the Sanskrit Commission (Azad 1957), which laid down several recommendations for preserving and promoting Sanskrit as a spoken language, some of which have been successfully introduced. Deshpande (2011) explains how, due to the changes in the three-language educational policy, Sanskrit has fared better in the Hindi speaking states than in the non-Hindi speaking states, where a dramatic reduction in students studying Sanskrit occurred once it became optional in 1968 (see Azad 1957, 99). This does not stop activists from trying to install Sanskrit as the national language (*rāṣṭra-bhāṣā*) and global *lingua franca* (*viśva-bhāṣā*) (Ramaswamy 1999). As well, these language politics have long encoded Hindi as a hegemonic language yet raise the status of Sanskrit. Complementing this, Babu (2017, 113) explains that in India Sanskrit sits atop the linguistic hierarchy and caste system by invoking the notion of *catur-varṇa* ('four-classes') by positing Sanskrit as occupying a privileged position and English (which is a rank outsider in the constitutional scheme), as a language with emancipatory capacity due to its positioning outside the legitimised hierarchy.

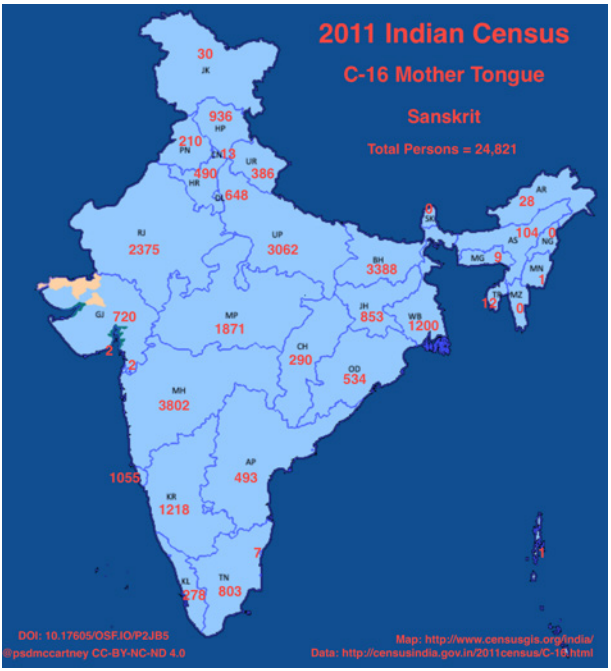


Figure 1 2011 India: States L1-Sanskrit. (India 2020)

A deeper historical review of Sanskrit's enumeration across all 15 censuses occurs in **[tab. 1]**, which provides a cursory glance at Sanskrit's India-wide total Persons' results from 1881 to 2011.

Table 1 Total Persons L1-Sanskrit returns, 1881-2011[11]

Year	Total Sanskrit	Total population	%
1881	1,308	224,000,000	0.0006
1891	308	287,000,000	0.0001

11 The Sanskrit data at each census is located in Plowden 1883, 132; Baines 1893, 144; Risley, Gait 1903, 164, 174; Gait 1913, 106; Marten 1923, 96; Hutton 1933, 492; Yeats 1943, 9; Mitra 1994; Office of the Registrar General 1954, 7; Mallikarjun 2002; Breton 1976, 304; Office of the Registrar General 2020. Yeats explains that "The language and script questions have not been tabulated and I make now a recommendation to the Government of India that they be not tabulated even if the suspended operations are resumed" (1943, 9). Mitra (1994, 3207) explains that war, communal tensions, and Yeat's transition to self-enumeration from household enumeration resulted in a completely botched census that produced incoherent results left unpublished. Mallikarjun (2002) shows that of the 2,554, one person identifies as a speaker of 'Vedic' Sanskrit and another as a speaker of VedPali, while 93 and 5 people respectively claim to speak Pali and Prakrit. Breton (1976, 304-8) provides a good overview of the 1961 census.

Year	Total Sanskrit	Total population	%
1901	716	238,396,327	0.0003
1911	360	252,093,390	0.0001
1921	356	251,321,213	0.0001
1931	1,181	278,977,238	0.0004
1941	N/A	318,660, 580	N/A
1951	555	361,088,090	0.0001
1961	2,554	439,234,771	0.0006
1971	2,212	548,159,652	0.0004
1981	6,106	683,329,097	0.0009
1991	49,736	846,427,039	0.006
2001	14,135	1,028,737,436	0.001
2011	24,821	1,210,854,977	0.002

This paper has three main sections. The first (§ 2) discusses key theo-political points related to Sanskrit's imaginative power. § 3 explicates Sanskrit's national-level enumeration comparing the 2011 and 2001 censuses. § 4 burrows down to the lowest administrative levels of four states to show which districts, sub-districts (*tehsil*/*taluk*, or *taluq*), and in some cases, blocks, returned the highest numbers of L1-L3 Sanskrit tokens.

2 Sanskrit, Theo-Politics, and Faith-Based Development

Inspection of faith-based competitive diplomacy, in relation to Yoga and Sanskrit, is sparse.[12] The ways in which Sanskrit is imbricated is often under appreciated. While for some, Sanskrit might be a dead language and a symbol for millennia of oppression, for others it is a treasure trove of untapped knowledge that might just save humanity and a heritage language one might like to speak (McCartney 2021a). Sanskrit helps define and chart one's path toward a moral horizon. This speaks more about temporalities of becoming rather than being (Fahy 2020). It helps an individual link to an archaic modernity or futured-past and potentially return to an imagined Vedic 'Golden Age' fuelled by re-enchanting, eco-sustainable, neo-Romantic, mystical holism (Hebden 2011; Subramaniam 2019). However, Sanskrit's reclamation and acquisition are indelibly constrained by substrate

12 The most closely related is Jacobs' documentation (2016) of the world's largest volunteer-based charity, the Art of Living Foundation, that originates from India. Watanabe (2019) explores a Japanese organisation's operations in south-east Asia. Haynes's (2021) edited volume explores international relations across several religions. Nelson (2021), however, provides the most comprehensive account of faith-based NGOs as non-state political and moral actors.

interference from the L1s influencing how it, as a L2, is spoken. One of the key assets used to justify Sanskrit's role as a tool for development is its perceived linguistic purity. It is argued that only a 'pure' Sanskrit can deliver the utopian world it inspires. What, exactly, might a pure Sanskrit sound like and how might it power anything? Answering this question is a particularly vexing matter, especially when considering that the earliest layers of the Vedic corpus contain hundreds of loan words from other Indo-European and non-Indo-European languages.[13] During the post-Vedic Period (c. 500 BCE), vernacular Sanskrit, otherwise known as *bhāṣā*, was referred to as the language of the conquerors (Burrow 1965). Burrow showed through a study of the morphology of Classical Sanskrit how the diversity of forms prevalent in the earlier Vedic language reduced significantly, even though it did become the language of Vedic and later Hindu texts, commentaries, rituals, and literature (Subbarao 2008). In the contemporary revival movement, this simplification has continued to the point where vernacular Sanskrit can be feasibly equated with a Sanskritised form of Hindi (Deshpande 2011; unpublished).

The production, reception, and consumption of the 'Sanskrit-speaking' village narratives across various media appear to function in a similar way to *phalaśruti* paratexts. The *phalaśruti* is the textual component listing the benefits of hearing or reciting the particular text. Taylor (2012, 94-5) explains that these "fruits of hearing" texts often include potential punishments and dangers listed along with the promise of heavenly rewards and that this "is a way of enabling the discourse to function as 'true' and is at least partly driven by a distinctly earthly agenda". Another similarity between Taylor's discussion of paratexts and the 'Sanskrit-speaking' village discourse is the way in which quantifiable and qualifiable measurements seem illusive. Nonetheless, the 'Sanskrit-speaking' village boots this signal. Consider the example of this faith-based development narrative that has evolved over the past decade in the state of Uttarakhand, which, in 2010, voted in Sanskrit as its second official language (Trivedi 2010; McCartney 2021c). Even though this project was implemented a decade ago, and has endured changing governments and allegations of corruption, by 2013 ₹21 crore (USD 275 million) had already been spent on promoting Sanskrit education in Uttarakhand. Regrettably, there is very little to show for it (Singh 2015). Recently, however, an updated policy has increased this imposition of language shift toward the target language, Sanskrit (Ahmad 2020). This new policy aims to

13 Historical linguistics demonstrates that the oldest intangible artefact of the Sanskrit canon, the Ṛgveda (c. 1600-1100 BCE), contains approximately two percent non-Aryan vocabulary, idiomatic expressions and phonemic influences, which are derived from the Dravidian language family, the Bactria-Margiana Archaeological Complex, and the Harappan Kubhā-Vipāś Substrate (see Witzel 2010).

create a Sanskrit village in every 'block' (an administrative division often confused with sub-districts) of Uttarakhand (News Desk 2020; Upadhyay 2020). The state of Uttarakhand consists of two Divisions, 13 Districts, 79 Sub-districts, and 97 Blocks. One wonders how much more investment might be needed to transform 97 villages scattered across the Himalayas into *saṃskṛta-grāmāḥ* ('Sanskrit-villages'). After all this investment not one L1-Sanskrit token comes from any village or rural area.[14] While 70% of the state's total population live in rural areas, 100% of the state's total (n=246) L1-Sanskrit tokens are linked to urban areas.

The ideology behind this top-down development project aims to reverse engineer India and the world through implementing a 'dharmic' lifestyle predicated by Sanskrit and Yoga. It aims to reform society toward an imagined Sanskritland, where just like the following song that attendees at Samskrita Bharati language camps learn, the aspiration is that Sanskrit will be spoken "*gṛhe gṛhe* [in every home], *grāme grāme* [in every village], *nagare nagare* [in every city], and *deśe deśe* [in every country]". This might seem terribly banal and optimistically utopian, yet it is part of a yoga-oriented, faith-based, competitive diplomatic, soft power initiative. Evidence of this includes propositions such that Yoga and Sanskrit can solve climate change (Chauhan 2015; King 2015; Jacobs 2016; United Nations India 2019; Miller 2020). The final aspiration of Samskrita Bharati, however, is evidenced through its road map, which aims to build on simple Sanskrit (*saṃskṛtaṃ saralam*) utterances toward it being spoken all the time (*saṃskṛtaṃ sarveṣām*). This potentially leads to an

14 Sanskrit and Yoga are used to brand the nation (McCartney 2021a). The production of legitimacy and authority in diplomatic and economic arenas involves interweaving narratives involving a product, a place, and a nation (Aroncyzk 2013) through which nations work to control their own images by implementing strategic communication strategies (Ermann, Hermanik 2018). This narrative is only a few clicks away from confirming one's bias. The following quote, from Soumitra Mohan (2020), encapsulates the sentiment around its didactic potential: "The language deserves to be treated much better than it has been so far, more so when it has been called the best 'computerable' language. Sanskrit's credentials to be a language of future India are definitely better and greater than we have realised so far. Its revival will not only renew and revive the pride in our own cultural heritage but will also bring about spiritualism and the concept of a meaningful society and polity, thereby bringing order and peace all across the country, a desideratum for any developed society". A similar sentiment comes from Sampūrṇānanda Saṁskṛta Viśvavidyālaya (SSV)'s homepage (2019) (`https://ssvv.ac.in/brief-history`). SSV is one of India's best-known Sanskrit universities, which is located in Benares, Uttar Pradesh. SSV explains on its homepage that "Sanskrit is the most ancient and perfect among the languages of the World. Its storehouse of knowledge is an unsurpassed and the most invaluable treasure of the world. This language is a symbol of peculiar Indian tradition and thought, which has exhibited full freedom in the search of truth, has shown complete tolerance towards spiritual and other kind of experiences of mankind, and has shown catholicity towards universal truth. This language contains not only a rich fund of knowledge for people of India, but it is also an unparalleled way to acquire knowledge and is thus significant for the whole World".

unavoidable critical mass (*saṃskṛtaṃ anivāryam*), which will inevitably lead to a language shift (*saṃskṛta sarvatra*), and adoption of the new global *lingua franca* (*viśva bhāṣā*) (McCartney 2017e). The perceived net-positive outcome (*abhyudaya*) has India positioned as the next global superpower and paragon of moral virtue, its ultimate dispenser (*viśva-guru*) (Bharati 2014; Singh R.K. 2014; Press Trust India 2018; 2019).

3 Comparing 2001 and 2011 Census Results

Now, it is time for some bookkeeping (*pusta pālana*). This section provides a straightforward analysis of the 2001 and 2011 census results. The Persons category is further disambiguated by two binaries, Rural/Urban and Males/Females. In the 2011 result (n = 24,821), fewer L1-Sanskrit tokens were returned from Rural areas (10,908), as opposed to Urban areas (13,913). This gives a ratio of 44:56. However, Bihar completely reverses this with an 89:11 ratio in favour of rural areas. See **[tab. 2]** for a comparison of the 2011 top ten Rural:Urban states.

Table 2 Rural and Urban Top Ten States, 2011 (Office of the Registrar General 2020)

2011 Urban Top 10		**2011 Rural Top 10**	
Bihar	3,041	Maharashtra	3,555
Rajasthan	1,461	Uttar Pradesh	1,668
Uttar Pradesh	1,394	Madhya Pradesh	1,020
West Bengal	1,066	Karnataka	1,016
Himachal Pradesh	857	Rajasthan	914
Madhya Pradesh	851	Tamil Nadu	743
Goa	385	Gujarat	680
Odisha	306	Goa	670
Jharkhand	264	NCT of Delhi	594
Maharashtra	247	Jharkand	589

The sex difference splits 55:45. A total of 13,636 Male tokens were returned compared to 11,185 Female tokens. Every state/union territory has more Males to Females at a 60:40 average. However, Tamil Nadu has a 50:50 split and Puducherry (Pondicherry) is the only instance where Males are fewer than Females at 29:71 (Office of the Registrar General 2020); table 3 compares the L1-L3 (total) 'Sanskrit speakers' between the 2001 and 2011 censuses **[tab. 3]**. Self-reported L1 speakers increased from 14,135 to 24,821; the L2 figure has stayed almost the same, while the L3 figure has dropped by 48%.

Table 3 2001-2011 Sanskrit L1-L3 (Office of the Registrar General 2020)

	L1	L2	L3
2001	14,135	1,234,931	3,742,223
2011	24,821	1,134,362	1,963,640
% change	43 %	-9 %	-48 %

We can still use these data to locate sub-district administrative zones which have the highest numbers. Table 4 elaborates on the previous table, demonstrating that even though the total number of L1 rose between 2001-11 the total L3 has almost halved **[tab. 4]**. This suggests that the total number of L1-L3 for 2011 decreased by 37%, even though the 2011 L1 increased by 76%.

Table 4 L1-L3 Sanskrit 2001 and 2011 (Office of the Registrar General 2020)

		2001 Total	2001 Total	% Change
	L1	14,135	24,821	43%
	L2	1,234,931	1,134,362	-8%
	L3	3,742,223	1,963,640	-48%
	L1-L3	4,991,289	3,122,823	-37%
Male	L1	8,189	13,636	67%
	L2	875,107	713,772	-18%
	L3	2,751,121	1,266,098	-54%
	L1-L3	3,634,417	1,993,506	-45%
Female	L1	5,946	11,185	47%
	L2	359,824	420,590	14%
	L3	991,102	697,542	-30%
	L1-L3	1,356,872	1,129,317	-17%

Another significant point relevant across the L1-L3 range relates to the relationship between Hindi, English and Sanskrit. The reason the L2-Sanskrit figures are different between table 4 and table 5 is due to table 4 consisting of all the L2-Sanskrit speakers **[tab. 4]**. In contrast, table 5 shows only the L1-Hindi_L2–Sanskrit *and* L1-Hindi_L2-English figures **[tab. 5]**. This figure is significant, as 95% of L2-Sanskrit speakers are L1-Hindi speakers (1,174,019 / 1,234,931).[15] This locates the L2-Sanskrit phenomenon within an exceptionally Hindi-centric con-

15 The Hindi language category consists of 57 sub-languages and dialects (Office of the Registrar General 2020).

text. This is similar to Breton's (1976, 304) observation based on the 1961 census, that "Le centre de la connaissance du sanskrit à notre époque est de loin la plaine gangétique (Uttar Pradesh: 79,000, Bihar: 29,000, Punjab et Haryana: 20,000, Delhi: 9,000)". However, it is important to appreciate that this group only comprises 0.3% of the total number of L1-Hindi speakers. The Male figures for both L2-Sanskrit and L2-English have fallen (24 and 11%) while the Female figures have increased (6 and 19%). The Totals have respectively decreased 15 and 1.2%.

Table 5 L1-Hindi_L2-Sanskrit/English between 2001 and 2011 (Office of the Registrar General 2020)

	L	TOTAL	Male	Female
2001	Sanskrit	1,174,019	830,827	343,192
	English	32,399,287	21,931,407	10,467,880
2011	Sanskrit	994,863	631,099	363,764
	English	32,018,890	19,592,236	12,426,654

Table 6 begins with the 2011 Total Persons figures for L1-Sanskrit **[tab. 6]**. Below, in the next part of the table the L1-Sanskrit_L2-Hindi_English combination equates to 69% of the total L1-Sanskrit_L2-α figure. In the bottom part of the table the combined L1-Sanskrit_L2-α_L3-Hindi_English portion is 77%. These data show the intimate relations that Hindi and English have with Sanskrit.

Table 6 2011 L1-Sanskrit_L2-[Hindi-English]_L3[English-Hindi] (Office of the Registrar General 2020)

	TOTAL	Male	Female
L1-Sanskrit	24,821	13,636	11,185
L2	19,712	11,075	8,637
L2 Hindi	12,221	6,960	5,261
L2 English	1,347	727	620
H+E Total	13,568	7,687	5,881
H+E %	69	69	68
L3	7,910	4,600	3,310
L3 Hindi	2,267	1,327	940
L3 English	3,796	2,211	1,585
H+E Total	6,063	3,538	2,525
H+E %	77	77	76

Table 7 below presents data related to L1-Sanskrit_L2-α [tab. 7]. It shows Hindi's clear L1-Sanskrit_L2-α majority with 62% of the L2 category.[16] While Hindi is one of India's official languages (alongside English, which also ranks high), the data show the intimate relation with the Hindi heartland. Notice that Female L2-Odia tokens were higher than Male (180:132), as well as Konkani (39:37), and Others (83:76).

Table 7 L1-Sanskrit/L2-α rankings, 2011 (Office of the Registrar General 2020)

L2	TOTAL	Male	Female
Hindi	12,221	6,960	5,261
Marathi	1,934	1,071	863
English	1,347	727	620
Bengali	900	462	438
Kannada	839	502	337
Tamil	551	281	270
Gujarati	365	254	111
Urdu	325	193	132
Odia	318	138	180
Telegu	271	150	121
Malayalam	163	86	77
Konkani	159	76	83
Others	76	37	39

Table 8 shows the fluctuations between the dominant Sanskrit states by comparing the results from the 2011 and 2001 censuses [tab. 8]. What might explain Uttar Pradesh's fewer speakers and Maharashtra's increase?

Table 8 2011 and 2001 State-level L1-Sanskrit Census Results (Office of the Registrar General 2020)

	2001	2011	% change
Maharashtra	408	3,802	832
Bihar	754	3,388	349
Uttar Pradesh	7,048	3,062	-57
Rajasthan	989	2,375	140
Madhya Pradesh	381	883	132
Karnataka	830	1,218	47

16 Note: this list is only a portion of the total.

Did all the speakers from Uttar Pradesh move to Mumbai? How can the dramatic rise in states like Rajasthan, Madhya Pradesh, and Karnataka be explained? These data show that the Sanskrit-speaking sentiment predominates in the Hindi belt. Having provided some national level statistics, what follows is a closer look at these top-ranking states, pinpointing L1-L3 Sanskrit tokens down to district and sub-district administrative levels. This begins in Maharashtra.

4 An Explication of Census Data in Several States

4.1 Maharashtra

Table 9 shows the top ten districts in the state **[tab. 9]**. Making sense of all the data is challenging because there is an overwhelming amount of overlap between tables. For ease of comprehension, table 9 is a modified version of the original. Several columns have been deleted.[17] Respectively, both Pune District (28%) and Pune City Sub-district (22%) represent the highest L1-Sanskrit administrative zones in their category for the state. It is clear from the Urban 3554:Rural 145 ratio that 96% of Maharashtra's Sanskrit tokens are located in urban areas.

Table 9 2011 Maharashtra C-16 Sanskrit–Top 10 Districts (Office of the Registrar General 2020)

C-16 POPULATION BY MOTHER TONGUE
State Code – 27 / Mother Tongue Code – 17 / Mother Tongue Name – Sanskrit

District code	Area name	Total			Rural			Urban		
		P	M	F	P	M	F	P	M	F
000	MAHARASHTRA	3,802	1,984	1,818	247	145	102	3,555	1,839	1,716
521	Pune	1,091	567	524	51	33	18	1,040	534	506

17 As an example, before the token numbers are given, on the right of each table, the complete line for Pune District would read, C0116 (table code, denotes that these data are from the C-16 'mother tongue' table), 027 (state code), 521 (district code), 0000 (sub-district code), Pune (area name), 017000 (mother tongue code), Sanskrit (mother tongue name). This would look like the following, C0116-027-521-0000-Pune-017000-Sanskrit. Where it gets confusing is when districts, sub-districts, and towns have the same or similar names and coding. For example, the district of Pune (27-521-00000) is similar to its sub-district constituent, Pune City (27-521-04194). The same logic applies to every other district/sub-district across the entire suite of census tables. Another example is Nashik District (27-516-00000), which has a similarly named sub-district, Nashik (27-516-04152). This raises some methodological issues. We can use these fields to find all sub-districts in each state and rank them. We can also filter just the districts and rank them in the same way.

C-16 POPULATION BY MOTHER TONGUE State Code – 27 / Mother Tongue Code – 17 / Mother Tongue Name – Sanskrit										
District code	**Area name**	**Total**			**Rural**			**Urban**		
517	Thane	710	385	325	14	9	5	696	376	320
518	Mumbai Suburban	556	263	293	0	0	0	556	263	293
516	Nashik	442	239	203	21	12	9	421	227	194
505	Nagpur	195	98	97	6	6	0	189	92	97
519	Mumbai	109	54	55	0	0	0	109	54	55
511	Nanded	61	30	31	5	2	3	56	28	28
515	Aurangabad	55	29	26	5	3	2	50	26	24
514	Jalna	51	27	24	47	24	23	4	3	1
504	Wardha	42	21	21	22	10	12	20	11	9

Table 10 shows the L1-L3 relations found in the C-17 tables **[tab. 10]**. In particular, the relations between the state language, Marathi, with Hindi, English, and Sanskrit are located. While Sanskrit clusters with Hindi and English, at the state level it becomes a bit more complicated since it is important to consider the relevant state language in this clustering. Hindi and English, however, still have an overwhelming presence. The left column shows L1-Marathi at the national level,[18] the middle column shows how many of these returned L2-Hindi, while the right column shows the figures for L3-Sanskrit. This repeats for each line demonstrating the L1-Marathi_L2-α_L3-β scenarios. Using the same formula, the second part of the table zooms in to the state level figures. L1-Marathi_L2-Hindi_L3-α equates to 42% of the L2 category. The final part of this table focuses on L1-Sanskrit_L2-α_L3-β.

Table 10 2011 Maharashtra C-17 Bi-/Trilingualism (Office of the Registrar General 2020)

Marathi across India		
L1	**L2**	**L3**
Marathi 83,026,680	Hindi 34,650,142	Sanskrit 57,070
	Kannada 1,468,221	Sanskrit 559
	English 1,395,659	Hindi 870,985
		Kannada 26,907
		Sanskrit 13,142
	Gujarati 361,327	Sanskrit 238
	Khandeshi 292,555	Sanskrit 124
	Konkani 98,318	Sanskrit 168

18 Notice C17-0000_2011 compared to C17-2700 in the middle and right panels, respectively. The state code for Maharashtra is 2700 and the national code is 0000.

	Sanskrit 56,360	Hindi 22,477
		English 19,780
L1-Marathi in Maharashtra		
L1	L2	L3
Marathi77,461,172	Hindi 32,660,911	Sanskrit 54,500
	English 1,202,810	Sanskrit 12,673
	Kannada 512,117	Sanskrit 471
	Khandeshi 292,035	Sanskrit 124
	Telegu 150,921	Hindi 31,881
		Sanskrit 38
	Gujarati 60,552	Sanskrit 142
	Sanskrit 54,100	Hindi 21,474
		English 19,415
L1-Sanskrit in Maharashtra		
L1	L2	L3
Sanskrit3,802	Hindi 1,782	English 888
		Marathi 365
	Marathi 1,259	Hindi 598
		English 306
	English 334	Marathi 49

This demonstrates that the L1-SOL_L2_L3 cluster with Hindi_English is a predictor for the relative number of L2 and L3 Sanskrit speakers. This is consistent across every national and state-level enumeration. If we reverse the order (right panel) and begin with L1-Sanskrit, the numbers for L2_L3 also follow a similar order, with the State Official Language (SOL), Hindi, and English in superior positions compared to other possibilities.

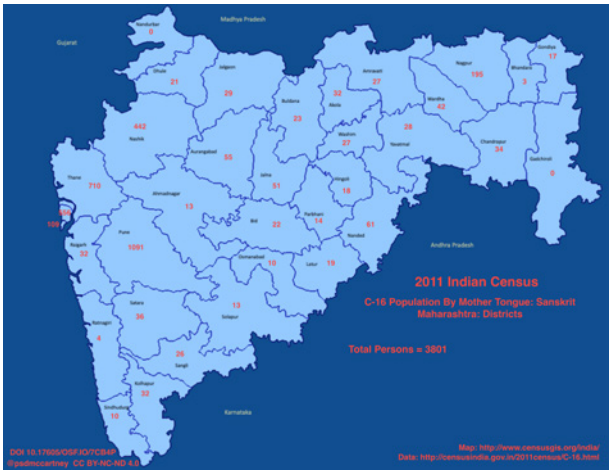


Figure 2 2011 Maharashtra: Districts L1-Sanskrit. (India 2020)

Figure 2 demonstrates the 2011 District-level L1-Sanskrit figures for Maharashtra **[fig. 2]**. The top five districts are Pune 1091, Thane 710, Mumbai Suburban 555, Nashik 442, and Nagpur 195.

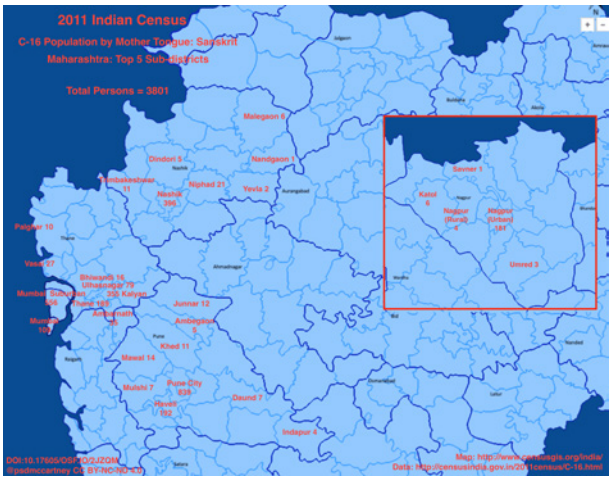


Figure 3 2011 Maharashtra: Sub-districts L1-Sanskrit. (India 2020)

Figure 3 zooms down to the sub-district level to show the top five districts in more detail **[fig. 3]**. Notice how Nagpur District has been superimposed in the red box. Next, we move onto the state of Bihar.

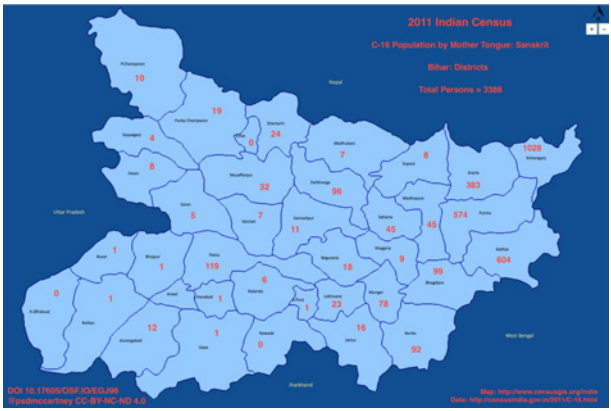
4.2 Bihar

In Bihar, the highest number of L1-Sanskrit tokens are located in the eastern administrative area of Kishanganj District (210). The top five districts are listed in **[tab. 11]**.

Table 11 2011 Top 5 Districts: Bihar (Office of the Registrar General 2020)

C-16 POPULATION BY MOTHER TONGUE **State Code – 10 / Mother Tongue Code – 17 / Mother Tongue Name – Sanskrit**										
District code	**Area name**	**Total**			**Rural**			**Urban**		
		P	M	F	P	M	F	P	M	F
000	BIHAR	3,388	1,845	1543	3,041	1,654	1387	347	191	156
210	Kishanganj	1,028	543	485	962	507	455	66	36	30
212	Katihar	604	324	280	598	321	277	6	3	3
211	Purnia	574	326	248	544	309	235	30	17	13
209	Araria	383	220	163	381	219	162	2	1	1
230	Patna	119	67	52	4	3	1	115	64	51

From within these districts, Bihar's top ten sub-districts are Dighalbank 558 (Kishanganj District), Terhagachh 307 (Kishanganj District), Baisi 246 (Purnia District), Azamnagar 154 (Katihar District), Sikti 120 (Araria District), Dagarua 108 (Purnia District), Palasi 105 (Araria District), Falka 92 (Katihar District), Thakurganj 83 (Kishanganj District), and Patna Rural 82 (Patna District). In **[figs 4-5]** notice the clustering in the far east of the state across four districts.



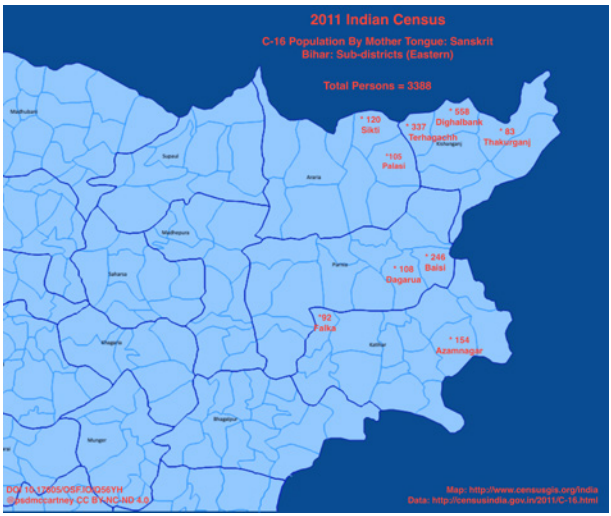


Figure 4 2011 Bihar: Districts L1-Sanskrit (India 2020)

Figure 5 2011 Bihar: Sub-districts L1-Sanskrit (India 2020)

Unlike Maharashtra, 90% of Bihar's L1-Sanskrit tokens are Rural. Where might they be located? The 'village-level' C-16 tables are not available. One cumbersome triangulation method requires scrutinis-

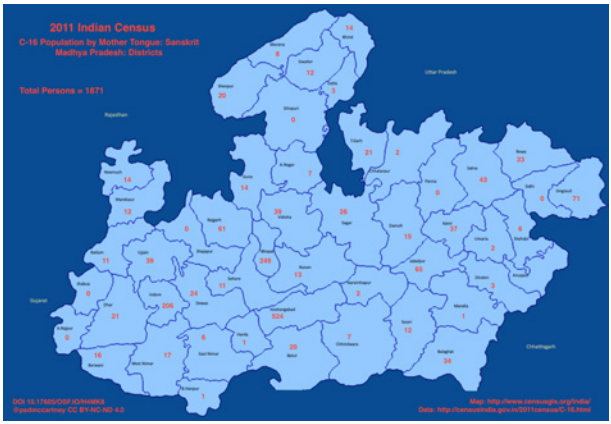
ing other tables. The state's highest returning district is Kishanganj. Beginning with the 'town level' C-16 table for Bihar, we learn that Kishanganj Nagar lists 66 Urban L1-Sanskrit tokens (Office of the Registrar General 2020). We can combine this with the Village Directory of Kishanganj District, Bihar (Office of the Registrar General 2020).[19] The relevant code for Dighalbank's Block is 002. We can sort and filter all the listed towns and villages with this code and then sort further for all the places with a population under 5,000 inhabitants (the town/village population boundary). While we currently are unable to locate with more accuracy, we are, nonetheless, left with a list of 'rural villages' from which Bihar's highest L1-Sanskrit sub-district is potentially comprised of. Let us move on to Madhya Pradesh.

4.3 Madhya Pradesh

Madhya Pradesh returned more L1-Sanskrit Urban tokens (54%) [Rural 851 (451 M/400 F) and Urban 1020 (537 M/483 F)]. This further complicates the fabled 'Sanskrit-speaking' village narrative. However, the highest-ranking sub-district, Pipariya, accounts for 96% (Rural 485/Urban 5) of Hoshangabad District's total 524 (496 Rural/28 Urban) (Office of the Registrar General 2020). Compared to the state Rural total, Pipariya Sub-district equates to 57% of all the 367 sub-districts in Madhya Pradesh and 26% of the state's total. At the village level this can be narrowed down to 114 villages in Pipariya Sub-district (Office of the Registrar General 2020).

If this sub-district does have such a high number of Sanskrit speakers it is certainly unclear as to why these places are not as famous as the three so-called 'Sanskrit-speaking' villages (Jhiri, Sarangpur Sub-district, Rajgarh District; Mohad, Gadawara Sub-district, Narsimhapur District; and Baghuwar, Kareli Sub-district, Narsimhapur District). What is more curious is the fact that the districts these three 'Sanskrit-speaking' villages are located in barely returned any L1-Sanskrit tokens. The internet contains countless sites that assert that everyone, or nearly everyone, in these villages speaks Sanskrit. The district and sub-district L1-Sanskrit tokens are represented in **[figs 6-7]**.

19 Kishanganj District is comprised of seven 'blocks', which is the same number of sub-districts. The difference between a block and a sub-district is its function. A block is a geographical unit for rural development, whereas a sub-district (*tehsil*) is a geographical unit for revenue collection (Maheshwari 1984).



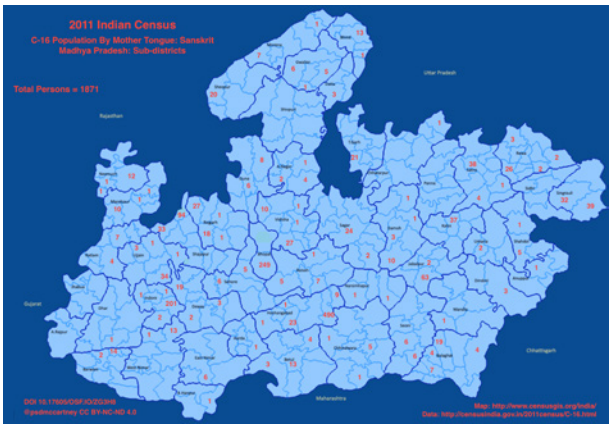


Figure 6 2011 Madhya Pradesh: Districts L1-Sanskrit. (India 2020)

Figure 7 2011 Madhya Pradesh: Sub-districts L1-Sanskrit. (India 2020)

Like Bihar, filtering through the village/town lists for Pipariya Sub-district the towns with 5,000+ inhabitants can be removed. The result is 142 villages, ranging in size from 4,767 (Khapar Kheda) down to two (Pathi Thekredri) inhabitants. It is this list that the majority of Madhya Pradesh's Sanskrit tokens come from (Office of the Registrar General 2020).

The ST-15 (Scheduled Tribes) table is a subset of the C-16 tables relating to Scheduled Tribes' mother tongues. These are available at the state/district levels. 6.7% of the L1-Sanskrit state total (149/1871) is comprised of people who identify as members of Scheduled Tribes. Of this, 126 tokens are from Hoshangabad District (Office of the Registrar General 2020).

Table 12 2011 and 2001 Madhya Pradesh L1-Sanskrit (Office of the Registrar General 2020)

Total Persons – Sanskrit	L1	L2	L3
2011	1,871	246,940	454,245
2001	381	210,400	960,176
% change	391	17	-53

Table 12 scratches beneath the surface to show the changes between 2001 and 2011 across the L1-L3-Sanskrit range **[tab. 12]**. This figure reflects the national data. As we have discussed, the L1 change is aspirational. It really makes not much sense, otherwise. How could a 391% increase have occurred in just one decade? The L2 figure is more feasible, however the L3 decline of 53% is a worrying predictor for the vitality of Sanskrit. Could it be that all the Sanskrit village mythology circulating might be harming Sanskrit's role? In relation to the Hindi/English/Sanskrit trinity, in Madhya Pradesh L1-H_L2-E_L3-S comprises 98% of all the L3-Sanskrit possibilities. This decline is also considerable among the Scheduled Tribes of Madhya Pradesh. Table 13 highlights the changes of L1-Bhili/Bhilodi_ and L1-Gondi_L2-α_L3-Sanskrit **[tab. 13]**. Both have declined dramatically between 2001 and 2011.

Table 13 2011 and 2001 Madhya Pradesh L1-ST_L2-α_L3-Sanskrit (Office of the Registrar General 2020)

	2011	2001	% change
L1-Bhili/Bhilodi_L2-α_L3-Sanskrit	779	1,677	-54
L1-Gondi_L2-α_L3-Sanskrit	661	889	-26

Madhya Pradesh's 2011 L2-Sanskrit total is 246,940. This is comprised of the Scheduled Tribes' total of 13,540 (5% of state total). This 13,540 is predominated by 84% Rural tokens. Of this Scheduled Tribes' total 52% come from a group of 60 ST-L1 languages, including Gond, Arakh, and Agaria (prevalent in Hoshangabad District). The next group of languages include Bhil and Bhilala (15%), and Kol (10%) (Office of the Registrar General 2020). Having located the districts and sub-districts within Madhya Pradesh where L1-L3 Sanskrit tokens were returned, a clearer image of where future fieldwork could be directed emerges. The final state is Uttar Pradesh.

4.4 Uttar Pradesh

Uttar Pradesh is India's most populous state with 199,581,477 people (Office of the Registrar General 2020). While Uttar Pradesh experienced a 57% decrease in L1-Sanskrit tokens, in comparison, the total nationwide increase for L1-Sanskrit is 76% (Government of Office of the Registrar General 2020). What has caused Uttar Pradesh to decrease when other states witnessed large increases? Three districts are worthy of closer scrutiny because of how they changed dramatically between 2001 and 2011. The state capital, Lucknow, reduced by -82%, while Unnao reduced by -95%, and Gorakhpur, the electorate of UP's Chief Minister, Yogi Adityanath, decreased by -99% (Office of the Registrar General 2020). The top ten districts are Kanpur Nagar 932, Sitapur 722, Sultanpur 323, Ghaziabad 128, Saharanpur 85, Ballia 79, Lucknow 55, Varanasi 55, Bijnor 50, and Agra 41 (Office of the Registrar General 2020); figure 8 provides the total Person numbers for each district **[fig. 8]**.

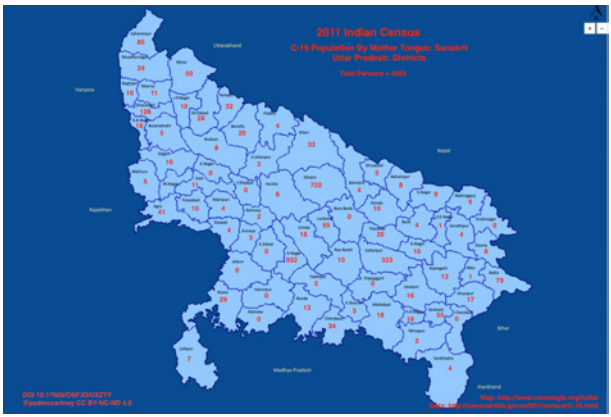


Figure 8 2011 Uttar Pradesh: Districts L1-Sanskrit. (India 2020)

Figure 8 shows the numbers for each district in which L1-Sanskrit tokens were returned. In Uttar Pradesh, 1,697 Male L1-Sanskrit tokens were returned compared to 1,365 Female. Like Madhya Pradesh and Maharashtra, the Urban L1-Sanskrit (1668) token count is slightly higher than the Rural equivalent (1394). The urban area of Varanasi (Benares), which is considered one of the holiest cities of Hinduism, which is home to the famous Sanskrit university, Sampūrṇānanda Saṁskṛta Viśvavidyālaya (SSV), only returned 54 tokens (36 M/18 F). Similarly, Breton (1976, 306) wonders why the famous traditional seats of Sanskrit learning have fewer speakers of Sanskrit. Notice that Kanpur is the city with the highest number of returned

L1-Sanskrit tokens (Office of the Registrar General 2020). The top ten sub-districts are Kanpur 931, Mahmudabad 482, Lambhua 310, Ghaziabad 127, Sidhauli 94, Biswan 83, Ballia 62, Varanasi 55, Deoband 51, and Bijnor 50 (Office of the Registrar General 2020). Kanpur is also the highest ranked town. Zooming in, figure 9 shows the unique L1-Sanskrit results for each of the six tehsils in Sitapur District **[fig. 9]**.

Figure 9 2011 Sitapur District within Uttar Pradesh. (India 2020)

In 2001, Sitapur District, Uttar Pradesh (430 kilometres east of New Delhi), reportedly had the highest number of all the districts in the country, at 558 (Priyanka 2014). The 2011 total of 722 (M 378/F 344) is a 28% increase. Sitapur District also represents 24% of the state's total. It splits Rural 1275 (M 676/F 599) and Urban 73 (M 33/F 40). This means that the district of Sitapur provides 46% of Uttar Pradesh's L1-Sanskrit tokens, which, itself, is comprised of 90% Rural from across the district's six tehsils (Office of the Registrar General 2020). When compared with the 2001 results for the same the district the fluctuations are quite remarkable. Where did 2,604 Sanskrit 'speakers' from Biswan tehsil go? Sitapur District has the second highest L1-Sanskrit number in Uttar Pradesh.

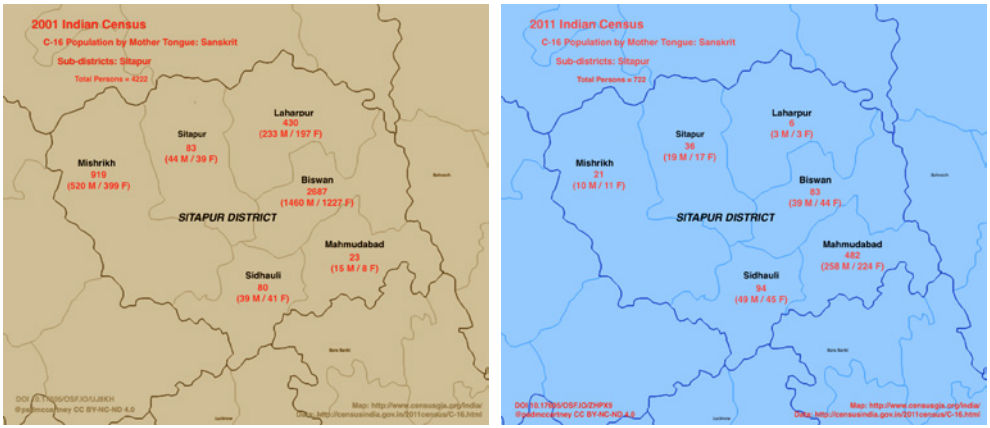


Figure 10 2001 and 2011 Sitapur District: Sub-districts L1-Sanskrit. (India 2020)

Figure 10 shows the 2011 figures for each tehsil of Sitapur District, which indicates that Mahmudabad Tehsil has the most with 482 (M 258/F 224) **[fig. 10]**; table 14 displays the top five sub-districts **[tab. 14]**. Three of the list are located in Sitapur Tehsil and Biswan Town is the sixth-highest ranked in the state, as well as the fifth-highest ranked tehsil. Both Biswan and Sidhauli tehsils return more Female than Male tokens. While Lambua returns the same number for both sexes.

Table 14 2011 Top 5 Rural Tehsils (Office of the Registrar General 2020)

Tehsil	Total	Male	Female
Mahmudabad	482	258	224
Lambhua	310	155	155
Sidhauli	94	49	45
Ballia	62	32	30
Biswan	83	39	44

In contrast to Kanpur, Mahmudabad Tehsil (Rural) is the second highest tehsil in the state with 482 (258 M/224 F). If all the L1-Sanskrit 'speakers' do not live in towns then is it safe to assume they live in villages? Mahmudabad Tehsil has 341 villages (Office of the Registrar General 2020). The question is, in which villages are L1-Sanskrit 'speakers' living?

Table 15 2011 Religious-Political affiliations and Literacy-Sex ratio comparisons (Office of the Registrar General 2020)

State	Tehsil name	Population	Literacy	Sex ratio	Political preference	Hindu %
MP	Pipariya	181, 261	63.8	896	BJP	95
MP	Sarangpur	186,082	56.4	950	BJP	84
UP	Kanpur	3,470,334	72.8	860	BJP	83
MP	Indore	2,389,511	73.6	925	BJP	74
UP	Mahmudabad	596,252	47.1	884	Samajawada	74
MP	Huzur	2,107,523	72	920	BJP	72

Finally, table 15 lists some of the main tehsils mentioned above **[tab. 15]**. They are ranked, first, by the percentage-age of Hindus in each tehsil. Most display a preference for the Bharatiya Janata Party (BJP), regardless of the size of the Hindu majority. The curious thing is that some of the highest L1-Sanskrit tehsils in Madhya Pradesh and Uttar Pradesh return literacy and sex ratios well below national and state averages. What does this tell us about the ability of Sanskrit to 'transform lives' and the development grand narrative it serves?

5 Concluding Remarks

This paper compares the 2001 and 2011 Sanskrit census results paying closer attention to the top-ranking states Maharashtra, Bihar, Madhya Pradesh and Uttar Pradesh. However, the results of this study are unexpected. Further analysis across the remaining states will be released in forthcoming articles. Still, we are able to determine that Maharashtra returned the highest number for L1-Sanskrit; that most of the L1-Sanskrit speakers live in urban areas; that the 'Sanskrit-speaking' village is an aspirational myth not borne out in the government data; that more men claim to speak Sanskrit; that the overwhelming majority also speak Hindi and English, regardless of L1-L3 combinations; that the L1-Hindi/L2-Sanskrit combination amounts for 92% of the total L2-Sanskrit speakers; and that, regardless of the rhetorical and ideological bluster, Sanskrit continues to fall short of its alleged capacity for attaining the #Sanskrit4ClimateAction goals related to key indicators, such as sex, literacy, and health development.

Sanskrit, alongside Yoga, is a key instrument for branding the nation. This is a domestic as well as international project that has a symbiotic relation to the global wellness and leisure markets, which is fertile ground for cultivating banal nationalism. From a language acquisition perspective, the media's preference to provide pithy and inaccurate data from the census seems counterproductive, if not misplaced. While

moods certainly lift with hearing the L1 level rose between 2001 to 2011, the more interesting categories relate to L2 and L3 levels, both of which have reduced, however the L3 level fell by almost 90%.

With this finer-grained analysis, a clearer map of where people who have an affinity to identify as L1-L3-Sanskrit speakers were located at the last census. Regardless of whether they do in fact speak Sanskrit, these data will aid future research related to in-country field work, enabling strategic sorties down to the sub-district tehsil level. Finally, the 2021 Indian census will the first digitised census the nation will embark on. Hopefully, this allows for data to be enumerated, rationalised, and published much sooner than the seven-year lag that occurred at the last census. Unlike the botched 1941 census, which first introduced self-reporting of data, it is anticipated that this new age of demography in India will not suffer the same fate. From a linguistic perspective, building on demographic data potentiates future exploration of various aspects of language revitalisation and second-language acquisition through in-depth study of particular linguistic features related to substrate interference and imperfect learning of the target language.

Bibliography

Ahmad, R. (2020). "Why Must the Uttarakhand Govt Undermine Urdu to Promote Sanskrit?". *The Wire*. https://thewire.in/communalism/sanskrit-urdu-uttarakhand.

Aklujkar, A. (1996). "The Early History of Sanskrit as Supreme Language". Houben, J.E. (ed.), *Ideology and Status of Sanskrit. Contributions to the History of the Sanskrit Language*. Leiden: Brill, 59-85.

Amaravāṇī (2020). "Amaravāṇi. Saṃskṛtam for All… forever". https://amaravaani.org/visvabha%E1%B9%A3a-sa%E1%B9%83sk%E1%B9%9Btam.

Aralikatti, R.N. (1989). *Spoken Sanskrit in India. A Study of Sentence Patterns*. Tirupati: Kendriya Sanskrit Vidyapeetha.

Aralikatti, R.N. (1991). "A Note on the Word Order in Modern Spoken Sanskrit and Some Positive Constraints". Hock, H.H. (ed.), *Studies in Sanskrit Syntax*. Delhi: Motilal Banarsidas, 13-17.

Arel, D. (2004). "Language Categories in Censuses. Backward- or Forward-looking?". Kertzer, D.I.; Arel, D. (eds), *Census and Identity. The Politics of Race, Ethnicity, and Language in National Censuses*. Cambridge: Cambridge University Press, 92-120. https://doi.org/10.1017/cbo9780511606045.005.

Aroncyzk, M. (2013). *Branding the Nation. The Global Business of National Branding*. New York: Oxford University Press.

Azad, M.A.K. (1957). *Report of the Sanskrit Commission 1956-1957*. Ministry of Education and Culture. Poona: Government of India. http://www.wilbourhall.org/sansknet/general/reportonsanskritcommission/S_comm2.htm.

Babu, H. (2017). "Breaking the Chaturvarna System of Languages The Need to Overhaul the Language Policy". *Economic and Political Weekly*, 7(23), 112-19.

https://m.epw.in/journal/2017/23/special-articles/breaking-chaturvarna-system-languages.html.

Baines, J.A. (1893). *General Report of the Census of India, 1891*. London: Eyre & Spottiswoode.

Bedewi, J. (2020). "10 Sanskrit Phrases You Often Hear in a Yoga Class (& What They Mean)". *TheTalko*. https://www.thetalko.com/sanskrit-phrases-hear-yoga-class-meaning.

Bharati, S. "Our Mission & Vision". https://samskritabharatiusa.org/index.php/about/our-mission-vision.

Breton, R.J.-L. (1976). *Atlas géographique de langues et des ethnies de l'Inde et du Subcontinent. Bangladesh, Pakistan, Sri Lanka, Népal, Bhoutan, Sikkim*. Québec: Les Presses de l'Université Laval.

Burrow, T. (1965). *The Sanskrit Language*. London: Faber & Faber.

Chatterjee, M. (2020). "The Evolution of the Yajurvedic Words in Yaska's View". *International Journal of Sanskrit Research*, 6(3), 93-5.

Chauhan, R. (2015). "Modi and Buddhism. Between Cultural and Faith-Based Diplomacy". *Observer Research Foundation*, 77, 1-18.

Dasgupta, P. (1995). "Sanskrit and Indian English. Some Linguistic Considerations". Singh, R.; Dasgupta, P.; Lee, J.K. (eds), *Explorations in Indian Sociolinguistics. Language and Development*. New Delhi: Sage Publications, 160-75.

Deopujari, S. (2009). "Sanskrit Can Become the Language of the Masses in Rural Areas". *Organiser*. https://organiser.org/2009/09/06/41332/uncategorized/rc0668762/.

Deshpande, M.M. (2011). "Efforts to Vernacularize Sanskrit. Degree of Success and Failure". Fishman, J. (ed.), *Handbook of Language and Ethnic Identity. The Success-Failure Continuum in Language and Ethnic Identity Efforts*. Oxford: Oxford University Press, 218-19.

Deshpande, M.M. (2013). "Aryan Origins. Brief History of Linguistic Arguments". Thapar, R. (ed.), *India. Historical Beginnings and the Concept of the Aryan*. Delhi: National Book Trust, 98-156.

Deshpande, M.M. (unpublished). "Scope of Early Sanskrit Usage. A Wider Approach". Paper presented at the 17th World Sanskrit Conference (Vancouver, 9-13 July 2018).

Deumert, A. (2009). "Tracking the Demographics of (Urban) Language Shift. An Analysis of South African Census Data". *Journal of Multilingual & Multicultural Development*, 31(1), 13-35. https://doi.org/10.1080/01434630903215125.

Dizon, J. (2016). "Remote Indian Village Leading Sanskrit Language Revival". *World Yoga News*. https://web.archive.org/web/20161130174015/http://www.worldyoganews.com:80/remote-indian-village-leading-sanskrit-language-revival.

Duchêne, A.; Humbert, P.N.; Coray, R. (2018). "How to Ask Questions on Language? Ideological Struggles in the Making of a State Survey." *International Journal of the Sociology of Language*, 252, 45-72. https://doi.org/10.1515/ijsl-2018-0014.

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

Ermann, U.; Hermanik, K.-J. (2018). *Branding the Nation, the Place, the Product*. London; New York: Routledge.

Fahy, J. (2020). *Becoming Vaishnava in an Ideal Vedic City*. New York: Berghahn.

Foucault, M. (2007). *Security, Territory, Population. Lectures at the Collège de France, 1977-1978*. Transl. by G. Burchell. Ed. by M. Senellart, F. Ewald, A. Fontana. London: Palgrave Macmillan.

Gait, E.A. (1913). *Census of India, 1911*. Vol. 1, Part 2, *Tables*. Calcutta: Superintendent of Government Printing.

Ghosh, A. (2008). "Sanskrit Boulevard". *Hindustan Times*. https://www.hindustantimes.com/india/sanskrit-boulevard/story-u8M0ylD1druUeSPAd404hN.html.

Goswami, H. (2012). "Mother Language 'Sanskrit' Needs Urgent Protection". *GoI Monitor*. http://goimonitor.com/story/mother-language-sanskrit-needs-urgent-protection.

Hastings, A.M. (2004). *Past Perfect, Future Perfect. Sanskrit Revival and the Hindu Nation in Contemporary India* [PhD dissertation]. Chicago: University of Chicago.

Hastings, A.M. (2008). "Licked by the Mother Tongue. Imagining Everyday Sanskrit at the Home and in the World". *Journal of Linguistic Anthropology*, 18(1), 24-45.

Haynes, J. (ed.) (2021). *Handbook on Religion and International Relations*. Cheltenham (UK): Edward Elgar.

Hebden, K. (2011). *Dalit Theology and Christian Anarchism*. Farnham (UK): Ashgate.

Hindutvainus (2011). "Sanskrit Village. Jhiri, Madhya Pradesh". *YouTube*. https://www.youtube.com/watch?v=lHLIy-WHDew.

Hock, H.H. (1991). "Review. R.N. Aralikatti. Spoken Sanskrit in India. A Study of Sentence Patterns". *Studies in the Linguistic Sciences*, 21(1), 161-5.

Hock, H.H.; Pandaripande, R. (1976). "The Sociolinguistics Position of Sanskrit in Pre-Muslim South Asia". *Studies in Language Learning*, 1(2), 106-38.

Hutton, J.H. (1933). *Census of India, 1931*. Vol. 1, *India*, Part 2, *Imperial Tables*. Delhi: Manager of Publications.

India Education Diary Bureau (2020). "Sanskrit Is a Gift of India for Entire Humanity. Ramesh Pokhriyal 'Nishank'". *India Education Diary*. https://indiaeducationdiary.in/sanskrit-is-a-gift-of-india-for-entire-humanity-ramesh-pokhriyal-nishank.

Indian Eagle (2020). "Sanskrit. Language of Knowledge, Love and Peace in India". *Indian Eagle*. https://www.indianeagle.com/travelbeats/facts-of-sanskrit-in-india.

Indian Yug (2020). "Amazing Facts About the Sanskrit Language that You Didn't Know". *Indian Yug*. https://indianyug.com/amazing-facts-sanskrit-language/#comment-38307.

Jacobs, S. (2016). *The Art of Living. Spirituality and Wellbeing in the Global Context*. London: Routledge.

Jaffrelot, Ch. (2017). "From Holy Sites to Web Sites. Hindu Nationalism, from Sacred Territory to Diasporic Ethnicity". Michael, P.; Possamai, A.; Turner, B.S. (eds), *Religions, Nations, and Transnationalism in Multiple Modernities*. New York: Palgrave Macmillan, 154-74.

Jha, M.K. (2017). *Language, Politics and Public Sphere in North India: Making of the Maithili Movement*. New Delhi: Oxford University Press.

King, E. (2015). "Narendra Modi. Yoga can Help the Fight Against Climate Change". *Climate Home News*. https://bit.ly/3l65OrB.

Kavi, M.R. (1934). *Nāṭyaśāstra. With the Commentary of Abhinavagupta*. Baroda: Oriental Institute.

Mallikarjun, B. (2002). "Mother Tongues of India According to the 1961 Census". *Language in India*, 2. http://www.languageinindia.com/aug2002/indianmothertongues1961aug2002.html.

Maheshwari, S. (1984). "Administering the Planning System". *The Indian Journal of Public Administration*, 30(3), 603-12.

Marten, J.T. (1923). *Census of India, 1921*. Vol. 1, *India*, Part 2, *Tables*. Calcutta: Superintendent Government Printing.

McCartney, P. (2011). "Spoken Sanskrit in a Gujarat Ashram". *The Journal of the Oriental Society of Australia*, 43, 61-82.

McCartney, P. (2014a). "Ek Din Hamaare Ashram mein. A Day in Our Ashram". *YouTube*. https://www.youtube.com/watch?v=xZHJVkhVBPc.

McCartney, P. (2014b). "The Sanitising Power of Spoken Sanskrit". *Himāl Southasian*. http://himalmag.com/sanitising-power-spoken-sanskrit.

McCartney, P. (2015). "Episode 3: Jalam ≠ Chillum". *YouTube*. https://www.youtube.com/watch?v=cLZeuCT_mwQ.

McCartney, P. (2016a). "Episode 1: Imagining Sanskrit Land". *YouTube*. https://www.youtube.com/watch?v=jMi7tkPBbJ4.

McCartney, P. (2016b). "Episode 2: Total-am". *YouTube*. https://www.youtube.com/watch?v=7tAp8m9RHPU.

McCartney, P. (2016c). "Episode 4: It's Time to get Married". *YouTube*. https://www.youtube.com/watch?v=_B3un7aHEAc.

McCartney, P. (2017a). "Suggesting *Śāntarasa* in Shanti Mandir's *Satsaṅga*. Ritual, Performativity and Ethnography in Yogaland". *Ethnologia Actualis*, 17(2), 81-122. https://doi.org/10.2478/eas-2018-0005.

McCartney, P. (2017b). "Jhirī. A 'Sanskrit-Speaking' Village in Madhya Pradesh". *Journal of South Asian Languages and Linguistics*, 4(2), 167-209. https://doi.org/10.1515/jsall-2017-0007.

McCartney, P. (2017c). "Speaking of the Little Traditions. Agency and Imposition in 'Sanskrit-Speaking' Villages in North India". den Boer, L.; Cuneo, D. (eds), *Puṣpikā. Tracing Ancient India Through Text and Traditions*. Philadelphia: Oxbow Books, 62-88.

McCartney, P. (2017d). "Imagining Sanskrit Land". *UTNE Reader*. https://www.utne.com/politics/imagining-sanskrit-land-zm0z17fzcmar.

McCartney, P. (2017e). "Yoga Practitioners and the Unspoken Global Aspirations of Indian Ethno-Nationalism". *Political Theology Today*. https://bit.ly/3sxRMDx.

McCartney, P. (2017f). "Is the Global Prestige of Sanskrit Aiding and Abetting Hindu Nationalists and Supremacists?". *Political Theology Today*. https://bit.ly/3MhYzcb.

McCartney, P. (2017g). "Politics beyond the Yoga Mat. Yoga Fundamentalism and the 'Vedic Way of Life'". *Global Ethnographic*, 4, 1-18. https://oicd.net/ge/wp-content/uploads/Politics-byond-the-Yoga-Mat-P.-McCartney.pdf.

McCartney, P. (2018a). "Śāntamūrti. The Legitimate Disposition(s) of the 'Temple of Peace' Social Network". *Bulletin of the Nanzan Institute for Religion and Culture*, 8, 65-104. https://doi.org/10.17605/OSF.IO/4MP3K.

McCartney, P. (2018b). "Notes from the Field. Reflections on the Imagining Sanskrit Land Project". *Global Ethnographic*, 4, 2-9. https://bit.ly/3la6Gf8.

McCartney, P. (2019a). "Spiritual Bypass and Entanglement in Yogaland (योगस्तान). How Neoliberalism, Soft Hindutva and Banal Nationalism Facil-

itate Yoga Fundamentalism". *Politics and Religion Journal*, 13(1), 137-75. https://doi.org/10.54561/prj1301137m.

McCartney, P. (2019b). "When Cows Can Speak Sanskrit. Yoga, Cultural Revival, and Animal Testing Ethics". *Medium*. https://psdmccartney.medium.com/when-cows-can-speak-sanskrit-yoga-cultural-revival-and-animal-testing-ethics-22afb0c7ae86.

McCartney, P. (unpublished). "Sustainably-Speaking Yoga. Comparing Sanskrit in the 2001 and 2011 Indian Censuses". Paper presented at *The Conference on Asian Linguistic Anthropology* (Siem Reap, Cambodia, 23-6 January 2019).

McCartney, P. (2020). "Yoga-Scapes, Embodiment and Imagined Spiritual Tourism". Palmer, C.; Andrews, H. (eds), *Tourism and Embodiment*. London: Routledge, 86-106.

McCartney, P. (2021a). "The not So United States of Yogaland. Post-Nationalism, Environmentalism, and Applied + Yoga's Sustainable Development". Keaney, T.J. (ed.), *Nationalism. Past, Present, and Future*. New York: Nova, 1-122.

McCartney, P. (2021b). "The Abuse of Yoga and Proposition YSB. Herd Immunity for Troubling Times". *Abuse in Yoga and Beyond. Cultural Logics and Pathways for the Future* (Yoga and Abuse). https://sacredmattersmagazine.com/special-issue-yoga-abuse.

McCartney, P. (2021c). "Where Does India Stand When It Comes to Yoga Tourism?". *The Wire*. https://thewire.in/culture/india-yoga-tourism-uttarakhand.

McCartney, P.; Zuckermann, G. (eds) (2019). "Unsanitizable Yoga. Revivalistics and Hybridic Reclaimed Sanskrit". *Mentalities Journal*, 33(1), 1-48. https://doi.org/10.17605/OSF.IO/4GFN5.

Miller, Ch.P. (2020). "Soft Power and Biopower. Narendra Modi's 'Double Discourse' Concerning Yoga for Climate Change and Self-Care". *Journal of Dharma Studies*. https://doi.org/10.1007/s42240-020-00068-w.

Mitra, A. (1994). "Census 1961. New Pathways". *Economic and Political Weekly*, 29(51-2), 3207-21.

Mohan, A. (2012). *Utopia and the Village in South Asian Literature*s. New York: Palgrave Macmillan.

Mohan, S. (2020). "Revival of Sanskrit". *Eleven Myanmar*. https://elevenmyanmar.com/news/opinion-revival-of-sanskrit.

Nakamura, H. (1973). *A Companion to Contemporary Sanskrit*. Delhi: Motilal Banarsidas.

Nandy, A. (2001). *An Ambiguous Journey to the City. The Village and Other Odd Ruins of the Self in the Indian Imagination*. New Delhi: Oxford University Press.

Nelson, P.J. (2021). *Religious Voices in the Politics of International Development. Faith-Based NGOs as Non-state Political and Moral Actors*. Cham (CH): Springer Nature.

News Desk (2020). "Initiative. A Sanskrit Village Will Be Built in Every Block of Uttarakhand, People Will Be Taught the Language". *Amar Ujālā*. https://www.amarujala.com/dehradun/sanskrit-village-will-made-in-every-block-in-uttarakhand.

Office of the Registrar General and Census Commissioner, India (2020). "Census India". *Ministry of Home Affairs*. http://censusindia.gov.in.

Office of the Registrar General and Census Commissioner, India (1954). "Census of India. Paper no. 1. Languages. 1951 Census". New Delhi: Ministry of Home Affairs.

Ollet, A. (2017). *Language of the Snakes. Prakrit, Sanskrit and the Language Order of Premodern India*. Oakland (CA): University of California Press.

Olshansky, A.; Peaslee, R.M.; Landrum, A.R. (2020). "Flat-Smacked! Converting to Flat Eartherism". *Journal of Media and Religion*, 19(2), 46-59. https://doi.org/10.1080/15348423.2020.1774257.

Patherpanchali (2018). "Cows Will Talk in Tamil, Sanskrit Nithyananda Claims". *Infovision*. https://www.youtube.com/watch?v=LmYi_wNqR5s.

Plowden, W.C. (1883). *Indian Empire Census of 1881. Statistics of Population*, vol. 2. Calcutta: Superintendent of Government Printing India.

Posel, D.; Zeller, J. (2015). "Language Shift or Increased Bilingualism in South Africa. Evidence from Census Data". *Journal of Multilingual & Multicultural Development*, 37(4), 357-70. https://doi.org/10.1080/01434632.2015.1072206.

Prabhu, J. (2014). "Here's Why Sanskrit Is not a Dead Language and Should Be at the Core of School Curricula". *DNA*. https://bit.ly/3FLYGu3.

Priyanka (2014). "Where Are the Sanskrit Speakers?". *Gurumavin*. http://gurumavin.com/where-are-the-sanskrit-speakers [2020-07-12].

Press Trust India (2018). "India Is on the Path of Becoming 'Vishwa Guru'. Union Minister Harsh Vardhan". *Hindustan Times*. https://bit.ly/3l8Oz9g.

Press Trust India (2019). "India on Way to Become 'Vishwa Guru' Under Pm Narendra Modi. Amit Shah". *The Indian Express*. https://bit.ly/3wK48d5.

Ramanan, K.V. (1987). *Nāgārjuna's Philosophy. As Presented in the Mahā Prajñāpāramitā Śāstra*. New Delhi: Motilal Banarsidas.

Ramaswamy, S. (1999). "Sanskrit for the Nation". *Modern Asian Studies*, 33(2), 339-81.

Risley, H.H.; Gait, E.A. (1903). Census of India, 1901. Vol. 1-A, Part 2, *Tables*. Calcutta: Office of the Superintendent of Government Printing.

Samskrit101 (2009). "Samskrita Bharati ITV Interview Sri Chamu Krishna Shastry and Sowmya Joisa". *YouTube*. https://www.youtube.com/watch?v=VhpRKY4a3ds.

Sanskriti (2018). "Chanting Sanskrit Increases Brain Cognitive Areas. 'Sanskrit Effect'". *Sanskriti*. https://bit.ly/3MsvVFz.

Sharma, G. (2009). "Bajaj Discover Advertisement in Sanskrit". *YouTube*. https://www.youtube.com/watch?v=8FTVGNdDp2o.

Singh, R.K. (2014). "India Can Be a Vishwa Guru". *The Pioneer*. http://www.dailypioneer.com/todays-newspaper/india-can-be-vishwa-guru.html.

Singh, S. (2015). "New Life, Old Death for Sanskrit in Uttarakhand". *Archive*. http://archive.indianexpress.com/news/new-life-old-death-for-sanskrit-in-uttarakhand/1126773.

Skutnabb-kangas, T. (2012). "Linguistic Human Rights". Solan, L.M.; Tiersma, P.M. (eds), *The Encyclopedia of Applied Linguistics*. https://www.doi.org/10.1093/oxfordhb/9780199572120.013.0017.

Subbarao, K.V. (2008). "Typological Characteristics of South Asian Languages". *Language in South Asia*. Melbourne: Cambridge University Press, 49-78.

Subramaniam, B. (2019). *Holy Science: The Biopolitics of Hindu Nationalism*. New Delhi: Orient Blackswan.

Tare, K. (2010). "RSS Wants Sanskrit Registered in Census". *DNA*. http://www.dnaindia.com/mumbai/report-rss-wants-sanskrit-registered-in-census-1387878.

Taylor, M. (2012). "Heavenly Carrots and Earthly Sticks. How Phalaśruti Paratexts Empower Purāṇic Discourse". *Journal of Hindu Studies*, 5, 92-111. https://doi.org/10.1093/jhs/his012.

TNN (2019). "NASA Believes Sanskrit is Scientific Language for Programming". *The Times of India*. https://bit.ly/3w905HU.

Trivedi, A. (2010). "Sanskrit is Second Official Language in Uttarakhand". *Hindustan Times*. https://bit.ly/3LgevKG.

United Nations India (2019). "Yoga Is Force for Fighting Climate Change, Fostering Global Harmony: United Nations". *Hindustan Times*. https://bit.ly/39iAza8.

Upadhyay, V. (2020). "Uttarakhand to Soon Have 'Sanskrit Speaking' Villages". *The New Indian Express*. https://bit.ly/3wa7syK.

Vāgbhaṭa (2020). "Aṣṭāṅgasaṅgraha By Vṛddhavāgbhaṭa." *SARIT*. https://bit.ly/3sxYgSE.

Vertovec, S. (1994). "'Official' and 'Popular' Hinduism in Diaspora. Historical and Contemporary Trends in Surinam, Trinidad and Guyana". *Contributions to Indian Sociology*, 28(1), 123-47. https://doi.org/10.1177/006996694028001005.

Watanabe, C. (2019). *Becoming One. Religion, Development, and Environmentalism in a Japanese NGO in Myanmar.* Manoa: University of Hawai'i Press.

Witzel, M. (2010). "Gandhāra and the Formation of the Vedic and Zoroastrian Canons". Rotaru, F. (ed.), *The International Symposium. The Book*. Bucharest: Bucharest Library, 490-534.

Yeats, M.W.M. (1943). *Census of India, 1941*. Vol. 1, *India*; Part 1, *Tables*. Delhi: Manager of Publications.

e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

The Buddhist Text Known in Pāli as *Milindapañha* and in Chinese as *Nàxiān bǐqiū jīng* 那先比丘經 Some Philological Remarks and the Problem of the Archetype

Bryan De Notariis
Università Ca' Foscari Venezia, Italia

Abstract This article is conceived as an introduction to questions concerning the relationship between various versions of a Buddhist text known in its Pāli variant as *Milindapañha*, and in its Chinese versions as *Nàxiān bǐqiū jīng* (那先比丘經; T 1670 versions A and B). After a brief account of the conjectures about its redactor(s) and its public of the original Indian environment, the Chinese versions of the text will be dealt with in more detail, with particular attention to the Western reception and the problem related to the reconstruction of a possible archetype. The guidelines provided by Gérard Fussman will be taken into consideration, with some additional comments regarding the suggestion, in the case of the Chinese versions, of taking the Chinese audience into account. To confirm this point, a passage with an eristic dialogue, attested in both Pāli and in Chinese, will be analysed in detail to show how the Chinese translator(s) modified the text for the benefit of the public.

Keywords Buddhism. Milindapañha. Menander. Serindia. Gandhara. Pāli. Chinese.

Edizioni Ca'Foscari

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

Open access

Citation De Notariis, B. (2022). "The Buddhist Text Known in Pāli as *Milindapañha* and in Chinese as *Nàxiān bǐqiū jīng* 那先比丘經. Some Philological Remarks and the Problem of the Archetype". *Bhasha*, 1(1), 111-132.

DOI 10.30687/bhasha/8409-3769/2022/01/008

1 A Glance at the Literary Success of the Pāli *Milindapañha*

> I venture to think that the 'Questions of Milinda' is undoubtedly the master-piece of Indian prose; and indeed is the best book of its class, from a literary point of view, that had then been produced in any country. (Rhys Davids 1890, XLVIII)

Thomas W. Rhys Davids wrote these words in his introduction to the first full English translation of the Pāli text called *Milindapañha* 'Questions of Milinda'.[1] At that time, he did not know of the existence of a Chinese rendition of the text called *Nàxiān bǐqiū jīng* 那先比丘經 (T 1670 versions A and B), which would correspond to the Sanskrit **Nāgasenabhikṣusūtra* 'the Sūtra of the Monk Nāgasena'.[2] Interestingly enough, Rhys Davids probably bestows to the text more importance than that granted by the Buddhist tradition itself. Rhys Davids was writing "from a literary point of view", but the importance of the text from a doctrinal perspective surely deserves equal attention. The significance of the *Milindapañha* within the Theravāda tradition is, indeed, quite controversial. The *Milindapañha*, together with texts such as the *Nettippakaraṇa* and *Peṭakopadesa*, is regarded as a canonical text which is part of the *Khuddakanikāya* only by the Burmese tradition.[3] However, it was considered important enough to be quoted as an authority in Pāli commentaries: some of them even define it as a *sutta* (= *sūtra*),[4] in line with its nomenclature in the Chinese versions

An earlier version of this paper entitled "Remarks on the Pāli and Chinese Versions of the Buddhist *Milindapañha*" was presented for the first time at the international conference *Il re ellenistico e il saggio indiano. Il Milindapañha e il suo contesto / The Hellenistic King and the Indian Wise Man. Putting the Milindapañha in Its Context*, University of Bologna, 19-20 September 2019. I am grateful to Alberto Pelissero and Saverio Marchignoli for their remarks on the first draft. I should also thank Ven. Bhikkhu Anālayo and Giuliano Giustarini for having sent me valuable bibliographical sources to consider; Kenji Takahashi for the help in dealing with a Japanese source; and the two anonymous reviewers for their detailed feedbacks. It goes without saying that all remaining errors are my own responsibility. All translations from Pāli and Chinese are my own unless otherwise noted.

1 The Pāli title of the text, as known today, seems to be due to the editorial choice made by Trenckner (1880, VI) in the first full edition of the text. A variety of titles in some manuscripts is provided by Ooi (2021, 181-4). It is also worth highlighting that a thorough revision of the current *editio princeps* made by Trenckner in light of the Siamese edition of the text is still a desideratum and a worthwhile future task to accomplish. In this regard, see Skilling (2010) and Ooi (2021, 174).

2 It would seem that we can say either that the two Chinese versions "stem from a single original rendition" (Anālayo 2021a, 15) or, in other words, that "the two extant Chinese versions are the same work, one simply an amplification of the other" (Levman 2021, 108).

3 See Norman 1983, 31 and Allon 2018, 237-8.

4 See Nidd-a I, 166 = As, 108 quoted and discussed by Mori 1997-98, 297-8.

(*jīng* 經).[5] Then, was this text an important text or not? We can say that it was important enough to reach us, a fact that should not be underestimated since the method of transmission at that time involved a great deal of effort. Another Theravāda text called **Vimuttimagga*, for example, disappeared in the Indian mainland and Sri Lanka and has reached us in its entirety only in its Chinese translation (*Jiětuō dào lùn* 解脫道論; T 1648). Now, we are legitimised to wonder for whom and why this text was important. As it was theorised since the nineteenth century, the text is a product of Northwest India (Trenckner 1880, VII) and, as recently sustained by Stefan Baums (2018, 42), was created

> for the conversion of an audience that was neither Indian nor Greek, but part of the cosmopolitan melting pot of Gandhāra that was Indianized enough for the literary form of the Questions to appeal to it, Hellenized enough to be persuaded by its Greek style of argumentation and worldly enough to identify with the figure of the most famous foreign ruler of Gandhāra as he undergoes conversion to Buddhism.

Some elements could additionally lead us to hypothesise that it was a text created by monks for lay people or for people unfamiliar with Buddhism (the text might also have had the purpose of evangelising).[6] In this regard, it is interesting to consider the comparison made by Rhys Davids (1894, XX-XXVII) between the *Milindapañha* and *Kathāvatthu*. Both texts deal with controversial points in Buddhism, but they do it in quite different ways. It is worth mentioning the words of Rhys Davids (1894, XXVI):

> the controversy in the older book [i.e. *Kathāvatthu*] is carried on against members of the same communion, whereas in the Milinda we have a defence of Buddhism as against the outsider. The Kathā Vatthu takes almost the whole of the conclusions reached in the

5 The character *jīng* 經 does not purely mean *sūtra* but generically means 'scripture' (or originally, 'classic', including the many non-Buddhist classics revered in Chinese tradition). As such it was also used to translate the Indian word *sūtra*/*sutta*, but it does not necessarily point to this as the underlying term. Instead, it was often added by Chinese translators whether or not it had a counterpart in the title of the underlying Indic-language text. The Pāli commentaries' evidence might tip the scales in favour of the assumption that *jīng* 經 here actually means *sūtra*/*sutta* and, indeed, **Nāgasenabhikṣusūtra* has been a widespread reconstruction for the original Indian title of the text (e.g. Nanjio 1883, 304; Thich 1964; Guang 2007; 2008; 2009). For the sake of the present study, I will adopt this last rendering as a scholarly convention, bearing in mind the complexity behind this issue.

6 Similarly, Levman (2021, 113, 125) suggests that the *Milindapañha* was a sort of Buddhist 'catechism', implying with this term that the text was orally transmitted (from the Greek meaning of *katechízein* 'to instruct orally'), a fact that, in my opinion, we should be cautious to endorse.

> Milinda for granted, and goes on to discuss further questions on points of detail.

It is clear that the *Milindapañha* deals with topics that are more basic than the ones treated in the *Kathāvatthu*. Moreover, we also know that the *Milindapañha* refers silently to many Buddhist texts (Rhys Davids 1890, XXVII-XXXI), and more directly to others (Rhys Davids 1890, XXXI-XXXVI).[7] Therefore, on the one hand it seems that the *Milindapañha* was composed by people that were erudite in the Buddhist doctrine;[8] on the other hand, the topics treated are not sophisticated disputes on minor issues (the kind of things that would interest scholar monks), but concepts that are at the very core of Buddhism (e.g. *anātman*, rebirth, *karman*, Buddha, *nirvāṇa* etc.) and that would interest a person who knows a little about Buddhism.[9] This is also reflected by the history of the text in China. According to the conclusion reached by Guang Xing (2009), the text was translated into Chinese in a very early stage and so it seems to be among the first Buddhist texts arrived in China.[10] This is informative about the nature of the text, which, evidently, was able to satisfy the expectations and needs of a Chinese audience. In this regard, Kōgen Mizuno (1982, 46) wrote that

> Although many Chinese were curious about Buddhism and were interested enough in the sutras to want to study them, they could not really comprehend the alien Buddhist doctrines or philosophy; thus they read primarily the general moral teachings and stories that neither contain technical terms nor expound doctrine. Those simple teachings and stories, presented in ordinary language, were comprehensible, interesting, and useful.

Therefore, it seems that the *Milindapañha*'s fate was governed by the fact of being both a simple text and a text that can appeal a huge audience. This may be the past and present's good fortune of this oeuvre, a text that was also adopted by the Theravāda Buddhist school. This

7 See also the updates provided by Horner (1969, XI).

8 Horner writes that the *Milindapañha* "has a wide range and covers much ground, denoting deep erudition on the part of its compiler" (1969, IX).

9 This, according to Thich (1964, 32), seems to be especially true for the first sections of the *Milindapañha* (following the Pāli version's division into seven sections), whereas the last three sections are more sophisticated than the previous ones.

10 Notably, the Gandhāra region – from which the *Milindapañha* is supposed to have originated – was a pivotal area for the transmission of Buddhism in central Asia and China. In this regard, see Neelis 2011, 42-7, 229-56. As reported by Richard Salomon (2018, 26) "[r]ecently a few small fragments have been discovered of a Gāndhārī text that has some resemblance to the *Questions of Milinda*, including a reference to Nāgasena, but they seem to belong to some related tradition rather than to the *Questions* itself".

tradition preserved a Pāli version, which, however, is longer than the Chinese ones. The latter show an earlier stage of development of the work and cover only Mil, 1-89, leaving the remaining part (Mil, 90-420) without any other parallel. The Pāli version is not only longer, but is also 'Theravādised' (may the reader forgive my neologism),[11] although it maintains some odd passages which are clearly referring to the doctrines of other Buddhist traditions, such as the Sarvāstivāda school.[12] For some reasons, the Pāli version had more popularity in the West than the Chinese versions (Guang 2008, 237), and so it would make sense, not only now but also in the future, to further investigate the Chinese versions in order to shed new light on such an amazing, and to some extent unique, piece of literature.

11 A good example to demonstrate that the Theravāda school modified the text is the presence of the concept of *bhavaṅga* in the *Milindapañha* (Mil, 299-300). The term is, indeed, peculiar to the Theravāda tradition and is found primarily in the Pāli texts. As stated by Kim (2018, 754), Vasubandhu also wrote within his **Karmasiddhiprakaraṇa* (*Dàshèng chéngyè lùn* 大乘成業論; T 1609) that *bhavaṅga* (*yǒufēn shì* 有分識) originated among the *Tāmraparṇīya*(*-nikāya*) (= *chìtóngyè* 赤銅鍱) (the full passage reports: 赤銅鍱部經中建立有分識名; T1609.31.0785a14). Kim also writes that "*Tāmraparṇīya* refers to, or is at least closely related to Sri Lankan Theravāda tradition" (2018, 754). It could be of some interest here to highlight that the manuscripts used by Trenckner for his edition of the *Milindapañha* were mostly copied in Sri Lanka (Trenckner 1880, III-VII). Another element of the Theravāda within the *Milindapañha* is the interpretation of the term *kappa* 'aeon' in connection with the possibility on the Buddha's behalf to extend his life through the *iddhipāda* 'the foundation of psychic power'. In the canonical literature it is written that "anyone who has cultivated the four *iddhipāda*s, who has practiced them assiduously, mastered them, made them as a base, established them, become acquainted with them, properly undertaken them, he can last, as he wishes, for a *kappa* or what remains of a *kappa*" (*yassa kassaci cattāro iddhi-pādā bhāvitā bahulīkatā yānikatā vatthu-katā anuṭṭhitā paricitā susamāraddhā, so ākaṅkhamāno kappaṃ vā tiṭṭheyya kappāvasesaṃ vā*; D, II, 103). In this passage, the term *kappa* is interpreted, according to the Pāli commentarial literature, as the *āyu-kappa*, i.e. the 'life-span' (*ettha ca kappan ti āyu-kappaṃ*; Sv, II, 554), whereas the interpretation of the term *kappa* as indicating a *mahā-kappa* 'cosmic aeon' would seem the right one (Gethin 2001, 94-7). The fact that the *Milindapañha* (see Mil, 141), in the same context, also sustains the reading *āyu-kappa* may be an indicator of the Theravāda's influence. Finally, as highlighted by Thich (1964, 23), it is worth noting that the Pāli version mentions the names of Theravāda Abhidhamma.

12 There is, indeed, a mention of the existence of three times in the *Milindapañha* (see Mil, 49-50), a clear sign of the Sarvāstivāda's influence (see also Guang 2008, 239). Another of Sarvāstivāda's characteristics is found in Mil, 268-71, in which *nibbāna* (Sanskrit: *nirvāṇa*, the well-known ultimate goal of Buddhists) and *ākāsa* (= Sanskrit: *ākāśa* 'space') are described as *akammaja* 'not born of *kamma*', *ahetuja* 'not born of cause', *anutuja* 'not born of physical change'. This would remind the Sarvāstivāda's tenet according to which only *ākāśa* and two kinds of *nirvāṇa*s are considered *asaṃskṛta* 'unconstructed', whereas for the Theravāda tradition the *nibbāna* only is considered as such (see Lamotte 1988, 609-10; Horner 1969, XVIII). Finally, another point that differs from the orthodox Theravāda tradition is the eight investigations (*aṭṭha mahāvilokanāni*) at Mil, 193, because these investigations are only five in the Pāli commentaries (Horner 1969, XVI-XVII).

2 Western Reception of the Chinese Versions and the Problem of the Archetype

> A comparative study of the P[āli] *Milindapañha* and the C[hinese] *Na-hsien-pi-ch' iu ching* shows clearly that both versions derive from the same source as they have many points in common between them […] the trend of the dialogues is almost identical, the dialogues veer round the same theme, with unimportant divergences scattered unevenly. (Thich 1964, 1)

> [A] detailed comparison of the Chinese and Pāli texts does not support translation from a common text, as the vast majority of the translations are quite different, being not literal but paraphrases; the overall content is generally held in common, but the details of the similes are often quite different. (Levman 2021, 113)

As it might be noted from these quotations, scholars can have different inclinations concerning earlier sources underlying the extant versions of our text, being them either a common ancestral source or even the *Urtext*. The resulting judgment might seem *prima facie* based on either giving pre-eminence to similarities or differences. However, the recognition that there is something in common leaves little doubt. Here, this section aims not to establish a definitive answer to the conundrum of the existence of an archetype, but is of a more modest scope. Following the introduction of some historical data on the Chinese versions of the text, their Western reception is analysed, highlighting how, since the very beginning, some scholars showed a certain anxiety in establishing which one among the versions is closer to the original. This quest prompted the establishment of a methodological approach, exemplified by some applicable guidelines suggested by Gérard Fussman. A corollary to a guideline will be proposed and to prove its usefulness an eristic dialogue shared by all versions will be analysed.[13]

2.1 Dating the Chinese Versions of the Text

The text in its Chinese versions is called *Nàxiān bǐqiū jīng* (那先比丘經; T 1670) – which would correspond to the reconstructed Sanskrit form **Nāgasenabhikṣusūtra* – and the compilers of Chinese catalogues of Buddhist scriptures ascribed it to the Eastern Jin dynas-

13 I use the term 'eristic' to describe the dialogue that will be analysed because the disputers will exploit the ambiguity of words to win in the debate, rather than using logic to approach a more objective truth.

ty (東晉; 317-420 CE).[14] As this date was established in retrospect by later catalogues, it would make sense to also consider the Japanese scholarship, according to which the text was translated into Chinese no later than the third century CE, probably around the second century.[15] The text has been handed down to us in two versions, A and B, whereas another lost version was translated into Chinese around the third century (Demiéville 1924, 8, 21; Fussman 1993, 67). In the Western academic environment, the text established itself gradually, partly obscured by the success of its counterpart in Pāli.

2.2 Western Reception

In his *A Catalogue of the Chinese Translation of the Buddhist Tripiṭaka*, Bunyu Nanjio (1883) cautiously wrote that the **Nāgasenabhikṣusūtra* "seems to be a translation of a text similar to the Milinda-pañho, though the introductory part is not exactly the same as that of the Pāli text" (304). Nanjio was cautious in his statement since he com-

14 See Demiéville 1924, 9, 21; Fussman 1993, 66-7; Guang 2009, 227. It is not entirely clear to me the reason why von Hinüber (1996, 83) dated the translation of the text to the fourth century despite relying upon Demiéville, who reported the Eastern Jin dynasty (東晉; 317-420 CE) as the period of its translation, namely the fourth and, potentially, fifth centuries (Demiéville 1924, 21). Norman, who also was relying on Demiéville, wrote that "[t]here is also a Chinese version, which can be dated to a time earlier than the fourth century A.D." (1983, 111). It seems likely that Norman was considering the existence of the lost version.

15 "Modern Japanese scholar Mizuno Kogen argues convincingly that the Nāgasena Bhikṣu Sūtra was translated into Chinese in Latter Han dynasty (25-220 CE), not later than San Guo Dynasty [Three Kingdoms Period] (220-280 CE), even conservatively" (Guang 2009, 236). Guang bases his statement on Mizuno 1959, 29-33. In an English work of Mizuno it is written that he dates the Chinese translation around 200 CE (Mizuno 1982, 196), whereas de Jong (1996, 383) reports that Mizuno dates the translation around the second or third century CE. Mori (1997-98, 292 fn. 3) and Thich (1964, 104-5) also follow Mizuno's study. Given that the article by Mizuno is written in Japanese, I asked a Japanese colleague, Dr. Kenji Takahashi, to check the relevant pages for me and he kindly sent me the following Japanese quotation with its translation: "To give a conclusion first, considering the various points that I will describe in what follows, I [argue] that the translation of this text/*sūtra* is much older than Eastern Jin (東晋) period and should be placed during the period of the Later Han (後漢) and that at the latest it is not later than the Three Kingdoms (三國) period" (結論的に云えば、次に述べるような種々の點から見て、私は本經の譯出は東晋時代よりも遙かに古く、後漢代に置かるべきであり、おそくとも三國時代を下るものではないということである。; Mizuno 1959, 30). It goes without saying that a thorough examination of Mizuno's findings and Japanese scholarship in general would be of great benefit for future studies. At the moment, many scholars (including myself) can only rely on second-hand reports. The reasons adduced by Mizuno to predate the text seem to be stylistic in nature as according to Guang (2009, 236-43), Mizuno provides three reasons to support his argument: 1) the terminology used in the **Nāgasenabhikṣusūtra* is comparable with the translation made by Ān Shìgāo 安世高; 2) some proper terms and pronouns are quite archaic and were often used during the Han dynasty; 3) the *gāthā*s were translated into prose.

pared the Chinese rendition with the translation provided by Vilhelm Trenckner in his *Pali Miscellany* (1879), that reported a specimen of the Pāli version of the text. The Pāli text was edited in full by Trenckner only in 1880 and the first complete English translation was made by Thomas W. Rhys Davids only in 1890. For a full recognition of the Chinese versions, we should wait until 1893, when Edouard Specht and Sylvain Lévi clearly identified the two Chinese translations as a parallel text to the Pāli *Milindapañha* or, to be more precise, as different recensions of a text of which the Pāli version represents only one recension (Specht, Lévi 1893).[16] The discovery of the Chinese translations of the *Milindapañha* influenced the question about the existence of an original text. Trenckner already believed that the Pāli version was a translation from another text:

> It [i.e. *Milindapañha*] must have been imported from northern India, where alone the name of the conqueror [i.e. Milinda] can have been preserved. In all probability the original was in Sanskrit, and our text is a translation. (Trenckner 1880, VII)[17]

However, after the discovery of the Chinese translations, it is possible to wonder which one among the Pāli and Chinese versions is closer to the original work or if it is possible to recover an archetype comparing the recensions.[18] Just one year after the publication of Specht and Lévi's article, Rhys Davids published the second volume of his *Milindapañha*'s translation, taking into account the existence of the Chinese versions. He had the feeling that Specht and Lévi wanted to demonstrate that the Chinese versions were older recensions than the one in Pāli. Conversely, Rhys Davids seems to suggest that the Pāli version, despite being longer, might represent the closest one to the original work.[19] However, more accurate comparisons between the Pāli and

16 "Un simple examen suffit pour constater que nous avons trois rédactions du même ouvrage qui a été successivement remanié" (Specht, Lévi 1893, 521).

17 It is worth noting that the early assumption made by Trenckner that the original language of the text was Sanskrit has been replaced by the hypothesis that it was Gāndhārī. In this regard, see Demiéville 1924, 11; Fussman 1993, 66; von Hinüber 1996, 83; Kubica 2014, 188; 2021, 430; Baums 2018, 33; Salomon 2018, 26; Levman 2021, 110.

18 Here, it is worth mentioning the recent contribution of Jonathan Silk (2021), who presents some theoretical remarks on how to approach Mahāyāna Buddhist texts. He highlights that if our sources lack a 'unique redactorial moment', it would be impossible to recover the *Urtext* simply because it never existed. However, the versions of our text share a common core that is synoptically consistent. In many cases, its analysis helps us to move close, if not to the archetype, to the best reading based on evidence (see the discussion below). This approach would be appropriate even when considering the recent contribution of Bryan Levman (2021), who rejects the idea that Pāli and Chinese versions were based on a common text but recognises the existence of a common core.

19 "Both M. Specht and M. Sylvain Lévi seem to think that the two Chinese books were translations of older recensions of the work than the one preserved in Pāli. This

Chinese versions (notably, Demiéville 1924; Fussman 1993) suggest that the Pāli version is an enlarged version and so, to some extent, the Chinese recensions are closer to the archetypal work, if any.[20] This does not mean that it is the original work and in this regard it seems useful to quote the words of Gérard Fussman, who, in a succinct but very informative passage, provides a sketch for interpreting the relationship between the different versions of the text:

> [T]out détail commun au text chinois et au texte pāli a chance de remonter à la source originale. Tout détail attesté par un seul de ces deux textes est suspect d'être une addition, surtout lorsque ce détail se trouve dans la version pāli que l'on sait avoir été profondément remaniée et amplifiée. Il existe néanmoins une possibilité théorique que l'un des deux textes, chinois ou pāli, ait conservé une information disparue de l'autre texte; dans ce cas-là il est du devoir de l'historien de prouver que cette information remonte à la source originale avant de songer à l'utiliser. Enfin, si une information livrée à la fois par le texte chinois et le texte pāli doit être décodée ou interprétée pour être pleinement utilisable, cette interprétation doit convenir aux deux textes à la fois. J'ajouterai aussi que cette interprétation doit tenir compte du fait que ces textes sont des textes indiens, utilisant une phraséologie et des procédés littéraires indiens et s'inspirant nécessairement de la conception du monde et de l'imaginaire indiens. (Fussman 1993, 68)

This passage provides many interesting guidelines which can be summarised as follows:

1. Details shared by both versions may derive from the original source.
2. Details attested in only one version may be additions (especially if occurring in the Pāli version).
3. Theoretical possibility that some details survived in only one version and disappeared in the other.

argument does not seem to me, as at present advised, at all certain. It by no means follows that a shorter recension, merely because it is shorter, must necessarily be older than a longer one. It is quite as possible that the longer one gave rise to the shorter ones" (Rhys Davids 1894, XII). A similar position is expounded in Rhys Davids 1916, 632. In this regard, I quite agree with Olga Kubica who wrote that "when Rhys Davids expressed his opinions concerning Pāli literature, his conclusions were very reasonable, but when Chinese literature entered the discussion, it seems that the desire to emphasize the superiority of Pāli literature over Chinese prevailed" (2014, 195-6).

20 "[L]a plus proche de l'original a chance d'être la chinoise, plus anciennement attestée que le *Milindapañha*, et surtout beaucoup moins remaniée que le texte pāli y compris dans ses parties narratives" (Fussman 1993, 68). A list of reasons according to which the Chinese versions should be regarded as an older record than the Pāli version is provided by Thich (1964, 24-35) and Guang (2008, 242-3).

4. Information that occurs in both versions should be carefully interpreted and the interpretation should fit both versions.
5. The Indian context of the text should be taken into account during the process of interpretation.

I would add an additional item to this list, which can be regarded as a sort of corollary to the last item, namely that in the case of the Chinese versions, the Chinese audience should be taken into account.

2.3 An Explicative Example. A Buddhist Eristic Dialogue

In this regard, there is an interesting eristic dialogue between Milinda and Nāgasena survived in both recensions. This episode is reported in the Pāli version as follows:

> The king said: "Venerable Nāgasena, is the Buddha one who observes celibacy [*brahmacārin*, lit. 'one who has the brahma-conduct']?"
> "Yes, great king, the Buddha was one who observes celibacy!"
> "Then Venerable Nāgasena, the Buddha is a pupil of Brahmā!"
> "Do you have, great king, a state elephant [*hatthipāmokkho*]?"
> "Yes, Venerable, I have it."
> "Does this elephant, great king, make the trumpet noise [*koñcanāda*] at times?"
> "Yes, Venerable, it does it."
> "Then, great king, this elephant is a pupil of herons [*koñca*]!"
> "It is not so, Venerable!"
> "And is Brahmā, great king, intelligent [*sabuddhika*] or stupid [*abuddhika*]?"
> "He is intelligent, Venerable!"
> "Then, great king, Brahmā is a pupil of the Blessed one [*bhagavant*, epithet of Buddha]!"
> "You are witty Venerable Nāgasena".[21]

In this passage, there are many puns and it is here that this way of playing on words seems to bear the weight of a logical argument, al-

21 *rājā āha: bhante Nāgasena, Buddho brahmacārī ti. – āma mahārāja, Bhagavā brahmacārī ti. – tena hi bhante Nāgasena Buddho Brahmuno sisso ti. – atthi pana te mahārāja hatthipāmokkho ti. – āma bhante, atthī ti. – kin-nu kho mahārāja so hatthī kadāci karahaci koñcanādaṃ nadatī ti. – āma bhante, nadatī ti. – tena hi mahārāja so hatthī koñcānaṃ sisso ti. – na hi bhante ti. – kim-pana mahārāja Brahmā sabuddhiko abuddhiko ti. – sabuddhiko bhante ti. – tena hi mahārāja Brahmā Bhagavato sisso ti. – kallo si bhante Nāgasenā ti* (Mil, 75-6).

beit there is nothing which appears clearly logical.[22] The first question is put by Milinda to Nāgasena, asking if the Buddha is a *brahmacārin*, a term that came to mean 'celibacy' or 'chastity', in the sense of a total abstention from sexual intercourse.[23] However, since this term is composed by the words '*brahma*' and '*cāra*' plus the suffix *-in* used to create adjectives, literally it means 'one who has the brahma-conduct'. So, according to Milinda, if the Buddha has a brahma-conduct this means that he can be regarded as a follower of Brahmā (one Indian god).[24] This reasoning is certainly deceptive since it does not take into account the real meaning of the word, applying an overly literalistic interpretation. Therefore, Nāgasena answered to the king using the same reasoning, showing, at first, that it would lead to ridiculous results: as the elephant trumpets (*koñcanāda*), this would mean that it should be regarded as a follower of herons (*koñca*). Secondly, he shows that it is possible to demonstrate in the same way that the god Brahmā is a follower of the Buddha because he is intelligent (*sabuddhika*).

Now, turning to the Chinese version B, it is possible to note a slightly different phrasing for the first part concerning the *brahmacārin*'s pun, which is meaningful. The Chinese recension reports:

> The king had again a question to Nāgasena: "Has the Buddha practised the conduct of Brahmā, who is the king of the seventh heaven, having not any sexual intercourse with women?" Nāgasena replied: "[The Buddha] keeps himself completely apart from women, he is pure, without any flaw or contamination".[25]

22 It seems, indeed, that we should look at this dialogue in light of the ancient Indian way of debating. In this regard, and with particular reference to the debates in the *Milindapañha*, see Anālayo 2021a.

23 See Gombrich 2009, 202-3. Besides, as highlighted by Neri and Pontillo (2014, 160) the meaning of *brahmacariya* "cannot be limited to a life of chastity, but includes a 'path of life' and has other important links with the highest achievements of the Buddha's path". However, the meaning of 'chastity' is certainly relevant to our context.

24 A discussion on the cult of Brahmā in ancient India is provided by McGovern (2012). To be thorough, it can be worth mentioning that in addition to the interpretation of the stem *brahma/ā* as the Indian god Brahmā, another widespread meaning is that of 'excellent' or 'foremost' just as when the Pāli commentaries gloss *brahma* as *seṭṭha*. For instance: *brahman ti seṭṭhaṃ uttamaṃ visiṭṭhaṃ* (Ps, II, 27); *brahmabhūto ti seṭṭhasabhāvo* (Mp, V, 72); *brahmapattiyā ti seṭṭhapattiyā* (Spk, I, 265). Or, more specifically on *brahmacariya*: *brahmacariyan ti seṭṭhaṭṭhena brahmabhūtaṃ cariyaṃ brahmabhūtānaṃ vā buddhādīnaṃ cariyan ti vuttaṃ hoti* (Sv, I, 179). In some passages, being a *brahmacārin* is even equated with the attainment of the *arahant* state: "A pure *brahmacārin* is a monk who has destroyed the noxious influxes (i.e. an *arahant*)" (*suddhaṃ brahmacārin ti khīṇāsavabhikkhuṃ*; Sp, II, 484). According to some canonical passages, the word *brahma-* in some compounds can even be synonym with the word *dhamma*, as in *brahma-kāya*, *brahma-bhūta*, *brahma-yāna* (cf. Neri, Pontillo 2014, 170-1).

25 王復問那先:"佛審如第七天王梵所行,不與婦女交會不?"那先言然:"審離於婦女,淨潔無瑕穢。" (T1670B.32.0716b05-07), other translations of this passage are provided by Demiéville (1924, 158), Thich (1964, 87), Guang (2007, 177), Anālayo (2021b, 193).

Here, it seems as though the text is trying to explain the pun to the reader, by providing the two different meanings of '*brahma*' (= *fàn* 梵) involved.[26] At first, it is written that the so called 'brahma-conduct' (*fàn suǒxíng* 梵所行) is referring to the king of the seventh heaven (i.e. Brahmā)[27] and, secondarily, it is specified that the term is referring to the fact that there is total abstention from sexual intercourse (*jiāohuì* 交會) with women (*fùnǚ* 婦女).[28] Theoretically, we may wonder whether it was either the Chinese version that added new material or if it was the Pāli version that removed these parts. In this regard, it is useful to remember both Fussman's suggestion, namely, to take into account the Indian context, and my suggestion to take into account the Chinese audience. Considering these presuppositions, we should admit that the word *brahmacārin* would not require any explanation in India since it has formed part of the Indian culture for many years, given that it also occurs within the oldest Indian text recorded, i.e. the *Ṛgveda* (10.109.05). In the same way, we can assume that a native Chinese speaker would have some difficulty in grasping the meaning of the term, let alone the pun behind the passage. Therefore, the hypothesis that the Chinese version enlarged the text in order to better convey the pun might be more plausible than to suppose that the original work had these kinds of specifications.[29] The fact that the Chinese version B modified the text to satisfy the Chinese audience is also evident from another part of the same account. In this context, it is also useful to involve the version A of the Chinese translation. The point at issue is the pun based on the trumpet of the elephant (*koñcanāda*), which would lead to the (il-)logical result that the elephant is a pupil

26 The interpretation that the Chinese text gives to the stem *brahma/ā* is one among the many polysemantic uses and for further details see the seminal article of Neri and Pontillo (2014).

27 It is possible to compare the reconstruction of the Buddhist cosmology made by Gethin (1997, 195; 1998, 117-8) and De Notariis (2019, 66-7) to verify that the seventh world above the human realm is actually called Brahmā's retinue (*brahmapārisajja*), see the Appendix.

28 The term *suǒxíng* 所行 can also be translated as 'practice' as in the expression *shífǎ suǒxíng* 十法所行 (T0280.10.0445a28) that Jan Nattier (2007, 113) translates as 'ten practices'. This interpretation of the term would expand its scope beyond mere celibacy and would be in line with views of *brahmacariya* and *brahmacārin* as expounded in the Pāli canonical and commentarial literature (see fn. 24 above). This may shed some light upon the need to clarify the locution *fàn suǒxíng* 梵所行 that the Chinese redactors had.

29 Stefano Zacchetti highlighted that often in the process of translating, the text's interpretation or, let us say, its exegesis, was actively involved and put into the final translated version of the text. In this regard, he writes that: "Forse le prime traduzioni cinesi non erano lontane da questa situazione: in altre parole, l'elemento originariamente traducibile sarebbe stato non tanto il *sūtra*, quanto la sua esegesi orale" (Zacchetti 1996, 357-8; Author's transl.: "Perhaps the earliest Chinese translations were not far from this situation: in other words, the originally translatable element would have been not so much the *sūtra*, as its oral exegesis").

of herons (*koñca*). In **[tab. 1]** below, it is possible to compare the Pāli version with the Chinese versions A and B.

Table 1 Comparison between the Indian and Chinese versions on the elephant/bird song

Pāli version	**Chinese version A**	**Chinese version B**
"Does this elephant, great king, make the trumpet noise [*koñcanāda*] at times?" "Yes, Venerable, it does it." "Then, great king, this elephant is a pupil of herons [*koñca*]!" "It is not so, Venerable!"	Nāgasena asked to the king: "What is elephant [*xiàng* 象] song [*míngshēng* 鳴聲] like?" The king replied: "The elephant song is like the singing [*shēng* 聲] of a wild goose [*yàn* 鴈]." Nāgasena said: "If so, the elephant [*xiàng* 象] is the pupil of the wild goose [*yàn* 鴈], but each one of them is a different species."	Nāgasena asked to the king: "What is bird [*niǎo* 鳥] song [*niǎoshēng* 鳴聲] like?" The king replied: "The bird song is like the singing [*shēng* 聲] of a wild goose [*yàn* 鴈]." Nāgasena said: "If so, the bird [*niǎo* 鳥] is the pupil of the wild goose [*yàn* 鴈], but each one of them is a different species."
kin-nu kho mahārāja so hatthī kadāci karahaci koñcanādaṃ nadatī ti. – āma bhante, nadatī ti. – tena hi mahārāja so hatthī koñcānaṃ sisso ti. – na hi bhante ti. (Mil, 76)	那先問王言: "象鳴聲何等類?"王言: "象鳴聲如鴈聲!"那先言: "如是象爲是鴈弟子,各自異類。" (T1670A.32.0700c20-22)	那先問王言: "鳥鳴聲何等類?" 王言: "鳥鳴聲如鴈聲!"那先言: "如是鳥爲是鴈弟子,各自異類。" (T1670B.32.0716b11-13)

Concerning this passage, Demiéville believed that the Chinese translator(s) did not understand the pun. Therefore, he wrote that:

> Ce passage est corrompu; les copistes chinois ne pouvaient comprendre les jeux de mots (*buddha* et *buddhi*, *koñcanāda* «barrissement» et «cri du héron»), par lesquels Nāgasena réplique à la boutade étymologique du roi. (Demiéville 1924, 158 fn. 5)

In my opinion, it is the other way around. The Chinese translator(s) modified the passage just because they understood the pun and so tried to render it in the best way for their audience. It is worth noting that the first animal involved in the two Chinese versions is different. In the version A an elephant (*xiàng* 象) occurs, whereas version B replaces it with a bird (*niǎo* 鳥). In the Pāli version, we find an elephant, just as the Chinese version A. However, the second animal (i.e. the wild goose, *yàn* 鴈) is the same in both versions, and so also the term used to designate the animal's call (*míngshēng* 鳴聲). We can note from the Pāli version that the pun is due to the similarity between the trumpet *koñcanāda* (which literally means the sound *nāda* of the heron *koñca*), and the herons *koñca*. A kind of similarity is involved also in the Chinese phrasing, but there is not a phonetical similarity as in the Pāli version, but the similarity is here an ideographic one. It is possible, indeed, to note that the combination of characters used to denote the elephant's trumpet is *míngshēng* 鳴聲 and the second animal

involved is *yàn* 鴈, a wild goose. The characters *míng* 鳴 and *yàn* 鴈 have something in common, namely they share the same radical: *niǎo* 鳥 (which is, incidentally, the first animal involved in version B in place of the elephant of the version A). Here, the pun is more ideographic than phonetical as in the Pāli version, and this kind of rendition is certainly more suitable to a Chinese audience. It is also possible that the use of *míngshēng* 鳴聲 to indicate the elephant's trumpet in version A was a little forced. This can be inferred by the existence of *xiàngshēng* 象聲, probably a more appropriate term to designate the elephant's trumpet.[30] A search into *The SAT Taishō Shinshū Daizōkyō Text Database* shows that there are sixty-four occurrences for *xiàngshēng* 象聲, and only one more occurrence for *xiàng míngshēng* 象鳴聲, in addition to the occurrences in T 1670 A. It would seem that the choice of *míngshēng* 鳴聲 to designate the trumpet of the elephant (*xiàng* 象) is quite peculiar and might have sounded a bit odd. Perhaps, this was the reason why the Chinese version B emended the elephant (*xiàng* 象) with a bird (*niǎo* 鳥). It is indeed possible that for the other Chinese translator(s) the combination of characters in *míngshēng* 鳴聲 recalled something like a twitter or a chirp rather than a trumpet.[31]

Naturally, we cannot definitively exclude the possibility that the character *niǎo* 鳥 for *xiàng* 象 is the result of a hypercorrection by a scribe who thought that *xiàng* 象 must be a mistake, or even (given the overall similarity in shape of the two characters) a simple copying error. So, in this case we may wonder whether is better to assume an ancient dully scribe or skilful one. Similarly, considering in general the translation of the entire dialogue, we may wonder whether it is better to assume a dully translator – who either did not understand the pun or was not able to render it into Chinese – or a knowledgeable one who skilfully adapted the text to the target audience. Assuming, for the sake of argument, the latter case, we can read the remaining part[32] of this eristic dialogue in a new light.

30 *Xiàngshēng* 象聲 would correspond to the Sanskrit *nāgasvara* or *nāgaśabdā*, see Hirakawa 1997, 1105.

31 This fact may support the hypothesis that sees the version B as a more revised version when compared with version A, as sustained by Guang (2009). However, as a reviewer of this paper highlighted, the picture outlined by Mizuno (1959) – who has been one of the main Guang's sources – can be much more complex, and so future studies to understand and include his findings will be needed.

32 This section occurs in the middle of the account in the Chinese versions and as the last part of the Pāli version.

Table 2 Comparison between the Indian and Chinese versions of the section including the pun based on *sabuddhika*/*abuddhika* and *yǒu niàn* 有念/*wú niàn* 無念

Pāli version	**Chinese versions A and B**
"And is Brahmā, great king, intelligent *sabuddhika*] or stupid [*abuddhika*]?" "He is intelligent, Venerable!" "Then, great king, Brahmā is a pupil of the Blessed one [*bhagavant*, epithet of Buddha]!" "You are witty Venerable Nāgasena."	Nāgasena asked to the king: "The king of the seventh heaven is intelligent [*yǒu niàn* 有念] or stupid [*wú niàn* 無念]?" The king replied: "Brahmā, the king of the seventh heaven, is intelligent!" Nāgasena said: "For this reason, Brahmā, the king of the seventh heaven, as well as all the high gods, should be considered a disciple of the Buddha [*fó* 佛]!"
kim-pana mahārāja Brahmā sabuddhiko abuddhiko ti. – sabuddhiko bhante ti. – tena hi mahārāja Brahmā Bhagavato sisso ti. – kallo si bhante Nāgasenā ti (Mil, 76)	那先問王: "第七天王者有念無念?"王言: "第七天王梵有念!"那先言: "是故第七天王梵及上諸天皆爲佛弟子!" (T1670A.32.0700c18-20 = T1670B.32.0716b08-11)

On the surface, we may wonder why the translator(s) did not operate any change in explaining the connection of *sabuddhika* and *abuddhika*, translated respectively into Chinese as *yǒu niàn* 有念 and *wú niàn* 無念, to the word Buddha, in Chinese *fó* 佛. The term *niàn* 念 has always been interpreted as translating the word *buddhi*, probably, considering the evidence from the Pāli version and, possibly, underpinned by the fact that *niàn* 念 as much as *buddhi* broadly relates to the mental dimension (as the radical *xīn* 心 of *niàn* 念 would suggest). The act of providing a modern translation to the Chinese passage, in addition to the convention of following the Pāli version, obscures the fact that to the Chinese reader the passage, as it is written, might already convey the pun. If we step for a moment into the Chinese readers' shoes, could we really assume that the first concept that would rise in their mind when reading *niàn* 念 is something alike the Indic term *buddhi*? Rather, arguably, other concepts involving the word *niàn* 念 were more popular and probably the foremost was *niàn* 念 understood as one of the practices of 'recollection', the first of which is traditionally the 'Recollection of the Buddha' (see, for instance, Vism, 197) that in Chinese is *niàn fó* 念佛. In this case, *niàn* 念 translates the Pāli word *anussati* (= Sanskrit: *anusmṛti*) which indicates the systematic exercise of recollecting or calling to the mind[33] something that, in the case

33 "Because it is a mindfulness (*sati*) that arises again and again is [called] recollection (*anussati*)" (*punappunaṃ uppajjanato sati yeva anussati*; Vism, 197). Interestingly enough, according to Rupert Gethin (2001, 37) "[t]he *Milindapañha* contains what is perhaps the earliest attempt in Buddhist literature to state fully just what *sati* is. Questioned by king Milinda as to the characteristic (*lakkhaṇa*) of *sati*, the monk Nāgasena

of the *Buddhānussati*, concerns the Buddha and his qualities.[34] This technique is transversal to many Buddhist traditions and, especially along the Silk Road, developed in popular forms that arrived till the far East Asia.[35] Even our Chinese versions B testifies the existence of that practice:

> You ascetics say: "People who during their life practices evil for one hundred years [can], once approached the time of death, recollect the Buddha [*niàn fó* 念佛] [and so] all of them will after death be born in high heavens".[36]

And also:

> Although a person has been evil in the past, having recollected the Buddha [*niàn fó* 念佛] [even] one time, he will therefore not enter into the hells but promptly he gains rebirth in high heavens.[37]

This evidence tells us that at the time of the Chinese translation of the text the practice of the *niàn fó* 念佛 was already in existence in the cultural milieu in which the text circulated and, so, likely well known by the translator(s). Therefore, it is not implausible to think that the translator(s) of the text adopted the character *niàn* 念 since it implicitly recalls the idea of the Buddha thanks to the widespread practice of the *niàn fó* 念佛. The *Digital Dictionary of Buddhism* (DDB) reports the possibility that the character *niàn* 念 can stand alone with the implied meaning of *niàn fó* 念佛, being a sort of its abbreviation.[38] This fact might suggest to us that for one who is acquainted with the practice of the recollection of the Buddha, the simple reference to *niàn* 念 can somehow recall the whole locution *niàn fó* 念佛 which includes the term *fó* 佛 (= Buddha), given that this is the first among the recollections and one that had great success in the religious market along the

replies that it has both the characteristic of calling to mind (*apilāpana*) and the characteristic of taking hold (*upagaṇhana*)". Here, Gethin is referring to Mil, 37-8.

34 "The recollection of the Buddha is the recollection that arises with reference to the Buddha" (*Buddhaṃ ārabbha uppannā anussati Buddhānussati*; Vism, 197). Essentially, the practitioner has to recollect the qualities of the Buddha as expressed in the famous *iti pi so* formula (Vism, 198-213).

35 In Japan, for instance, this practice is known as *nenbutsu* (念仏). A nice overview on the *Buddhānussati*/*Buddhānusmṛti* is provided by Harrison 1992.

36 卿曹沙門言:"人在世間作惡至百歲,臨欲死時,念佛,死後者皆得生天上。" (T1670B.32.0717b12-13).

37 人雖有本惡, 一時念佛,用是故不入泥犁中, 便得生天上。 (T1670B.32.0717b18-19).

38 See DDB, s.v. "念" (`http://buddhism-dict.net/cgi-bin/xpr-ddb.pl?q=%E5%BF%B5`) in which, incidentally, is reported a quotation from Frédéric Girard: "Abréviation de *nianfo* 念佛, acte d'attention de la pensée".

Silk Road. Thus, we can hypothesise that the Pāli wordplay of *sabuddhika*/*abuddhika* with Buddha is mirrored in the Chinese text by *yǒu niàn* 有念/*wú niàn* 無念 with *fó* 佛, assuming a stretched and creative interpretation for *niàn* 念 as implicitly paired with *fó* 佛. Endorsing this understanding means to assume the existence of a skilful translator, who played with the Chinese characters as much as the Indian creator(s) did with the Indic words.

3 Conclusion

After some consideration regarding the possible audience and author(s) of the text known to us as *Milindapañha* in Pāli and *Nàxiān bǐqiū jīng* (那先比丘經) in Chinese, the relationship between the extant versions has been analysed. Beginning with some guidelines to compare the Pāli and Chinese versions provided by Gérard Fussman, a further guideline was suggested, namely, to take into account that the Chinese versions were written for a Chinese audience. In order to corroborate this point, a passage which involves a pun was analysed, showing that the Chinese translator(s) of the text adapted the translation in order to satisfy the target audience. This fact can of course have some important implications for any attempt to reconstruct the archetype, whose very existence could be questioned on the basis of some recent publications.[39] However, the question would undoubtedly deserve further inquiry. What is clear from the present study is that we can still scrutinise the extant versions comparing similar accounts and reasoning on the differences attested. This effort is not worthless, and the lack of certainty about the existence of the archetype has not negatively affected the knowledge gained by philologically working as if there were one. The findings also have some implications for our comprehension of the translator(s)' strategies in adapting foreign Indian Buddhist literature to the Chinese milieu. In the example taken into account, the Indian word *brahmacārin* conveys the ambiguity on which the pun is based since it means 'celibacy' but literally is 'brahma-conduct', and Brahmā is also a preeminent Indian god. However, we cannot expect that a non-Indian audience would easily grasp the *jeu de mots* and, indeed, a Chinese version of the passage specifies that the term is referring to the 'king of the seventh heaven' (i.e. Brahmā) and to the 'abstention from sexual intercourses' (i.e. celibacy). In this regard, during a potential attempt to reconstruct the archetype, we should assume that the Pāli version conveyed a more reliable reading since the specifications provided by the Chinese version are only necessary for a Chinese audience. It is, indeed, part of the

39 Here, I refer to Levman 2021 and Silk 2021.

very nature of the puns to be understood with ease and immediacy, otherwise not only would the humorous intent not be grasped, but also the general meaning of the passage would remain obscure. Therefore, there is little doubt that a pun in a text from Northwest India was intelligible for an Indian audience and not for a Chinese one. The differences in the exposition concerning the trumpet of the elephant (*koñcanāda*; *míngshēng* 鳴聲) that would make the elephant (*hatthin*; *xiàng* 象) or the bird (*niǎo* 鳥) a follower of the herons (*koñca*) or wild goose (*yàn* 鴈) should be interpreted in a similar way. Also, on this occasion, the Chinese translator(s) adapted the text in order to render the pun in the best way, using the similarity of the radicals of the characters (radical *niǎo* 鳥 in *míng* 鳴 of *míngshēng* 鳴聲 and in *yàn* 鴈). The odd choice of *míngshēng* 鳴聲 to designate the elephant's trumpet may have also influenced the substitution of the elephant (*xiàng* 象) with the bird (*niǎo* 鳥) in version B, assuming that *míngshēng* 鳴聲 would better convey the meaning of a twitter or a chirp than a trumpet. Finally, it has been suggested that the way in which the Indic pun based on *sabuddhika*/*abuddhika* as recalling the word Buddha (thanks to the assonant term *buddhi*) has been aptly rendered into Chinese in a way that preserved the mechanics of the wordplay. The term *niàn* 念 used to translate *buddhi* can similarly recall the Buddha (in Chinese *fó* 佛), due to the widespread practice of the 'Recollection of the Buddha' called *niàn fó* 念佛. Surely, it would seem hard to demonstrate beyond doubt that this is the only univocal interpretation since we cannot check into the mind of the ancient translator(s), but this hypothesis prompts us to ask at least one question: should we let the Pāli version level out our reading of the Chinese text? The analysis of the Buddhist eristic dialogue proposed in the present study introduces us into a new, different scenario, one in which the ancient Chinese translator(s) did not impersonate the role of a dully translator but acted skilfully and creatively in presenting sophisticated foreign puns to his own audience. All in all, is it not the creativeness we find at the very core of any pun? Gérard Fussman is, therefore, certainly right in highlighting the need to take into account the Indian origin of the text. As a logical corollary, we should also pay special attention to the Chinese adaptation and its cultural circumstances.

Appendix

This extremely simplified scheme is based on the reconstruction of the Buddhist cosmology made by Gethin (1997, 195; 1998, 117-8) and adopted also by De Notariis (2019, 66-7). It is not supposed to be comprehensive, but only aims to highlight as in the seventh realm starting from that of the humans the deities begin to be called *brahma*.

	REALM (*bhūmi*)	COSMOLOGICAL SPHERE
	.	World of Pure Form (*rūpadhātu*)
	.	
	.	
	mahābrahmā	
	brahma-purohita	
7	*brahma-pārisajja*	

6	*paranimmita-vasavattin*	World of the Five Senses (*kāmadhātu*)
5	*nimmāṇa-ratin*	
4	*tusita*	
3	*yāma*	
2	*tāvatiṃsa*	
1	*cātummahārājika*	
0	Human Being (*manussa*)	
	.	
	.	
	.	

This scheme is reflected in the Chinese version B (T1670B.32.0705a18-19), in which there is evidence that the first heaven is that of the four great kings (Pāli: *cātummahārājika* = Chinese: *dì yī sì tiān wáng* 第一四天王 'first [heaven] of the four heavenly kings') and the second heaven corresponds to that of the thirty-three [gods] (Pāli: *tāvatiṃsa* = Chinese: *dì èr dāo lì tiān* 第二忉利天 'second [heaven] of the thirty-three heavenly [gods]'). In this regard, see the translation of Demiéville (1924, 89).

Bibliography

Primary Sources and Abbreviations

All Pāli citations are from Pali Text Society Editions.
As = *Atthasālinī*.
D = *Dīghanikāya*.
DDB = *Digital Dictionary of Buddhism*. http://buddhism-dict.net/ddb.
Mil = *Milindapañha*.
Mp = *Manorathapūraṇī* (*Aṅguttaranikāya-aṭṭhakathā*).
Nidd-a I = *Saddhammapajjotikā* (*Mahāniddesa-aṭṭhakathā*).
Ps = *Papañcasūdanī* (*Majjhimanikāya-aṭṭhakathā*).
Ṛgveda = Aufrecht, Th. (ed.) (1877). *Die Hymnen des Ṛigveda*. Bonn: Adolph Marcus.
Sp = *Samantapāsādikā* (*Vinaya-aṭṭhakathā*).
Spk = *Sāratthappakāsinī* (*Saṃyuttanikāya-aṭṭhakathā*).
Sv = *Sumaṅgalavilāsinī* (*Dīghanikāya-aṭṭhakathā*).
T = *Taishō shinshū daizōkyō* (大正新修大藏經). Digital Edition. http://21dzk.l.u-tokyo.ac.jp/SAT. Database version 2015.
Vism = *Visuddhimagga*.

Secondary Sources

Allon, M. (2018). "The Formation of Canons in the Early Indian *Nikāya*s or Schools in the Light of the New Gāndhārī wManuscript Finds". *Buddhist Studies Review*, 35(1-2), 225-44. https://doi.org/10.1558/bsrv.36761.

Anālayo, B. (2021a). "The Opening Debate in the *Milindapañha*". *Sri Lanka International Journal of Buddhist Studies*, 7(2), 15-27.

Anālayo, B. (2021b). "The Scripture on the Monk Nāgasena". *Analysis of the Middle and Extremes & The Scripture on the Monk Nāgasena*. Moraga (CA): BDK America, 105-241.

Baums, S. (2018). "Greek or Indian? The *Questions of Menander* and onomastic patterns in early Gandhāra". Ray, H.P. (ed.), *Buddhism and Gandhara. An Archaeology of Museum Collections*. London; New York: Routledge, 33-46.

Demiéville, P. (1924). "Les versions chinoises du Milindapañha". *Bulletin de l'École française d'Extrême-Orient*, 24, 1-258.

de Jong, J.W. (1996). "Review. Édith Nolot (tr.), Entretiens de Milinda et Nāgasena. Traduit du pâli, présenté et annoté. 385 pp. Paris: Gallimard, 1995. Fr. 196". *Bulletin of the School of Oriental and African Studies*, 59(2), 382-3.

De Notariis, B. (2019). "The Concept of *Manomaya* in Early Buddhism and Upaniṣads. A Study with Particular Reference to the Pāli *Sīlakkhandhavagga*". *Journal of the International Association of Buddhist Studies*, 42, 47-81. https://doi.org/10.2143/JIABS.42.0.3287477.

Fussman, G. (1993). "L'indo-grec Ménandre ou Paul Demiéville revisité". *Journal Asiatique*, 281(1-2), 61-138.

Gethin, R.M.L. (1997). "Cosmology and Meditation. From the Aggañña-Sutta to the Mahāyāna". *History of Religion*, 36(3), 183-217. https://doi.org/10.1086/463464.

Gethin, R.M.L. (1998). *The Foundations of Buddhism*. Oxford; New York: Oxford University Press.

Gethin, R.M.L. (2001). *The Buddhist Path to Awakening*. Oxford: Oneworld.

Gombrich, R.F. (2009). *What the Buddha Thought*. London: Oakville, Equinox.

Guang X. (2007). "The *Nāgasena Bhikṣu Sūtra*. An Annotated Translation from the Chinese Version". *Journal of Buddhist Studies* (*Journal of the Centre for Buddhist Studies, Sri Lanka*), 5, 113-216.

Guang X. (2008). "Introduction to the *Nāgasena Bhikṣu Sūtra*". *Journal of Buddhist Studies* (*Journal of the Centre for Buddhist Studies, Sri Lanka*), 6, 235-51.

Guang X. (2009). "The Different Chinese Versions of the *Nāgasena Bhikṣu Sūtra*". *Journal of Buddhist Studies* (*Journal of the Centre for Buddhist Studies, Sri Lanka*), 7, 226-47.

Harrison, P. (1992). "Commemoration and Identification in *Buddhānusmṛti*". Gyatso, J. (ed.), *In the Mirror of Memory. Reflections on Mindfulness and Remembrance in Indian and Tibetan Buddhism*. Albany (NY): State University of New York Press, 215-38.

Hirakawa, A. (1997). *A Buddhist Chinese-Sanskrit Dictionary*. Tokyo: The Reiyukai.

Horner, I.B. (ed.) (1969). *Milinda's Questions*. 2 vols. London: Luzac & Co.

Kim, K. (2018). "Observations on the Term *Bhavaṅga* as Described in the *Jié tuō dào lùn* (**Vimuttimagga*). Its Proper English Translation and Understanding". *Journal of Indian Philosophy*, 46(4), 753-71. https://doi.org/10.1007/s10781-018-9361-7.

Kubica, O. (2014). "Beyond Influence. A Reflection on the History of Research on the Milindapañha with a Comparison of the Text to the Kitab al Khazari". *Eos. Commentarii Societatis Philologae Polonorum*, 101, 187-206.

Kubica, O. (2021). "Reading the *Milindapañha*. Indian Historical Sources and the Greeks in Bactria". Mairs, R. (ed.), *The Graeco-Bactrian and Indo-Greek World*. London; New York: Routledge, 430-45.

Lamotte, É. (1988). *History of Indian Buddhism. From the Origin to the Śaka Era*. Louvain; Paris: Peeters Press. Transl. of *Histoire du Bouddhisme Indien, des origines à l'ère Śaka*. Louvain: Bibliothèque du Muséon, 1958.

Levman, B.G. (2021). "Revisiting *Milindapañha*". *Journal Asiatique*, 309(1), 107-30. https://doi.org/10.2143/JA.309.1.3289430.

McGovern, N. (2012). "Brahmā. An Early and Ultimately Doomed Attempt at a Brahmanical Synthesis". *Journal of Indian Philosophy*, 40(1), 1-23. https://doi.org/10.1007/s10781-011-9146-8.

Mizuno, K. (1959). "Mirinda-monkyō-rui ni tsuite" 〈ミリンダ問經類について〉 (On the Recensions of Milindapañha). *Komazawadaigaku kenkyū kiyō*, 17, 17-55.

Mizuno, K. (1982). *Buddhist Sutras. Origin, Development, Transmission*. Tokyo: Kōsei Publishing Co.

Mori, S. (1997-98). "The *Milindapañha* and the Pāli Aṭṭhakathā Literature". *Indologica Taurinensia*, 23-4, 291-312.

Nanjio, B. (1883). *A Catalogue of the Chinese Translation of the Buddhist Tripiṭaka. The Sacred Canon of the Buddhists in China and Japan*. Oxford: Clarendon Press.

Nattier, J. (2007). "Indian Antecedents of Huayan Thought. New Light from Chinese Sources". Hamar, I. (ed.), *Reflecting Mirrors. Perspectives on Huayan Buddhism*. Wiesbaden: Harrassowitz, 109-38.

Neelis, J. (2011). *Early Buddhist Transmission and Trade Networks. Mobility and Exchange within and beyond the Northwestern Borderlands of South Asia*. Leiden; Boston: Brill.

Neri, C.; Pontillo, T. (2014). "Words Involving the Stem Brahman. Denoting the Achievement of Super-Human Status in Vedic and Sutta Piṭaka Sources". *Indologica Taurinensia*, 40, 151-94.

Norman, K.R. (1983). *Pali Literature, Including the Canonical Literature in Prakrit and Sanskrit of all the Hinayana Schools of Buddhism*. Wiesbaden: Harrassowitz.

Ooi, E.J. (2021). "Survey of the Pāli *Milindapañha* Manuscripts Kept at the National Library of Thailand. A Brief Catalogue". *Journal of the Siam Society*, 109(1), 169-210.

Rhys Davids, T.W. (ed.) (1890). *The Questions of King Milinda*, vol. 1. Oxford: Clarendon Press.

Rhys Davids, T.W. (ed.) (1894). *The Questions of King Milinda*, vol. 2. Oxford: Clarendon Press.

Rhys Davids, T.W. (1916). "Milinda". Hastings, J. (ed.), *Encyclopædia of Religion and Ethics*. New York: Charles Scribner's Sons, 631-3.

Salomon, R. (2018). *The Buddhist Literature of Ancient Gandhāra. An Introduction with Selected Translations*. Somerville (MA): Wisdom Publications.

Silk, J.A. (2021). "Editing without an Ur-Text. Buddhist Sūtras, Rabbinic Text Criticism, and the Open Philology Digital Humanities Project". *Annual Report of the International Research Institute for Advanced Buddhology*, 24, 147-63.

Skilling, P. (2010). "Problems with Milinda (I). The Opening Verses and Prose of the Printed Siamese *Milindapañhā*". *Journal of the Centre for Buddhist Studies Sri Lanka*, 8, 1-24.

Specht, E.; Lévi, S. (1893). "Deux Traductions Chinoises du Milindapañho". Morgan, D.E. (ed.), *Transactions of the Ninth International Congress of Orientalists (Held in London, 5th to 12th September 1892)*. Vol. 1, *Indian and Aryan Sections*. London: Printed for the Committee of the Congress, 518-29.

Thich M.C. (1964). *Milindapanha and Nāgasenabhikshusūtra. A Comparative Study through Pāli and Chinese Sources*. Calcutta: Firma K.L. Mukhopadhyay.

Trenckner, V. (1879). *Pali Miscellany*, part 1. London: Williams & Norgate.

Trenckner, V. (1880). *The Milindapañho. Being Dialogues Between King Milinda and the Buddhist Sage Nāgasena*. London: Williams & Norgate.

von Hinüber, O. (1996). *A Handbook of Pali Literature*. Berlin; New York: Walter de Gruyter.

Zacchetti, S. (1996). "Il *Chu sanzang ji ji* di Sengyou come fonte per lo studio delle traduzioni buddhiste cinesi. Lo sviluppo della tecnica di traduzione dal II al V secolo d.C.". *Annali di Ca' Foscari. Serie Orientale*, 35(3), 347-74.

e-ISSN 2785-5953

Bhasha
Vol. 1 – Num. 1 – April 2022

Predicative Possessive Constructions in Hindi

Lucrezia Carnesale
Università di Pavia; Università degli Studi Bergamo, Italia

Abstract Hindi has more than one construction for the encoding of possessive relationships: in particular, this language lacks a nominative-accusative construction and a 'have'-verb, and instead encodes possession through intransitive sentences with the Possessor in an oblique case (mainly locative or genitive). This paper describes and analyses the syntactic, semantic, and pragmatic properties of Hindi predicative possessive constructions. It shows that each possessive construction in Hindi is customised to encode particular semantic properties, and that the high iconicity of this language accounts for its lack of a nominative-accusative construction for the expression of possession.

Keywords Hindi. Predicative possession. Locative constructions. Genitive constructions. Iconicity.

Edizioni Ca'Foscari

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

Open access

Citation Carnesale, L. (2022). "Predicative Possessive Constructions in Hindi". *Bhasha*, 1(1), 133-172.

DOI 10.30687/bhasha/8409-3769/2022/01/009

1 Introduction

The expression of possession has been debated for many decades: since the 1980s, several monographs have been published, both typological (Seiler 1983; Heine 1997; Stassen 2009), and language-specific (see for example, Lehmann 2002 for Yucatec Maya; Taylor 1996 for English; Mazzitelli 2015 for Belarusian and Lithuanian). Despite the numerous contributions that have appeared in recent years, the study of possessive constructions continues to present significant analytical challenges.

On the one hand, possession is a fundamental domain of human experience: possessive constructions can be found in all the languages studied thus far and every human being can conceive – even if only intuitively – the difference between 'what belongs to me' and 'what belongs to someone else' (Heine 1997). On the other hand, it is very difficult to define the semantic and pragmatic parameters that lead scholars to collocate under the same label constructions that, from a purely syntactic point of view, have nothing in common. Indeed, as many typological studies on possession clearly show (Heine 1997; Stassen 2009), global language variations reveal a multitude of syntactic configurations expressing the notion of possession. Some languages (mainly SAE languages) use transitive constructions with 'have'-verbs to encode possessive notion; other languages use intransitive constructions. Many Indo-Aryan languages, for example, lack a verb equivalent to English 'have'. They use intransitive constructions with the Possessor marked in the oblique case and the Possessee in the nominative. This happens for example in Punjabi (Shackle 1972), in Bengali (Thompson 2010) and in Marathi (Dhongde, Wali 2009). The same happens in Hindi, where possession is encoded mainly through genitive or locative existential constructions. To better illustrate this point, let us consider the difference among the three sentences below (1)-(3):

(1)	*tya-la*	*tin*	*sədr-e*	*ahe-t*
	he-DAT	three	shirt-M.PL	be-3PL

'He has three shirts'.* (MARATHI)

* The example (including transliteration and glossing) is taken from Dhongde, Wali 2009, 197.

A special thanks to Silvia Luraghi for her precious insights and suggestions. This paper derives from my MA thesis at the University of Rome "La Sapienza" under the supervision of Maria Carmela Benvenuto, Flavia Pompeo and Giorgio Milanetti: my sincere thanks to them too, for their valuable guidance through my years of study. I am also grateful to Andrea Drocco for the long discussions we had and for his many insightful remarks and to the two anonymous reviewers for their comments. Any mistakes are my responsibility entirely.

(2)	*uske pās*	*tīn*	*kamīzeṁ*	*haiṁ*
	3SG.LOC(beside)	three	shirt.F.PL.NOM	to be.3PL.PRS
	'He has three shirts'. (HINDI)			

(3)	*lui*	*ha*	*tre*	*magliette*
	3SG.NOM	to have.3SG.PRS	three	shirt.F.PL
	'He has three shirts'. (ITALIAN)			

Constructions (1) and (2) are intransitive predications in which the first participant is encoded in an oblique case, while the second participant is the syntactic subject; the verb has an existential meaning. Note that in this case there is no lexicalisation of the possessive meaning: the semantics of the construction arises from the global structure of the sentence. The third construction is a transitive predication where the meaning of possession is lexicalised in the verb (*ha* 'has'), the first participant is encoded as a subject, and the second participant is the direct object of the verb. Despite the syntactic differences, these three constructions encode the same possessive meaning.

Further, important problems also arise from a semasiological point of view: what is intuitively identified as a possessive construction can frequently be used to express different semantic notions. To illustrate this point, let us consider some possessive predicates from English, a language with a single possessive verb *have* which covers much more than just the semantics of ownership.

(4) My doctor has a Volkswagen.

(5) Maria has a twin sister.

(6) Everyone has the right to speak.

(7) I have no idea.

(8) That woman has a lot of courage.

As we can see, the single English *have*-construction can express a range of semantic possibilities that goes from *material alienable possession* (see sentence (4), where the notion expressed is that of ownership) to *abstract possession* (sentence (6), where the entity possessed is immaterial) and, through a metaphorical extension, also covers some sub-domains of the *experiential domain* (like *cognition* in sentence (7), where the Possessee is a cognitive status). These extensive semantic uses of the verb *have* are not exclusive to the English language; rather they are quite common (Heine 1997). To further complicate matters, the opposite situation can also occur: a language

can display two or more constructions for the encoding of the possessive domain. In consideration of all these factors, a preliminary distinction needs to be made between genuine possessive constructions (as in the English sentence (4)) and formal possessive constructions (as in the English experiential predicate of sentence (7)), in order to keep distinct the notion of possession as a cognitive domain from possession intended as a linguistic structure (Heine 1997; Langacker 1995; Seiler 1983; Keidan 2008).

In this paper I investigate the syntactic, semantic, and pragmatic properties of Hindi predicative possessive constructions. Hindi has more than one construction for the encoding of possessive relationships, and I will attempt to show that each construction is customised for the encoding of particular semantic properties. The paper is organised as follows: § 2 deals with the formal taxonomies of possessive constructions that have been proposed in the last decades. This paves the way for the exposition of the *construal* of the domain of possession and of the consequences that this conceptualisation has for the linguistic encoding of this notion (§ 3). § 4 analyses the semantic prototype of possessive notions. §§ 5, 6 and 7 present Hindi data from a semasiological point of view and analyse the semantic and syntactic properties of Hindi possessive constructions. Lastly, § 8 draws some conclusions.

2 Formal Distinctions[1]

The first and most basic formal distinction is that between attributive possession (i.e. 'Mark's watch') and predicative possession (i.e. 'Mark has a watch' or 'The watch is Mark's'). Both types of construction are used to express some kind of relation between two entities, but while in the former the relation is presupposed, in the latter it needs to be established. For this reason, attributive constructions consist of a single NP and the relation is internal to it, while predicative possession requires two NPs and the relation is mediated by a predicate. Moreover, while in attributive possession both the Possessor and the Possessee have the same pragmatic role (either the topic or the focus), in predicative possession the Possessor is typically (but not always, as we will soon see) topical (Mazzitelli 2015, 33). The pragmatic difference emerges from the fact that only in attributive possession is the relation already given, while in predicative possession this relationship is established through a predication.

1 For an overview of the formal distinctions proposed on possessive constructions, see Heine 1997, 1-43 and Stassen 2009, 3-36.

The ways in which the relation between the Possessor and the Possessee is predicated lead us to two other major distinctions within the macro group of predicative possessive constructions: that between *ascription of possession* and *predication of belonging* (Heine 1997; Seiler 1983; Lehmann 2002; Stassen 2009),[2] and that between *have-possessives* and *be-possessives* (Isačenko 1974; Heine 1997; Keidan 2009). An *ascription of possession* encodes the relation between the two *relata* from a possessor-oriented point of view: it takes the Possessor as the topical item, while the Possessee is the new information and has the role of focus. In *predications of belonging*, instead, the Possessee is the topic, while the Possessor adds new information: this construction encodes the relationship from a possessee-oriented point of view. Two examples from English are given below (examples (9)-(10)):

(9) ASCRIPTION OF POSSESSION: *Sarah has a red coat*.

(10) PREDICATION OF BELONGING: *The red coat is Sarah's* / *The red coat belongs to Sarah*.

The most important difference between *ascriptions of possession* and *predications of belonging* is in the definiteness and topicality of the two items involved: the presence of an indefinite Possessee and of a topical Possessor seems to be the central characteristic of *ascriptions of possession* (Heine 1997, 30), whereas in a *predication of belonging*, the Possessee is typically definite, it being the topic of the sentence. However, it is worth noting that there is also a relevant semantic difference: as Taylor points out, *predications of belonging* allow "only limited extension from the prototype" (Taylor 1989, 205). This means that while *ascriptions of possession* lend themselves to the expression of a large range of semantic notions (as exemplified in the English sentences (4)-(8)), the use of *predications of belonging* seems to be restricted only to the expression of prototypical possession. We will see, in the paragraph dedicated to the Hindi *belong*-construction (§ 6.3), that although *predications of belonging* do not have the same wide semantic functionality of *ascriptions of possession*, they are not limited to the semantic area of prototypical *ownership*, but can also express different notions.

The last fundamental distinction is that between the two syntactic macro-types of *h-possessives* (i.e. have-constructions) and *e-possessives* (i.e. existential-constructions) (Isačenko 1974; Keidan 2008).

2 Terminology varies from linguist to linguist: for instance, Stassen (2009) uses the terms *indefinite possession* and *definite possession*, while Seiler (1983) and Lehmann (2002) prefer the terms *ascription of possession* and *predication of belonging*. Heine (1997) uses the terms *have-constructions* and *belong-constructions*. In this paper we will opt for the terms *ascription of possession* and *predication of belonging*: Heine's terminology could be confused with another formal distinction proposed in literature: that between *have-constructions* and *be-constructions*.

Have-possessive constructions are transitive-agentive configurations where the semantics of possession is lexicalised in a verb. In *have-possessives* languages, the Possessor and the Possesse are metaphorically related to the prototypical Agent and the Patient of a transitive action: the first is encoded as the subject, while the latter is encoded as the direct object of a transitive verb as in the English sentence above 'Sarah has a red coat'. *E-possessives* constructions instead are intransitive constructions with an existential-locative predicate or a copula, as in the Hindi example below (11). In these constructions, the Possessor is encoded in the oblique case and the Possessee always performs the syntactic role of subject.

(11)	*merī*	*bahan*	*ke pās*	*naī*	*sāṛī*	*hai*
	1SG.GEN.F	sister.F	LOC(beside)	new.F	sari.F.SG.NOM	to be.3SG.PRS

'My sister has a new sari'.

3 The Construal of Possession

Like any other situation, possession needs to be conceptualised before being expressed linguistically. A possessive event always involves at least two participants, the Possessor (henceforth PR) and the Possessee (henceforth PE). Even though these two participants are co-dependent (there can be no PR without a PE and no PE without a PR) they are in an asymmetrical relationship. The asymmetry is both semantic – the prototypical PR has the PE at its disposal and controls it, but not vice versa – and pragmatic – prototypically the PR is topical, while the PE is the focus. Normally, the relationship between a PR and a PE is not perceived through the senses by the speaker – as it happens for example in the case of a location, that can be visually perceived – so possession is a relatively abstract domain, quite complex to conceptualise.

According to cognitive studies (Lakoff, Johnson 1980), the conceptualisation of complex cognitive domains usually takes place through processes of simplification: complex and abstract domains are associated with simpler and more concrete ones through the mental processes of metaphor and metonymy. From a linguistic perspective, this means that the encoding strategy of a complex situation is based on the formal expressions of the concrete domains on which the conceptualisation is based. Furthermore, as has been demonstrated by many studies on the genesis of linguistic expressions (Hopper, Traugott 1993), morphological and syntactic elements are the outcomes of processes of grammaticalisation of lexical items referring to concrete concepts.

According to Heine's model (1997), in the genesis of linguistic expressions of possessive situations the same mechanism of simplification comes into play: the complex and abstract domain of possession

is conceived through more concrete domains. In particular, these simpler domains provide the basis for the emergence of what Heine (1997, 46) calls "event schemas", i.e. conceptual archetypes derived through the abstraction of a number of related events experienced through perception. Heine (1997, 46) states:

> What distinguishes an event schemas from simple concepts in particular is that the former are composed of more than one perceptually discontinuous entity. For example, an event schema like "X EATS Y" typically contains three entities, which are X, EAT, and Y. Simple concepts, on the other hand, consist of no more than one entity, even though they may imply the presence of other entities in addition.

From this definition, it follows that formally an *event schema* has the structure of a proposition (and not of a single lexeme), formed by a predicate and the arguments associated with it; **[tab. 1]** shows the eight *source schemas* theorised by Heine (1997).[3]

Table 1 Summary table of predicative possessive constructions (Heine 1997, 47)

Event schema	**Formula**
Action	X takes Y
Location	Y is located at X
Companion	X is with Y
Genitive	X's Y exists
Goal	Y exists for/to X
Source	Y exists from X
Topic	As for X, Y exists
Equation	Y is X's (property)

According to Heine's proposal, these *schemas* **[tab. 1]** are the conceptual archetypes most used by the languages of the world as the basis for the construal of possession; some examples are given below. One clear instance of the *Action Schema* is the construction with the verb *avere* 'to have' in Italian, as in sentence (12).

(12)	*io*	*ho*	*una*	*macchina*	*nuova*
	1SG	have.1SG.PRS	a	car.F.SG	new.F.SG
	'I have a new car'.				

3 For a detailed analysis of each *schema*, see Heine 1997, 45-76.

In this construction, the PR *io* is encoded as the subject of a transitive predication, and the PE, *una macchina nuova*, as the direct object; thus, the two participants are interpreted as an Agent and as a Patient of an agentive situation. As Givón (2001, 134) points out, this type of construction most commonly emerges as a consequence of the semantic shift of verbs such as 'take, 'grab', 'seize', 'get' which lose their original meaning and acquire a bleached meaning of possession. The same semantic bleaching process characterised the evolution of the Latin verb *habere* 'to have' (etymologically related to the Italian verb *avere*), which derives from the PIE root *$*g^{h}h_{1}b^{(h)}$-(e)i- meaning 'to take' (Baldi, Cuzzolin 2005, 29; de Vaan 2008, 277). Thus, in Italian, possessive relationships are construed through the conceptual archetypes of agentive events. Possession is expressed through a nominative-accusative construction, e.g. the construction associated with the prototypical agentive events: the PR is conceptualised as Agent, and the PE is conceptualised as Patient.

A completely different construal of the possessive event is at the basis of Heine's *Location Schema*. An example of this type of construction is the Hindi sentence in (13): the PR is in the oblique case followed by the postposition *-ke pās* 'beside' and the PE is in the nominative and agrees in number with the predicate.

(13)	*Sītā*	*ke pās*	*naī*	*gāṛī*	*hai*
	Sita	LOC(beside)	new.F	car.F.SG.NOM	to be.3SG.PRS
	'Sita has a new car'.				

In this type of construction, the PR is conceptualised as the Place where the PE is located: the construction is an intransitive predication where the PR stands in the locative case, while the PE is the syntactic subject; the verb has an existential meaning. Note that, as in the case of the Marathi dative construction in example (1), in this sentence there is no lexicalisation of the possessive meaning: the semantics of the construction is projected by the global structure of the sentence.

4 The Semantics of Possession

Although unanimous consensus on the description of possession has not yet been reached, most scholars discuss possession in light of the theory of prototypes (Seiler 1983; Langacker 1995; Heine 1997; Stassen 2009). According to them, the domain of possession consists of several notions hierarchically organised around a prototypical one. Thus, the primary thrust in the analysis of the semantics of possession has focused upon the individuation of the prototypical possessive notion.

In a cognitive theoretical approach, Langacker (2001; 2009) considers possession as a particular instance of the cognitive strategy that he calls "Reference-point strategy". Langacker starts from the assumption that human beings can create mental access to an indefinite entity, the Target, by directing their attention to another more definite entity functioning as a Reference Point. Thus, for example, in NP 'Mark's watch' the watch the speaker refers to is brought to the mind through the evoking of 'Mark' as a reference point: 'Mark' is evoked in order to establish a mental link with 'his watch'. Langacker identifies three prototypes: ownership, kinship and part-whole relationship, and he maintains that their prototypicality is a consequence of the fact that their possessors naturally lend themselves to the reference-point function. In these prototypical cases, the relationship that is used as the basis for the reference-point strategy is "objectively construed", meaning that it exists in the real world; whereas in non-prototypical cases, the reference-point strategy is applied through means of metaphorical or metonymical extensions, and it is "subjectively construed" (Langacker 2009, 84).

Heine (1997) and Stassen (2009), in contrast to Langacker, do not focus their analysis on the cognitive function of the possessive construction, but rather on its semantic parameters. Following an approach that can be reconnected to the semantic binary features approach, they ultimately detect only one prototypical notion. They assume that the fundamental parameters required by a possessive relationship are the control that the PR has over the PE, the proximity between them and the lack of a temporal limit. Following Taylor (1989), Heine proposes a wider range of semantic properties for the individuation of the prototypical notion, adding to the parameters listed above the concreteness of the PE and the humanity of the PR (Heine 1997, 39). Stassen's and Heine's proposals are quite similar: according to them both, *ownership*[4] is the prototypical possessive notion characterised by the maximum control of the PR over the PE, by spatial proximity and by the absence of a temporal limit. The further a

4 Note that the same notion has been identified with different labels: for example, Stassen (2009) uses the term "alienable possession", while Heine (1997) calls it "permanent possession".

possessive notion strays from this prototype, the less it evinces those three fundamental parameters. As a logical consequence of the theoretical differences between Heine and Stassen's approach and Langacker's, the conclusions resulting from their studies differ markedly from those derived from Langacker's model. For example, inalienable possession cannot be considered prototypical, as it is in Langacker's model, since there is a lack of control on behalf of the PR, who cannot decide to break the relationship, and the proximity between the PR and the PE is not necessarily spatial.

In this paper, I follow Heine and Stassen's proposal, and I assume that the notion of *ownership* as defined below is the prototypical one:

> OWNERSHIP
> An asymmetrical relationship between two entities: a PR which must be [+HUMAN] and a PE which must be [-ANIMATE] and [+CONCRETE]. The PR has control over the PE and the relationship has no temporal limit.

The range of notions conceived within the possessive domain varies from scholar to scholar: Heine (1997, 34-40) distinguishes seven possessive notions, e.g.: permanent possession (i.e. ownership), physical possession, temporal possession, abstract possession, inalienable possession, inanimate inalienable possession, inanimate alienable possession; Stassen (2009, 16), in contrast, proposes a conceptual space of four notions: alienable possession (i.e. ownership) inalienable possession, temporary or physical possession, and abstract possession. In this paper, I assume that one of the fundamental properties of possession is the humanness of the PR, and following Stassen (2009, 17) I exclude inanimate possession from my analysis, considering it to be merely a metaphorical extension of possession.

The following paragraphs focus on the Hindi expression of the four possessive sub-domains proposed by Stassen (2009). These possessive notions can be described with reference to the different values they assume in regard to the properties of control, temporal limit and spatial proximity. Specifically, they can be defined as follows:

- Ownership, as defined above.
- Temporary possession and physical possession: the PR can dispose of the PE, even if he/she does not own it, as in the sentence 'I have an apartment where I can stay when I spend the night in London; it belongs to my uncle'. In physical possession, the PR and the PE are physically associated, and the PE is available to be used by the PR even though not belonging to him/her, as in 'I have my sister's keys with me'.
- Inalienable possession: the relationship between the PR and the PE is considered to be inherent, and usually the PE is not a material object but a body-part or a person. Moreover, the PR has

no control in the relationship: the parameter of control implies that the PR can choose when to break the relationship, and in inalienable possession the PR does not have this power. Consider the English example: 'That girl has three brothers'.

- Abstract possession: the PE is an abstract entity or an experiential state like an emotion or a body-sensation, as in the English sentence 'I have a headache'.

5 The Hindi Expression of Possession

Unlike most SAE languages, Hindi does not have a single construction that covers all the semantic notions identified in the previous paragraph (§ 4), exemplified by the English sentences (4)-(8). Instead, it uses different constructions, each one specialised for the encoding of particular semantic features. Let us see the Hindi translation of the English sentences (4)-(8) (sentences (14)-(18)):

(14) My doctor has a Volkswagen.
[HUMAN-MATERIAL ENTITY]: [Ownership] → Locative construction

mere	*ḍākṭar*	*ke pās*	*Volkswagen*	*hai*
1SG.GEN.M	doctor.M	LOC(beside)	Volkswagen.NOM	to be.3SG.PRS

(15) Maria has a twin sister.
[HUMAN-HUMAN]: [Inalienable possession: Kinship] → Genitive construction

Maria	*kī*	*ek*	*juṛvāṁ bahan*	*hai*
Maria	GEN.F	one	twin-sister.F.SG.NOM	to be.3SG.PRS

(16) Everyone has the right to speak.
[HUMAN-ABSTRACT ENTITY]: [Abstract possession] → Dative construction

sab	*ko*	*bolne*	*kā*	*adhikār*	*hai*
everyone	DAT	to speak.INF.OBL	GEN.M	right.M.SG.NOM	to be.3SG.PRS

(17) I have no idea.
[HUMAN-COGNITION]: [Abstract possession: Experience] → Dative construction

mujhe	*khabar*	*nahiṁ*	*hai*
1SG.DAT	information.SG.NOM	not	to be.3SG.PRS

(18) That woman has a lot of courage. [HUMAN-QUALITY]: [Abstract possession: Quality] → Inessive construction

us	*aurat*	*meṁ*	*bahut*	*sāhas*	*hai*
that.OBL	woman.SG.OBL	LOC(in)	a lot of	courage.SG.NOM	to be.3SG.PRS

This behaviour is typical of Hindi, a language exhibiting a number of syntactic patterns selected on the basis of semantic parameters. This peculiarity of Hindi has led Montaut (2004a; 2013) to define these patterns as semantic alignments (following the definition given in Wichmann 2008) rather than syntactic ones. In fact, Hindi clearly encodes semantic roles in a rather iconic way and applies each syntactic pattern to specific semantic features, mainly related to the most salient participant and to the type of the event (Montaut 2004b).

Thus, for example, the use of the transitive-ergative pattern in Hindi is generally restricted to prototypical Agents volitionally acting and controlling the scene. When the Agent does not control the event, another pattern is chosen e.g. the instrumental one.[5] Similarly, when

5 One of the reviewers highlights that the ergative construction in Hindi has many deviations from the principle of iconicity. I agree with him/her that ergative marking in Hindi seems to be partly triggered by syntactic features, however the correlation between ergativity, syntax and semantics in Hindi is not easy to understand. For example, one of the main arguments used to demonstrate that ergativity depends also on syntax is that volitional Agents are not marked by ergative postposition if the predicate is expressed by compound verbs with intransitive light verbs. However, Drocco (2018) and Drocco, Tiwari (2020) showed that when transitive verbs are followed by intransitive light verbs (like *baiṭhnā* or *jānā*) "the meaning conveyed is that the Agent-like argument either acted foolishly, or unconsciously, or lost control over his actions, or was even forced to do something against his wishes" (Drocco, Tiwari 2020, 329). While the Agent argument is not marked with the ergative case, and this seems to be triggered by syntactic properties (since it happens when the light verb is intransitive), we cannot ignore the fact that compound verbs' constructions have also semantic consequences and when the light verb is intransitive the construction seems to express reduced transitivity.

Undoubtedly more work needs to be done to understand the correlation between ergativity and iconicity in Hindi, but many arguments can be given to support the thesis of iconicity. For example, the single argument of a set of 'body emission' predicates can be optionally marked with the ergative case. When this happens, the ergative case-marking encodes a more like prototypical Agent: volitional and in control of the event (Mohanan 1994; de Hoop, Narasimhan 2005). Moreover, there are clearly contrasting examples showing that the ergative marking brings with itself the semantics of agentivity, as opposed to other case markings:

Dative Experiencer vs Ergative Experiencer:

sahsā	*use*	*maṁṛēyā*	*ke sāmne*	*cūṛiyoṁ*	*kī*
suddenly	3SG.DAT	shed	PSP(in front of)	bracelet.F.OBL	PSP.GEN.F
jhaṁkār	*sunāī dī*		*usne*	*kān lagākar*	
tinkle.F.SG.NOM	to hear.PRF.F.SG		3SG.ERG	strain the ear.CVB	
sunā	*hāṁ, koī hē*				
listen.PRF.M.SG	Yes there was someone				

'Suddenly he heard the tinkle of bracelets outside the shed. He strained his ears and listened. Yes, there was someone'.

Instrumental Agent vs Ergative Agent:

A	*tum-hīṁ*	*ne*	*us-kā*	*khūn*	*kiyā*
	2SG-EMPH	ERG	3SG-GEN	blood	to do.PRF.M.SG

the perceiver of a visual or auditory perception is agentive and he/she controls the perception, the pattern selected is the transitive one and the perceiver is encoded as the Agent (in the NOM/ERG). However, when the perception is not controlled by the perceiver, s/he is encoded as an Experiencer e.g. in the dative case: the choice thus evolves from the semantic parameters of the event.

Literature on case (Mallinson, Blake 1981; Comrie 1989; Malchukov 2005; 2015; de Hoop, Narasimhan 2005) generally distinguishes two main functions of case-marking: the so-called indexing function and disambiguating function. The indexing function uses cases to express semantic roles (or specific semantic features of the argument), while the disambiguating function uses cases mostly or exclusively to mark core arguments and express grammatical relations. Following this distinction, Malchukov (2005; 2015) proposes two typological tendencies determining case-marking cross-linguistically:

- Iconicity, which implies the "choice of the most semantically fitting frame" (Malchukov 2005, 85) when encoding semantic roles, thus favouring the indexing function.
- Markedness, which implies the "choice of the transitive frame as a major default pattern" (Malchukov 2005, 85) for the expression of most events, thus favouring the distinguishing function.

Languages of the world vary in their ways of ranking these two parameters. Languages that rank Iconicity over Markedness are more concerned with the faithful encoding of the semantic features of their arguments: these languages tend to not extend the use of transitive constructions to non-transitive events, because in such languages transitive constructions are semantically constrained to prototypical transitivity. In contrast, languages that favour Markedness over Iconicity are more concerned with the differentiation of the two principal syntactic elements (the subject and the object) from peripheral arguments, and therefore they tend to use transitive patterns by default, regardless of the semantic properties of the event.

B	*sāhab*	*maiṁ-ne*	*us-kā*	*khūn*	*nahĩṁ*	*kiyā*		*mujhse*
	sir	1SG-ERG	3SG-GEN	blood	not	to do.PRF.M.SG		1SG.INS
	ho gayā							
	happen.PRF.M.SG							

AB 'It's you who murdered him *Sir, I did not kill him, it happened by myself (I did it unconsciously)*'.
(Example taken from Montaut 2004a, 211)

Unfortunately, a discussion on the correlation between ergativity and iconicity in Hindi is beyond the scope of the present paper; for a thorough investigation upon differential subject marking and indexing function in Hindi the reader can refer to de Hoop, Narasimhan 2005 and Mohanan 1994. For a detailed overview of the study of the interaction between ergativity and semantic transitivity, see Drocco 2008.

With possessive constructions, the interaction between these two parameters determines the typological variability between *have*-possessives and *be*-possessives languages. Languages that rank Markedness over Iconicity are nominative-accusative languages (like English and Italian) that tend to extend the transitive pattern (e.g. Heine's *Action Schema*) to non-transitive events, thus encoding the participants of most events with the nominative and accusative cases, without regard for the semantic properties of the arguments. Languages that rank Iconicity over Markedness, instead, do not extend the transitive construction to encode possessive notions, it being specialised for the encoding of Agentive Events.

In Tsunoda's Implicational Hierarchy of Transitivity (Tsunoda 1985; 2015; **tab. 2**), possessive events are the most distant from the prototypical transitive ones:[6] a prototypical transitive event is dynamic and concrete, characterised by an intentionally acting Agent and by a Patient that is directly affected in a perceptually salient way (Kittilä 2002, 190). A possessive event lacks both of these properties: it has neither an intentionally acting Agent nor an affected Patient, and in fact it is not even a dynamic event, but rather a stative one. This explains why the *Action Schema* is not often employed among the languages of the world since most languages encode possession with intransitive sentences having an oblique PR and a nominative PE. Heine (1997, 75) points out that remarkably only 13.6% of the languages in the world use the *Action Schema* as their major *schema* to express possession. Only highly nominative-accusative languages (like many SAE languages) allow the extension of transitive constructions to stative situations by using transitive verbs such as Eng. *have* or It. *avere* that lexicalise the semantics of possession.

Table 2 Tsunoda's Implicational Hierarchy of Transitivity (Tsunoda 2015, 1598)

Type	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Meaning	**Direct effect on patient**	**Perception**	**Pursuit**	**Knowledge**	**Feeling**	**Relationship**	**Ability**
Examples			search, wait, await	know, understand, remember, forget	love, like, want, need, fond, fear, afraid, angry, proud, boast	possess, have, lack, lacking, resemble, similar, correspond, consist	capable, proficient, good

6 Notably, Malchukov (2005; 2015) does not even include possessive verbs in his two-dimensional Transitivity Hierarchy.

Subtype	1A Resultative	1B Non-resultative	2A Patient more attained	2B Patient less attained
Examples	kill, break, bend	hit, shoot, kick	see, hear, find	look, listen

Montaut (2004a; 2004b) points out that the action model (i.e. transitive pattern) in Hindi is clearly marginal and it is restrained to action processes where the action chain is fully profiled (imperfective aspect). She lists six basic patterns in Hindi which she defines as follows (Montaut 2004b, 51):

1. the nominative accusative diathesis represents action processes;
2. the ergative diathesis encodes action processes but viewed from the viewpoint of the result (aspectual split), and not as an action;
3. the dative diathesis describes experiential processes;
4. the instrumental diathesis describes non-volitional actions in the affirmative and unfeasible actions in the negative, centred on actors lacking some of the features of the agent;
5. the locative and genitive diatheses describe states.

Montaut states that patterns from 2 to 5 are "absolute construals" (as defined by Langacker 1999), where the less salient entity is the starting point from the linguistic viewpoint and the most salient argument is dissociated from the predication and encoded iconically. Following Montaut, I propose that predicative possessive constructions in Hindi are realised as absolute predications where the less salient entity – the PE – is encoded as the subject, while the PR is dissociated from the predication and its case marking is semantically constrained and depends on the semantic properties of the relation. This would explain the variety of constructions that Hindi uses to translate English possessive sentences from (4) to (8): as we have seen each construction expresses a different possessive situation.

6 Presentation of Hindi Data

In the next paragraphs, I propose a semasiological presentation of possessive constructions in Hindi. The examples shown in these paragraphs are taken from a classic of modern Hindi literature, *Godān* by Munshi Premcand, published in 1936. This corpus has been interrogated through SketchEngine.[7]

7 https://www.sketchengine.eu.

As scholars of Hindi know well,[8] this language uses at least two different possessive constructions for the encoding of prototypical possession: both are existential constructions, the first encoding the PR in the oblique case followed by the locative postposition *-ke pās* 'beside', and the second encoding it in the genitive. First, the locative construction (§ 6.1), which can express the notion of ownership but can also express physical possession and temporary possession, is discussed. In §§ 6.2-6.4, an analysis of the genitive pattern, which can express the prototypical notion of ownership, and which is more frequently used for the encoding of inalienable relationships, is presented. In § 6.5 other non-prototypical uses of these constructions are considered. Finally, § 7 focuses on the analysis of the dative construction and inessive constructions, which can express only abstract possession and do not allow the encoding of more prototypical notions.

6.1 The Locative Construction

The pattern of the locative construction is as follows: the PR NP is in initial position, marked in the oblique case and followed by the postposition *-ke pās*, which means 'beside' and is normally used for the encoding of locations; next comes the PE, marked in the nominative case; and in the final position there is the verb *honā* 'to be', which agrees in number and person (in past tenses also in gender) with the PE and has an existential function. This construction can be schematised as follows: *Y is at X's place* > *X has, owns Y* and can be associated with Heine's *Location Schema* (Heine 1997, 51).

Before continuing with the exposition, it is essential to briefly examine the use of the compound postposition *-ke pās* to better understand the examples and their glosses. Like any other compound postposition in Hindi, the postposition *-ke pās* is composed of the simple genitive postposition (in this case *-ke*) followed by an adverb (in this case *pās* 'near'). When compound postpositions follow a noun, they are attached to its oblique form, as in example (19), but when they follow a personal pronoun, the possessive form of the pronoun, rather than its oblique form followed by the genitival postposition, is required, as in sentences (20)-(21).

(19) *Mehtā* *ke pās* *sāmān* *to jyādā* *na* *thā*
Mehta LOC(beside) belongings.M.SG.NOM many not to be.3SG.PST.M
'Mehta didn't have many belongings'.

8 Kachru 1970; Montaut 1997; 2004a; Mohanan 1994; Pandharipande 1981.

(20) *lekin* *mere pās* *nagad* *nahīṁ* *hai*
but 1SG.LOC(beside) cash.SG.NOM not to be.3SG.PRS
'But I have no cash'.

(21) *hamāre pās* *jo* *kuch* *hai* *vah*
1PL.LOC(beside) REL.ADJ.DIR something.NOM to be.3SG.PRS CRR.PRN.NOM
abhī *khalihān* *meṁ* *hai*
now barn PSP(in) to be.3SG.PRS
'What we have now is in the barn'.

Note that the possessive construction with the postposition *-ke pās* is formally identical to the Hindi locative construction. In truly locative constructions, the location argument is in the oblique case followed by the postposition *-ke pās* (as the PR in the possessive construction), with the entity located appearing in the nominative case (as the PE) and the predicate being the existential verb *honā* 'be'. The most important difference between these two sentence-types is the semantics of their two arguments: when the argument preceding the postposition *-ke pās* is [+HUMAN] and the second argument is [-ANIMATE], the resulting construction is a possessive one. Remarkably, the semantics of possession is not lexicalised in a lexical item, but rather it emerges from the instantiation of the locative construction through these specific semantic features. If these features are absent, then the resulting construction has a locative meaning. See the examples below (22a)-(22d):[9]

(22a) 1° argument: [+HUMAN]; 2° argument: [-ANIMATE]: Possession
uske pās *qalam* *hai*
3SG.LOC(beside) pen.SG.NOM to be.3SG.PRS
'He has a pen'.

(22b) 1° argument: [- HUMAN]; 2° argument: [- ANIMATE]: Location
kitāb *ke pās* *qalam* *hai*
book LOC(beside) pen.SG.NOM to be.3SG.PRS
'Next to the book, there is a pen'.

(22c) 1° argument: [- HUMAN]; 2° argument: [+HUMAN]: Location
gāṛī *ke pās* *Sītā* *hai*
car LOC(beside) Sita.NOM to be.3SG.PRS
'Next to the car, there is Sita.'

9 Examples (22a)-(22d), (23a)-(23b), example (43) from § 6.4 and examples (52)-(53) from § 7.1 are not taken from the corpus.

(22d) 1° argument: [+ HUMAN]; 2° argument: [+HUMAN]: Location

Sītā	*ke pās*	*merī*	*bahan*	*hai*
Sita	LOC(beside)	1SG.GEN.F	sister.F.SG.NOM	to be.3SG.PRS

'Next to Sita, there is my sister'.

Some other semantic and syntactic features also differentiate the possessive construction from the locative. In particular:

1. Only in a locative construction can the postposition *-ke pās*, normally meaning 'beside', be exchanged with some other synonymic postpositions, like *-ke bagal meṁ* or *-ke nikaṭ* 'next, near to'. The possessive construction does not allow for the interchangeability of *-ke pās* with other locative postpositions; if another locative postposition is selected, the resulting construction acquires an existential-locative meaning. See the contrasting examples below (23a) and (23b).

(23a)

Rām	*ke pās*	*naī*	*kitāb*	*hai*
Ram	LOC(beside)	new.F	book.F.SG.NOM	to be.3SG.PRS

'Ram has the new book'.

(23b)

Rām	*ke bagal meṁ*	*naī*	*kitāb*	*hai*
Ram	LOC(beside)	new.F	book.F.SG.NOM	to be.3SG.PRS

'Next to Ram there is the new book. *Ram has the new book'.

This phenomenon may be explained by the fact that the locative construction instantiated with a [+HUMAN] location and a [-ANIMATE] second argument has undergone a grammaticalisation process causing the desemantisation of the postposition *-ke pās*, which, in this context, has lost its original lexical meaning.

- Only in a possessive sentence is the element preceding the postposition *-ke pās* endowed with some non-nominative subjects' properties. First, in non-pragmatically marked possessive constructions, the PR is in initial position, whereas in non-marked locative constructions, the element followed by the postposition *-ke pās* is preverbal and the subject is in initial position. Secondly, only the PR governs coreference with the reflexive pronoun *apnā*: see the example (24).

(24)

mere pās	*apne*	*dost*	*kī*
1SG.LOC(beside)	REFL.ADJ.M.SG.OBL	friend.M.SG.OBL	GEN.F
kitāb	*hai*		
book.F.SG.NOM	to be.3SG.PRS		

'I have my friend's book'.

As mentioned in § 2, the prototypical information structure of an *ascription of possession* requires a topical PR with the PE as the comment. As Keidan (2008, 349) points out, "the topicality we are concerned with here belongs primarily to the cognitive domain, not simply to the grammatical level". An *ascription of possession*, indeed, encodes the relation between the two *relata* from a possessor-oriented point of view, taking the PR as the starting point for the predication of the relationship. Languages use different ways to mark the topicality of the PR: *have-possessives* languages, for example, raise it to the syntactic status of subject. In these languages, the PR is marked as a nominative and combines various syntactic properties of subjecthood (Kibrik 1997; Onishi 2001). As noted above, Hindi is a highly iconic language: it uses cases to faithfully encode the thematic properties of the most salient element. Moreover, as Montaut (2004b, 51) points out, in Hindi

> the profiled segment always leaves the cognitively more salient entity in a secondary position, so that the less salient entity is the starting point from the linguistic viewpoint. Hindi indeed shows a clear preference for profiling less salient entities as starting points in asymmetric relations.

This means that while Hindi encodes the more salient entity in the sentence through an iconic use of case marking, it assigns the nominative to the less salient entity by default. In *ascriptions of possession*, the PR is always more salient than the PE, as it is prototypically [+HUMAN] and the topical element. Consequently, in Hindi possessive sentences the syntactic properties of subjecthood are split between the PR and the PE: this explains why the PR, even if marked as a locative, is endowed with such syntactic properties as initial position in the unmarked sentence and the control of coreference with reflexive pronouns and adjectives.[10]

Let us now move on to the semantics of this type of possessive construction. As already mentioned above, the locative construction with the postposition *-ke pās* can express the notions of ownership, temporary possession, and physical possession. Examples are given in sentences (25)-(27). This pattern is thus characterised by a certain degree of ambiguity: it is only the context that helps us to understand what type of possessive notions the construction is encoding.

10 Remarkably, Montaut (2004b, 51) points out that "full subjecthood is restricted in Hindi/Urdu to action phrases and single arguments of simple verbs". Moreover, discussing the notion of subject in Hindi, Drocco (2008, 40-1) points out that "l'analisi relativa alla determinazione del soggetto in hindī è stata infatti effettuata basandosi non tanto sulle proprietà di codifica, bensì sulle proprietà relative al controllo dei diversi processi sintattici" (Author's transl.: 'The study of the notion of subjecthood in Hindi has been carried out not through the analysis of the coding properties of the argument, but through the analysis of its behavioral properties: i.e. through syntactic tests').

(25) *hamāre pās* *ilāke,* *mahal,* *savāriyāṁ*
1PL.LOC(beside) land.M.PL.NOM palace.PL.NOM carriage.F.PL.NOM
naukar-cākar *haiṁ*
servant.PL.NOM to be.3PL.PRS
'We have lands, palaces, carriages, servants'.

(26) *hamāre pās* *becne* *ko* *bhūsā* *nahīṁ*
1PL.LOC(beside) to sell.INF.OBL PSP(to) straw.SG.NOM not
hai
to be.3SG.PRS
'We have no straw to sell'.

(27) *jiske pās* *jo* *kuch* *ho,*
REL.PRN.LOC(beside) REL.ADJ.DIR INDF.PRN.NOM to be.3SG.SBJV
nikālkar *rakh de*
take.out.CVB to put.3SG.SBJV
'Take out what you have and put it here'.

6.2 The Genitive Construction

As noted by many scholars of Hindi (Caracchi 2002; Pandharipande 1981; McGregor 1972; Mohanan 1994), in this language the notion of *ownership* can also be expressed by a genitive construction. The pattern of this construction is as follows: the PR is in initial position and is marked in the genitive case, the PE is in preverbal position marked in the nominative case and the predicate is expressed by the existential verb *honā* 'be'. The verb agrees in number and person (in gender in past tenses) with the PE. Note that the PR is marked in the oblique case and followed by the genitive postposition *-kā* (*/-ke/-kī*) which agrees in gender and number with the PE, thus forming an adjectival unit with the PR: see examples (28a)-(28c). In particular, the genitive form *-kā* is the masculine singular form, while the masculine plural is *-ke*; the feminine form is *-kī* and it is the same for both the singular and the plural. As in the case of the locative postposition *-ke pās*, when the genitive postposition *-kā/-ke/-kī* follows a personal pronoun, the possessive form of the pronoun is required, as exemplified in (28d) and (28e).

(28a) *bacc-e* *k-ā* *ḍibb-ā*
child.M.SG.OBL GEN.M.SG.DIR box.M.SG.DIR
'The child's box'.

(28b) *bacc-e* *k-e* *ḍibb-e*
child.M.SG.OBL GEN.M.PL.DIR box.M.PL.DIR
'The child's boxes'.

(28c) *bacc-e* *k-ī* *kitāb*
child.M.SG.OBL GEN.F book[F.SG]
'The child's book'.

(28d) mer- ā ḍibb-ā
1SG.GEN.M.SG.DIR box.M.SG.DIR
'My box'.

(28e) *mer-ī* *kitāb*
1SG.GEN.F book[F.SG]
'My book'.

This construction corresponds to Heine's *Genitive Schema*, summarised in the formula: *X's Y exists > X has Y* (Heine 1997, 58). An example is given below (29):

(29) *unkī* *tīn* *laṛkiyāṁ* *thīṁ*
3PL.GEN.F three daughter.F.PL.NOM to be.3PL.PST.F
'He had three daughters'.

Notably, as in the case of locative constructions, the PR marked with the genitive case acquires some syntactic properties of subjecthood: it is always in the initial position, and it controls the coreference with the reflexive pronoun and coreferential deletion; see the example (30) taken from Montaut (2013, 93).

(30) *merā* *apnī* *bahan* *se* *milne* *dillī*
1SG.GEN.M.SG REFL.F sister COM to meet.INF.OBL delhi
jāne *kā* *irādā* *thā*
go.INF.OBL GEN.M.SG intention.M.SG.NOM be.PST.M.SG
'I intended to visit my sister in Delhi'. (Lit. 'I had the intention to visit my sister in Delhi')

In many contexts, genitive constructions and locative constructions are semantically interchangeable (Mohanan 1994, 178): a sentence of the type 'That man owns a huge house' can be translated into Hindi with either a genitive construction (example 31) or a locative construction (example 32):

(31) *us* *admī* *kā* *ek* *bahut* *baṛā*
that.OBL man.SG.OBL GEN.M.SG one very big.M.SG.DIR
makān *hai*
house.M.SG.NOM to be.3SG.PRS

(32) *us* *admī* *ke pās* *ek* *bahut* *baṛā*
that.OBL man.SG.OBL LOC(beside) one very big.M.SG.DIR
makān *hai*
house.M.SG.NOM to be.3SG.PRS

McGregor (1972, 51) suggests that there is a semantic difference between the two constructions: while the genitive construction expresses a permanent possessive relationship, the locative pattern is used for more contingent relationships, e.g. what has been identified here as temporary and physical possession. Even if it is true that the locative construction can be used to express non-permanent possession (as noted in § 6.1), some scholars have shown that such a semantic distinction is not applicable.

Mohanan (1994, 179), for example, points out that both constructions can be modified by a subordinate clause of the type "which he is trying to sell" thus implying that the genitive construction can also express permanent possession. In the same way, both sentences can be modified by the clause "which he will hand down to his children", thus implying that locative constructions, too, can encode permanent possession. Pandharipande (1981) suggests that the selection of the genitive construction for the encoding of alienable possession is determined by how the relation between the PR and the PE is perceived by the speaker. A peculiarity of the genitive construction is that most characteristically it involves concrete entities as PE (like *estates*, *buildings* and *lands*) that are normally perceived as being less alienable than the entities frequently involved in locative constructions (like *money*, *books*, etc.). Moreover, the genitive construction expressing *ownership* typically occurs when the PE is perceived to be particularly close to the personal sphere of the PR; in this regard, notice the contrasting following examples (33)-(34).

(33) *uske* *maurūsī* *pāṁc* *bīghe* *khet* *haiṁ*
3SG.GEN.M.PL inherited five bighe field.M.PL.NOM to be.3PL.PRS
'He has an inherited field of five bighe [Indian unit of measure]'.

(34) *hamāre pās* *ilāke*, *mahal*, *savāriyāṁ*
1PL.LOC(beside) land.M.PL.NOM palace.PL.NOM carriage.F.PL.NOM
naukar-cākar *haiṁ*
servant.PL.NOM to be.3PL.PRS
'We have lands, palaces, carriages, servants'.

In sentence (33), the speaker is answering another character who asked him whether his family owns land. In his answer, the speaker chooses to express the notion of ownership through a genitive construction: the PR is a peasant family who inherited the PE – a land – and

who is particularly attached to it, having cultivated it for generations. In sentence (34), the PR is a *zamīndār*. *Zamīndārs* were large landowners who did not have the same attachment to the land, as it was worked by others. Consequently, in this context the notion of ownership is expressed by a locative construction. Thus, the genitive construction is generally associated with a more intimate possessive relationship, in which the PR is perceived as emotionally attached to the PE; on the other hand, the locative construction seems to be used for the expression of mere legal ownership. Sentence (35) offers a further interesting example:

(35) *Agar vah ek bīghā bhī bec de, to sau mil jāyaṁ, lekin kisān ke liye jamīn jān se bhī pyārī hai, kul-maryādā se bhī pyārī hai;*
'If he sold even one bigha of land, he could get a hundred rupees. But to a peasant, land is dearer than life, dearer even than family reputation;'

aur	*kul*	*tīn*	*hī*	*bīghe to*	*uske pās*	*hai,*	*agar*
and	total.NOM	three	just	bighe	3SG.LOC	to be.3SG.PRS	if
ek	*bīghā*	*bec de*		*to phir*	*khetī*	*kaise*	*karogā*
one	bigha	to sell.3SG.SBJV		then	cultivation	how	to do.3SG.FUT

'And he had just three bighe of land, if he were to sell one bigha, how could he live off the land?'

In sentence (35), the PR is a peasant whose family has been living under the poverty threshold for a while, and who is considering the idea of selling a part of his land to make some money. However, his biggest concern derives from the emotional attachment he has with the land, notably the author even tells us that 'to a peasant, land is dearer than life'. The context makes it clear that the PE is here felt as strongly connected with the emotional sphere of the PR. Nonetheless, possession is here expressed through a locative construction and not through a genitive one. One might think that this example weakens the argument according to which the choice of the genitive is based on the intimacy of the possessive relationship, while the locative is used to encode mere legal ownership. However, note that the PR here is thinking to sell the land, so if on the one hand the PE is felt as intimately connected with the PR, on the other hand it is also conceptualised as a mere legal ownership, and that could explain the choice for the locative marking on the PR.

However, these semantic explanations are not always adequate to explain the preference of a construction over the other: sometimes the choice seems to be random. Consider the example in sentence (36), in which the speaker encodes three consecutive possessive constructions. The PR is always the same and it is encoded with the first-person plural pronoun, while the PEs are three different material entities: a land in the first construction, the crop of that land in the second construction, and money in the third construction. The possession of land,

that should be felt more intimate and less alienable that the possession of the crops is encoded through a construction that marks the PR as a locative, while the possession of the land's crop which should be far more alienable and less connected with the emotional sphere of the PR is encoded through a genitive construction.[11]

(36) *utne hī* *khet to* *hamāre pās* *bhī* *haiṁ.*
just as much.M.PL field.M.PL.NOM 1PL.LOC(beside) too to be.3PL.PRS
utnī hī *upaj* *hamārī* *bhī* *hai.* *phir*
just as much.F crop.F.SG.NOM 1PL.GEN.F too to be.3PL.PRS then
kyoṁ *hamāre pās* *kaphan* *ko* *kauṛī* *nahīṁ* *aur*
why 1PL.LOC(beside) shroud PSP(for) cent.SG.NOM not and
unke *ghar* *naī* *gāy*
3PL.GEN.M house[M.SG.OBL] new.F COW.F.SG.NOM
ātī hai
to come.3SG.PRS.F
'We own field of the same size as his, and we have crops as good as his. Then how come that we don't even have a cent to buy a shroud, while they have a new cow in their house?'

Lastly, consider the following example, in which once again the choice of the genitive construction seems not to be determined by an emotional attachment of the PR to the PE. See sentence (37):

(37) *"Mere sir meṁ jor kā dard ho rahā hai. ādhā sir esā phaṭā paṛtā hai, jaise gir jāyagā." Mehtā ne ākar kahā* […] *"Tumhāre sāth koī davā bhī to nahīṁ hai?" "Kyā maiṁ kisī marīj ko dekhne ā rahī thī, jo davā lekar caltī?"*
'"I have got a terrible headache. My head's bursting as half of it were about to drop off." Mehta walked over to her and said, "[…] Don't you have any pills with you?" "Was I supposed to be visiting a patient? Why should I have brought any pills?"'
merā *ek* *davāoṁ* *kā* *baks*
1SG.GEN.M a medicine.F.PL.OBL GEN.M box.M.SG.NOM
hai *vah Semrī meṁ hai*
to be.3SG.PRS it is in Semri
'I do have a box of medicine, (it is in Semri)'.

Example (37) is interesting because the reason for the use of a genitival construction here seems to be pragmatic. As the context makes clear, the PR – a doctor who is suffering from a bad headache – is asked whether she has a box of medicine with her or not. The speaker an-

11 Note that the sentences in example (36) are pragmatically marked, for this reason the order of constituents is here PE-PR-V.

swers that she *owns* a box of medicines, but unluckily she does not have it with her (she left it in Semri). As we have seen in § 6.1, the locative construction with the postposition *-ke pās* is characterised by a certain degree of ambiguity, since it allows the expression of all types of possession – temporary possession, physical possession, and ownership. Here, the speaker chooses to use the *genitive construction* in order to avoid any ambiguity: she wants to emphasise that the possession she is expressing is not of the physical type. Thus, it must be noted that while grammatical tradition of Hindi has mainly highlighted the semantic implications of the use of the genitive construction, sometimes the choice of this type of structure is influenced by pragmatic factors.[12]

6.3 Predications of Belonging in Hindi

As mentioned above, one fundamental distinction holds between *ascriptions of possession* (or *have*-constructions: i.e. 'I have a new sari') and *predications of belonging* (or *belong*-constructions: i.e. 'The new sari is mine'). The difference between these two types is pragmatic and primarily depends on the information packaging of the sentence. Thus far, possessive constructions where the PR is the topical element and the PE the comment have been discussed, i.e. how Hindi encodes *ascriptions of possession* with the possessive relationship profiled from the point of view of the PR. We now turn our attention to the ways in which Hindi expresses a possessee-oriented relation.

This pragmatic distinction has been claimed to be cross-linguistically valid (Heine 1997): every language has constructions that encode possessive relationship from the inverse perspective i.e. from the point of view of the PE. However, while some languages mark the difference between *belong*-constructions and *have*-constructions by lexical and syntactic means, other languages do not clearly encode this distinction. In English, for example, the verb *have* is used for a possessor-oriented expression, while the verb *belong* and the construction *X is Y's* are used to encode a possessee-oriented expression. In contrast, Hindi does not distinguish these two types of sentences through such lexical

12 One of the anonymous reviewers suggests that the genitive construction here is used to imply a particular connection with the personal sphere of the PR, and therefore to underline the intimate relation between the PR – a doctor – and the PE – a box of medicines (*davāoṁ kā baks*). I do not fully agree with this interpretation: even if the PR is a doctor and this could imply that she feels the PE as closer to her personal sphere, I believe that in this specific case the reason behind the use of a genitive construction instead of a locative one is to avoid any ambiguity: Dr Malti is saying that she has her medical box but at the moment it is not with her. In a non-marked context, a doctor would probably use a locative construction to encode the possession of a box of medicine and s/he would not map this relation as inalienable.

and syntactic strategies. In this language, *belong*-constructions have the following structure: the PR is marked with the genitive case and the PE is in the nominative; the verb, once again, is *honā* 'to be'. This construction can be associated with Heine's *Equation Schema* and it is summarised with the following formula: *Y is X's* (*property*) > *Y belongs to X* (Heine 1997, 65). Some examples of *predications of belonging* in Hindi are given in sentences (38) and (39):

(38) *jis* *makān* *meṁ* *rahtā hūṁ,* *vah*
REL.ADJ.OBL house.OBL LOC(in) to live.3SG.PRS.M CRR.PRN.NOM
ab *merā* *nahīṁ* *hai*
now 1SG.GEN.M not to be.3SG.PRS
'The house I am living in now does not belong to me'.

(39) *Jhuniyā* *ab* *hamārī* *ho gaī*
Jhuniya.NOM now 1PL.GEN.F become.1SG.AOR.F
'Now, Jhuniya has become ours (daughter)'.

Clearly, this construction is quite similar to the genitive construction discussed in § 6.2 and used to encode *ascriptions of possession*, but there are some fundamental differences.

1. First, while in *ascriptions of possession* the verb *honā* has an existential meaning, in *predications of belonging* it has a copular function: it only connects the PE to the PR and defines the tense and mood of the relationship. Indeed, like many other Indo-European languages, Hindi uses one verb, *honā* 'to be', for two major functions, the copular and the existential-locative.
2. Second, in predications of belonging, the PR is not endowed with any syntactic properties of subjecthood (e.g. initial position, control of reflexive pronouns and adjectives, control of coreferential deletion) as it is in *ascriptions of possession*. So notably while in *ascriptions of possession* the unmarked order of the constituents is PR-PE-V, in *predications of belonging* the unmarked order is PE-PR-V. What accounts for this is that in *ascriptions of possession*, the raising of the syntactic status of the oblique PR is a consequence of its topicality, but as previously noted, *belong*-constructions are characterised by an inverse informational structure, where the PE is the topic element. Notice that as a consequence of the topicality of the PE, in *belong*-constructions the PE is always definite.
3. Third, in belong constructions the PE is *always* definite and known, while *ascriptions of possession* can encode also possessive constructions in which the PE is indefinite and unknown.

e-ISSN 2785-5953

According to some scholars (Taylor 1989; Heine 1997), there is a strong correlation between *predications of belonging* and the concept of *ownership*: that is, *predications of belonging* do not share the same wide semantic extension that *ascriptions of possession* do, and they disallow the expression of such notions as temporary possession and physical possession. From the analysis of the corpus considered here, this correlation does appear to hold in part: when the construction involves concrete PEs, generally the notion encoded is that of ownership. However, evidence from Hindi tells us that this construction can also express other notions.

From the analysis of the semantic features of PR and PE of Hindi constructions found in the corpus, it emerges that while the PR is always [+HUMAN], the PE can be [±HUMAN], [±ANIMATE] or [±CONCRETE]: so, this construction can encode other expressions beside ownership. In particular, of the 23 occurrences of this construction in the corpus, a third encode kinship or social relationships (see example (39)), only two instances encode abstract possession (example (40)), and all the other instances are expressions of ownership of concrete entities (like *house*, *bank*, *assets*, as in sentence (38) above) and of animals (sentence (41)).

(40) *kānūn* *aur* *nyāy* *uskā* *hai*,
law.M.SG.NOM and justice.M.SG.NOM CRR.PRN.GEN.M to be.3SG.PRS
jiske pās *paisā* *hai*
REL.PRN.LOC(beside) money.SG.NOM to be.3SG.PRS
'Law and justice belong to the one who has money'.

(41) *gāy* *merī* *hogī*
COW.F.SG.NOM 1SG.GEN.F to be.3SG.F.FUT
'The cow will be mine'.

6.4 Genitive Constructions and the Expression of Inalienability

Hindi grammars (Hook 1979; Kachru 2006; Montaut 2004a; Milanetti, Gupta 2008) systematically associate the genitive construction with the notion of inalienability: this pattern is used to express intimate and inherent relationships, like kinship and body-part relationships. The locative construction, on the other hand, bears no such meaning. Examples for the encoding of kinship relationships and body-part relationships follow in (42) and (43):

(42) *unkī* *tīn* *laṛkiyāṁ* *thīṁ*
3PL.GEN.F three daughters.F.PL.NOM to be.3PL.PST.F
'He had three daughters'.

(43)	*us*	*laṛkī*	*kī*	*nīlī*	*ankheṁ*	*haiṁ*
	that.OBL	girl.F.SG.OBL	GEN.F	blue.F	eyes.F.PL.NOM	to be.3PL.PRS
	'That girl has blue eyes'.					

The terms *alienable* and *inalienable* (and their synonymic alternatives *inherent* vs. *established*; *separable* vs. *inseparable*) appear very frequently in the literature on possession. The term 'inalienability' indicates that the relationship between the two *relata* is conceived as inherent or indissoluble, like, for example, the relationship between family members and the relation between a body part and its possessor. The contents of the class of inalienable entities vary from culture to culture. However, Stassen (2009, 17) points out that when a language has a unique encoding for inalienability, "this encoding will almost always cover at least the relation between a 'possessor' and his or her body parts, and/or the relation between a 'possessor' and the members of his or her kinship circle", thus suggesting that these relationships are generally considered prototypical examples of inalienability. The fact that these two types of relationship seem to form the core of inalienable possession can be explained by the fact that body parts and family members are relational entities in the real world. Further extensions of inalienable encoding may then vary from culture to culture; in Hindi, for example, not only are blood-kinship relations seen as inalienable, but also intimate social relationships such as those with a friend or a spouse are usually encoded with the genitive construction. Professional relationships, in contrast, are not viewed as inalienable and are generally encoded with locative constructions (see examples (44) and (45), § 6.5).

Over the entire set of 21 sentences with genitive constructions found in the corpus, 15 were identified as expressions of inalienability and only six sentences conveyed the meaning of ownership. Moreover, it must be noted that no other syntactic schema can express inalienable relationships in Hindi. Considering this, it appears clear that genitive constructions have the unique ability to codify this notion; in this regard, Hindi aligns with the typological data presented by Heine (1997, 67): "in a number of languages, the *Genitive Schema* provides the primary means of expressing inalienable possession".

Recall from § 6.2 that Hindi genitive postposition forms an adjectival structure: the genitival postposition *-kā* (*/-ke/-kī*) is attached to the PR in the oblique case and agrees in gender, number and case with the PE, as exemplified in the examples (28a)-(28e). It is significant that the genitive construction is used to encode only inalienable possession, while alienable possession in Hindi is typically expressed through locative construction and does not require agreement. Discussing the use of adjectival constructions for the encoding of inalienability in Sanskrit, Viti (2004) points out that the category of adjective is generally

used to express inherent and permanent properties. Moreover, she remarks that the adjective is a relational category both from a syntactic point of view – an adjective cannot occur without a head-noun – and from a semantic point of view – the value of an adjective depends on the noun it modifies. The same can be said about inalienable relationships: inalienable relationships are seen as inherent and permanent, and the inalienable PE is prototypically a relational entity (a family member or a body-part). In short, the fact that the genitive construction uses a postposition that agrees in gender and number with the PE seems to be emblematic of the type of relationship that exists between the two *relata*. We can thus conclude that the Hindi genitival construction is iconic: the relational nature of inalienability is formally encoded through the use of the relational category of adjectives.

Given the fact that the genitive construction is specialised for the encoding of inalienability, and given the high iconicity of Hindi (§ 5), one might wonder why this construction allows the expression of non-inherent and non-inalienable relationships as ownership. However, recall from § 6.2 that, when expressing ownership, the genitive construction is generally associated with a more intimate possessive relationship. Moreover, as Pompeo (2010, 42) points out, in some contexts the notion of *ownership* is similar to that of inalienable possession in many respects: the relationship between the *relata* is particularly strong and it exists even without spatial proximity. Additionally, both relationships require an exclusive association between the PR and the PE. Thus, in Hindi, the use of the genitive construction to express *ownership* mirrors the similarity between this possessive notion and inalienable relationships. Once again, the choice of the syntactic pattern is dependent on the semantic properties of the event.

6.5 Other Uses of the Locative and Genitive Constructions

The constructions analysed in the previous paragraphs can also serve other semantic purposes. As in many other languages, Hindi possessive constructions can be used to encode non-possessive meanings, owing to the mental processes of metaphoric or metonymic extension. In the next paragraph, semantic extensions of the locative construction will be considered first and afterwards the semantics of the genitive construction.

The locative construction can be used to express professional relationships, as in example (44), 'We have servants'. Notice that if the relationship is not of a professional type, but a more general social relationship (as in 'I also have a friend', in example (45)) the use of the locative construction is disallowed, and the genitive construction is employed instead. It is worth noting that once again, the parameters that influence the choice between the genitive and the locative

constructions are the emotional attachment and the intimacy of the relationship.

(44) *hamāre pās* *naukar-cākar* *haiṁ*
1PL.LOC(beside) servant.M.PL.NOM to be.3PL.PRS
'We have servants'.

(45) *merā* *bhī* *koī* *hitū* *hai*
1SG.GEN.M too INDF.ADJ.DIR friend.M.SG.NOM to be.3SG.PRS
'I also have a friend'.

Furthermore, the locative construction can sometimes be used to express some metaphorical possessive notions. Specifically, it is allowed in expressing the possession of abstract entities such as *answer* in example (46) or *time* in example (47); notably, however, these uses are conventionalised and do not constitute systematic phenomena. They probably derive from metaphorical conceptualisations of abstract entities as concrete and material ones. For example, according to Lakoff and Johnson (1980), the metaphor TIME IS MONEYoccurs quite frequently across cultures and languages: it is present in English, in Italian and also in Hindi.

(46) *dhaniyā* *ke pās* *javāb* *taiyār* *thā*
dhaniyā LOC(beside) answer.M.SG.NOM ready to be.3SG.PST.M
'Dhaniya had a ready answer'.

(47) *unke pās* *lagan* *thī* *aur*
3PL.LOC(beside) passion.F.SG.NOM to be.3SG.PST.F and
samay *thā*
time.M.SG.NOM to be.3SG.PST.M
'He had passion and time'.

The genitive construction is also systematically used to encode non-core possessive notions. In particular, it is frequently used to encode relationships in which one of the two relata is an abstract entity, as in the example (48):

(48) *gharvāloṁ* *ke sāth* *uskā* *bhī* *kuch*
family.M.PL.OBL PSP(with) 3SG.GEN.M also some
kartavya *hai*
responsibility.M.SG.NOM to be.3SG.PRS
'He also has some responsibilities towards his family'.

Moreover, genitive constructions are also frequently used to encode events belonging to the domains of cognition and volition, as exemplified by sentences (49) and (50). Just as in the case of the extension of the semantics of locative constructions, a metaphorical explanation can also apply in the case of genitive constructions. The experiential cognitive domain is, cross-linguistically, intimately connected with the domain of possession; in a large number of world languages, an Experiencer can be encoded as a Possessor in a possessive construction. When this happens, the following metaphor is set in motion: EXPERIENCERS ARE POSSESSORS OF EXERIENCES AND EXERIENCES ARE THINGS POSSESSED (Luraghi 2014). This metaphor occurs, for example, in English, in Greek (Benvenuto 2014; Benvenuto, Pompeo 2017; Luraghi 2020), in Italian and in Latin (Fedriani 2014) among many other languages. Note that the extension of the functionality of the genitive construction to the expression of experience is far more systematic than the extension of locative constructions for metaphorical uses.

(49) *merā* *is* *vyavasthā* *par* *viśvās* *nahīṁ*
1SG.GEN.M this.OBL system.OBL on faith.M.SG.NOM not
hai
to be.3SG.PRS
'I have no faith in this system'.

(50) *unkī* *yah* *icchā* *hai* *ki*
3PL.GEN.F this.DIR desire.F.SG.NOM to be.3SG.PRS that
'They want that… (Lit. Theirs is the desire that… / They have the desire that…)'.

7 Other Constructions and the Notion of Abstract Possession

In the above exposition of Hindi possessive constructions, an analysis of dative construction and inessive-locative construction has been put aside. As mentioned in § 6, these two constructions are used to encode only the notion of abstract possession, and they disallow the encoding of prototypical possession. In some Hindi grammars (Hook 1979; Kachru 2006), these two sentence-types are classified as possessive constructions; however, it may be argued that they should not be considered as truly possessive, since they are prototypically used to express non-possessive situations. A brief discussion of these two patterns follows.

7.1 The Inessive Construction

In inessive constructions, the argument in the initial position is marked in the oblique case followed by the postposition *meṁ* 'in', while the second argument appears in the nominative case. The predicate is once again the verb *honā* 'to be' with an existential function. This construction is sometimes considered to be specialised for the encoding of 'possession of qualities'. See the example in sentence (51).

(51) *unkī* *patnī* *meṁ* *kyoṁ* *vahī* *ātmābhimān*
3PL.GEN.F wife.F LOC(in) why that same self-confidence.SG.NOM
nahīṁ *hai*
not to be.3SG.PRS
'Why doesn't his wife have that same self-confidence?' [Lit. 'Why in his wife there is not that same self-confidence?'].

The inessive construction can be also used to encode what Heine (1997, 35) defines "inanimate inalienable possession"; see the example in (52):

(52) *us* *ghar* *meṁ* *cār* *kamre* *haiṁ*
that.OBL house.OBL LOC(in) four room.M.PL.NOM to be.3PL.PRS
'That house has four rooms' [Lit. 'In that house, there are four rooms'].

An argument for the inessive construction not being included in the classification of Hindi possessive constructions can be made: this pattern, indeed, is not used to express prototypical possessive notions as defined in this paper (§ 4). The inessive construction in (52), for example, expresses a part-whole relationship rather than a possessive one. In many languages, like English, part-whole (or meronymic) relationships can be encoded through the constructions conventionalised for the expression of possession (as in the sentence 'That house has four rooms'); in these languages, these sentences express 'inanimate inalienable possession'. However, in such cases, we are not dealing with possessive relationships but rather with relational situations of a different type that are being conceptualised through possession. As Stassen points out, inanimate possession is to be "consider[ed] a metaphorical extension of possession, in the same way that the notion of possession can be extended into the domain of aspect or modality" (2009, 17).

Remarkably, the metaphorical extension of possessive constructions to part-whole relationships also takes places in Hindi: the sentence in (52) can also be encoded through a genitive construction, as exemplified in (53).

(53) *us* *ghar* *ke* *cār* *kamre* *haiṁ*
that.OBL house.OBL GEN.M.PL four room.M.PL.NOM to be.3PL.PRS
'That house has four rooms' [Lit. 'Of that house, there are four rooms'].

The genitive construction in (53) can be interpreted as expression of "inanimate inalienable possession": it expresses a meronymic relationship metaphorically conceptualised as possession. Since the semantics of the genitive construction is more relational than possessive (§ 6.4), it is not surprising that this pattern allows the encoding of both possessive relationships and meronymic relationships. In contrast, the same extension is disallowed by the locative-adessive construction with the postposition *-ke pās*, which is possessive (and not relational) in its prototypical use.

7.2 The Dative Construction

In Hindi grammars, dative constructions are sometimes numbered among possessive notions and are said to be specialised for the encoding of 'abstract possession'. In this type of sentence, the most salient argument is marked with the dative, a case that in Hindi is prototypically associated with the Experiencer/Beneficiary, not with the Possessor. The second argument is in the nominative and agrees with the existential verb *honā*. Some examples are given below:

(54) *unheṁ* *kuch* *bolne-kā* *adhikār* *hai*
3PL.DAT something say.INF.OBL-GEN.M right.M.SG.NOM to be.3SG.PRS
'They have the right to say something'.

(55) *mujhe* *sir-dard* *hai*
1SG.DAT headache.SG.NOM to be.3SG.PRS
'I have a headache'.

Once again, it is apparent that these constructions cannot really be considered possessive: what emerges from examples (54)-(55) is not an expression of possessive events but rather of other types of situations. Example (54) illustrates a beneficiary event: the first argument should not be seen as a Possessor of an abstract entity, but instead as the Beneficiary of a situation; while in (55), the dative construction expresses a body sensation: the argument in the dative is an Experiencer, not a Possessor.

From a typological perspective, the notion of 'abstract possession' is quite problematic in itself: it is very far from the possessive prototype, as it lacks both control and spatial proximity, and whether it can

even be considered as a possessive notion is debatable (Barðdal, Danesi 2018). In SAE languages, like Italian and English, it makes sense to assume the existence of such a notion, since the encoding strategy for the core possessive meaning is also used to express relation with abstract relata (as in example (6) 'That woman has courage'). In languages like English, then, the relation with an abstract entity, like the body sensation in the sentence 'I have a splitting headache', is metaphorically interpreted as possession. Note that the notion of abstract possession is not completely irrelevant in Hindi: as noted in § 6.5, while dealing with other uses of locative and genitive constructions, these patterns can encode relationships in which the first participant is [+HUMAN] and the second participant is [-CONCRETE]. In particular, the genitive construction is systematically used for the expression of the experiential domains of cognition and volition which in Hindi can be metaphorically conceptualised through possessive relationships. In these cases, it makes sense to talk about *abstract possession*, since the constructions under consideration are prototypically associated with the encoding of core possessive notions and are metaphorically extended to the expression of other situations. This explanation however does not work for a possessive interpretation of the Hindi dative construction or the Hindi inessive construction: as noted, the dative postposition *-ko* is prototypically associated with the Beneficiary or the Experiencer of an event, while the locative postposition *meṁ* is prototypically associated with inessive-locative meaning.

8 Conclusion

This paper has analysed the expression of core possessive notions in Hindi, demonstrating that two syntactic patterns can encode the notion of ownership, namely, the locative construction with the postposition *-ke pās*, and the genitive construction. The locative construction is clearly the more conventional. Locative marking on the PR is used to express the whole domain of alienable possession: it can encode the notion of ownership, and it is also the only type of sentence that allows the expression of temporary or physical possession. Moreover, it is highly specific in its semantics: except for the expression of professional relationships, it is rarely used for the encoding of other situations.

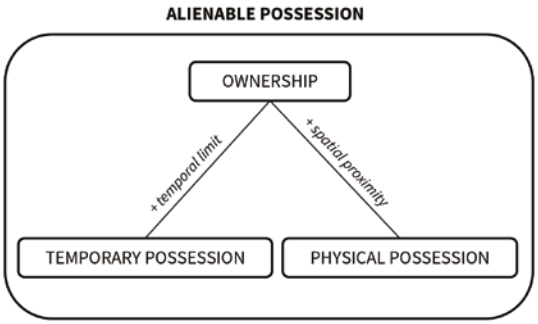


Figure 1 Semantic reconstruction of Hindi locative (possessive) construction

The genitive construction, instead, has vaguer semantics: its basic meaning is relational, not possessive. This construction is the canonical vehicle for expressing inalienable relationships, but it can also be used to encode ownership when the semantics of the event shows some specific relational properties, i.e. when there is a strong connection between the PE and the personal sphere of the PR. This explains why only the prototypical notion of ownership allows the use of genitive marking on the PR: temporary and physical possession are normally not characterised by an intimate relationship between the two entities. Additionally, given its semantic vagueness, the genitive construction allows more functional extension. It can also be used metaphorically to encode possession of abstract entities and possession of psychological states.

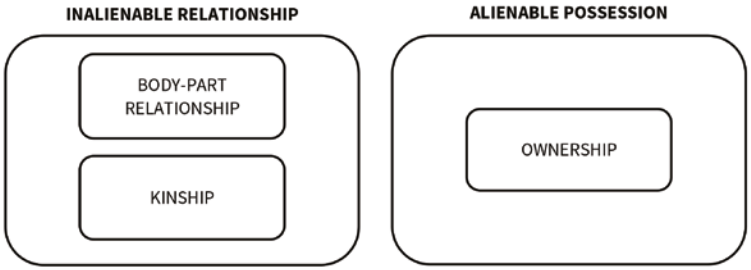


Figure 2 Semantic reconstruction of Hindi genitive possessive construction

The results of this investigation are depicted in the overall semantic map of Hindi possessive constructions shown in **[fig. 3]**.

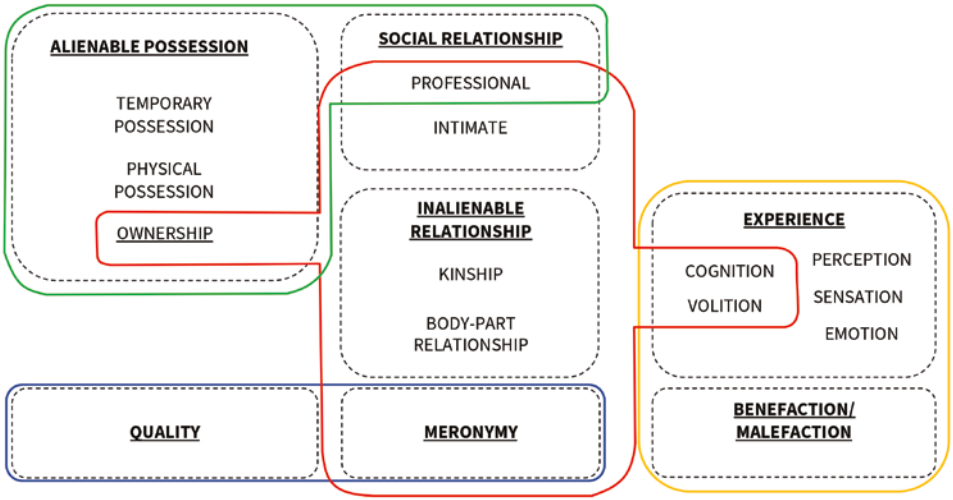


Figure 3 Semantic maps of possessive constructions in Hindi. In red, the genitive construction; in green, the locative construction; in blue, the inessive construction, and in yellow, the dative construction

List of abbreviations

1	First person
2	Second person
3	Third person
ACC	Accusative
ADJ	Adjective
AOR	Aorist
COM	Comitative
CRR	Correlative
CVB	Converb
DAT	Dative
DIR	Direct
EMPH	Emphatic
ERG	Ergative
F	Feminine
FUT	Future
GEN	Genitive
INDF	Indefinite
INF	Infinitive
INS	Instrumental
LOC	Locative

M	Masculine
NOM	Nominative
OBL	Oblique
PE	Possessee
PL	Plural
PR	Possessor
PRF	Perfect
PRN	Pronoun
PRS	Present
PSP	Postposition
PST	Past
REFL	Reflexive
REL	Relative
SG	Singular
SBJV	Subjunctive
V	Verb

Bibliography

Amberber, M.; de Hoop, H. (eds) (2005). *Competition and Variation in Natural Languages. The Case of Case*. Amsterdam: Elsevier.

Baldi, P.; Cuzzolin, P. (2005). "Considerazioni etimologiche, areali e tipologiche dei verbi di 'avere' nelle lingue indoeuropee". Kiss, S. et al. (eds), *Latin et langues romanes. Études de linguistique offertes à József Herman à l'occasion de son 80ème anniversaire*. Tübingen: Niemeyer, 27-35.

Barðdal, J.; Danesi, S. (2018). "Case Marking of Predicative Possession in Vedic. The Genitive, the Dative, the Locative". Barðdal J.; Pat-El, N.; Carey, S. (eds), *Non-Canonically Case-Marked Subjects. The Reykjavík-Eyjafjallajökull Papers*. Amsterdam; Philadelphia: John Benjamins, 181-212.

Benvenuto, M.C. (2014). "Possession". Giannakis, G.K. (ed.), *Encyclopedia of Ancient Greek Language and Linguistics* (EAGLL), vol. 3. Leiden; Boston: Brill, 115-22. http://dx.doi.org/10.1163/2214-448X_eagll_COM_00000285.

Benvenuto, M.C.; Pompeo, F. (2017). "Abstract Possession and Experiential Expression. Some Preliminary Remarks". Logozzo, F.; Poccetti, P. (eds), *Ancient Greek Linguistics. New Approaches, Insights, Perspectives*. Berlin; Boston: Walter de Gruyter, 507-22. https://doi.org/10.1515/9783110551754-519.

Caracchi, P. (2002). *Grammatica hindi*. Torino: Magnanelli.

Comrie, B. (1989). *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press.

Dhongde, R.V.; Wali, K. (2009). *Marathi*. Amsterdam; Philadelphia: John Benjamins.

Drocco, A. (2008). *L'ergatività in hindi. Studio diacronico del processo di diffusione della posposizione 'ne'*. Alessandria: Edizioni dell'Orso.

Drocco, A. (2018). "An Attempt to Understand the Encoding of Reduced Transitivity in Hindi. The Case of Compound Verbs with Jānā". Sharma, G.; Bhatt, R. (eds), *Trends in Hindi Linguistics*. Berlin: Walter de Gruyter, 265-86. https://doi.org/10.1515/9783110610796-011.

Drocco, A.; Tiwari, N. (2020). "Atypical Compound Verb Constructions in Hindi/Urdu. The Case of Transitivity Polar Verbs with Intransitive Vector Verbs". *Annali di Ca' Foscari. Serie Orientale*, 56, 307-34. http://doi.org/10.30687/AnnOr/2385-3042/2020/56/012.

Fedriani, C. (2014). *Experiential Constructions in Latin*. Leiden; Boston: Brill.

Givón, T. (2001). *Syntax. An Introduction*, vol. 1. Amsterdam; Philadelphia: John Benjamins.

Heine, B. (1997). *Possession. Cognitive Sources, Forces, and Grammaticalization*. Cambridge: Cambridge University Press.

Hook, P. (1979). *Hindi Structures. Intermediate Level*. Ann Arbor (MI): Michigan Papers on South and South East Asia.

de Hoop, H.; Narasimhan, B. (2005). "Differential Case Marking in Hindi". Amberber, de Hoop 2005, 321-45. https://doi.org/10.1016/b978-008044651-6/50015-x.

Hopper, P.J.; Traugott, E.C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Isačenko, A.V. (1974). "On Have and Be Languages. A Typological Sketch". Flier, M.S. (ed.), *Slavic Forum. Essays in Linguistics and Literature*. The Hague: Mouton, 43-77. https://doi.org/10.1515/9783110888386-006.

Kachru, Y. (1970). "A Note on Possessive Constructions in Hindi-Urdu". *Journal of Linguistic*, 6(1), 37-45. https://doi.org/10.1017/s0022226700002346.

Kachru, Y. (2006). *Hindi*. Amsterdam; Philadelphia: John Benjamins.

Keidan, A. (2008). "Predicative Possessive Constructions in Japanese and Korean". *Rivista degli Studi Orientali*, 81, 339-67. https://doi.org/10.1400/143142.

Kibrik, A.E. (1997). "Beyond Subject and Object. Towards a Comprehensive Relational Typology". *Linguistic Typology*, 1(3), 279-346. https://doi.org/10.1515/lity.1997.1.3.279.

Kittilä, S. (2002). "Remarks on the Basic Transitive Sentence". *Language Science*, 24(2), 107-30. https://doi.org/10.1016/s0388-0001(00)00043-7.

Lakoff, G.; Johnson, M. (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Langacker, R.W. (1995). "Possession and Possessive Constructions". Taylor, J.R.; MacLaury, R.E. (eds), *Language and the Cognitive Construal of the World*. Berlin; New York: Mouton de Gruyter, 51-79.

Langacker, R. (1999). *Grammar and Conceptualization*. Berlin; New-York: Mouton de Gruyter.

Langacker, R.W. (2001). "Topic, Subject, and Possessor". Simonsen, H.G.; Endresen, R.T. (eds), *A Cognitive Approach to the Verb. Morphological and Constructional Perspectives*. Berlin; New York: Mouton de Gruyter, 11-48.

Langacker, R.W. (2009). *Investigations in Cognitive Grammar*. Berlin; New York: Mouton de Gruyter.

Lehmann, C. (2002). *Possession in Yucatec Maya*. 2nd revised ed. ASSidUE, Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt 10. Erfurt: Seminar für Sprachwissenschaft der Universität.

Luraghi, S. (2014). "Plotting Diachronic Semantic Maps. The Role of Metaphor". Luraghi, S.; Narrog, H. (eds), *Perspectives on Semantic Roles*. Amsterdam: John Benjamins, 99-150. https://doi.org/10.1075/tsl.106.04lur.

Luraghi, S. (2020). *Experiential Verbs in Homeric Greek. A Constructional Approach*. Leiden; Boston: Brill.

Malchukov, A.L. (2005). "Case Pattern Splits, Verb Types and Construction Competition". Amberber, de Hoop 2005, 73-117. https://doi.org/10.1016/b978-008044651-6/50006-9.

Malchukov, A.L. (2015). "Valency Classes and Alternations. Parameters of Variation". Malchukov, A.; Comrie, B. (eds), *Valency Classes in World's Languages*, vol. 1. Berlin: De Gruyter Mouton, 74-130. https://doi.org/10.1515/9783110338812-007.

Mallinson, G.; Blake, B. (1981). *Language Typology. Cross-Cultural Studies in Syntax*. Amsterdam: North Holland Publishing.

Mazzitelli, L.F. (2015). *The Expression of Predicative Possession. A Comparative Study of Belarusian and Lithuanian*. Berlin; New York: Mouton de Gruyter.

McGregor, R.S. (1972). *Outline of Hindi Grammar. With Exercises*. Oxford: Clarendon Press.

Milanetti, G.; Gupta, S.T. (2008). *Corso di hindi*. Milano: Hoepli.

Mohanan, T. (1994). *Argument Structure in Hindi*. Stanford (CA): The Center for the Study of Language and Information Publications.

Montaut, A. (1997). "Constructions expérientielles et possessives en hindi", in "Les langages d'Asie du Sud", monogr. no., *Faits de langues*, 10, 229-36.

Montaut, A. (2004a). *A Grammar of Hindi*. Munich: Lincom.

Montaut, A. (2004b). "Oblique Main Arguments as Localizing Predications in Hindi/Urdu". Bhaskararao, P.; Subbarao, K.V. (eds), *Non Nominative Subjects*, vol. 2. Amsterdam; Philadelphia: John Benjamins, 178-210.

Montaut, A. (2013). "The Rise of Non-Canonical Subjects and Semantic Alignments in Hindi". Seržant, I.A.; Kulikov, L. (eds), *The Diachronic Typology of Non-Canonical Subjects*. Amsterdam; Philadelphia: John Benjamins, 91-118. https://doi.org/10.1075/slcs.140.05mon.

Onishi, M. (2001). "Non-Canonically Marked Subjects and Objects. Parameters and Properties". Aikhenvald, A.Y.; Dixon, R.M.W.; Onishi, M. (eds), *Non-Canonical Marking of Subjects and Objects*. Amsterdam; Philadelphia: John Benjamins, 1-51. https://doi.org/10.1075/tsl.46.03oni.

Pandharipande, R. (1981). "Interface of Lexicon and Grammar. Some Problems in Hindi Grammar". *Studies in the Linguistic Sciences*, 11(2), 77-100.

Pompeo, F. (2010). *Aspetti cognitivi e semantici dell'espressione del possesso predicativo in greco antico*. Cosenza: Pellegrini Editore.

Premcand, M. (2019). *Godān*. New Delhi: Finger Print Publishing.

Seiler, H. (1983). *Possession as an Operational Dimension of Language*. Tübingen: Narr.

Shackle, C. (1972). *Punjabi. Teach Yourself Books*. Hertford: English University Press.

Stassen, L. (2009). *Predicative Possession*. Oxford: Oxford University Press.

Taylor, J.R. (1989). *Linguistic Categorization. Prototypes in Linguistic Theory*. Oxford: Clarendon Press.

Taylor, J.R. (1996). *Possessives in English. An Exploration in Cognitive Grammar*. Oxford: Clarendon Press.

Thompson, H.R. (2010). *Bengali. A Comprehensive Grammar*. Abingdon: Routledge.

Tsunoda, T. (1985). "Remarks on Transitivity". *Journal of Linguistics*, 21(2), 385-96. https://doi.org/10.1017/s0022226700010318.

Tsunoda, T. (2015). "The Hierarchy of Two-Place Predicates. Its Limitations and Uses". Malchukov, A.; Comrie, B. (eds), *Valency Classes in World's Languages*, vol. 2. Berlin: De Gruyter Mouton, 1597-625. https://doi.org/10.1515/9783110429343-018.

de Vaan, M. (2008). *Etymological Dictionary of Latin and the Other Italic Languages*. Leiden: Brill.

Viti, C. (2004). "Funzioni semantiche e pragmatiche nelle strategie di possesso dell'antico indiano". *Archivio glottologico italiano*, 89(1), 41-83.

Wichmann, S. (2008). "The Study of Semantic Alignment. Retrospect and State of the Art". Donohue, M.; Wichmann, S. (eds), *The Typology of Semantic Alignment*. Oxford: Oxford University Press, 3-23. https://doi.org/10.1093/acprof:oso/9780199238385.003.0001.

Semestral journal

Department of Asian
and North African Studies

Università
Ca'Foscari
Venezia