Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Copula and Agent Marking in Nepali: Testing Areal Convergence and Substrate Explanations.

Nawa Raj Subba

dr.subba2015@gmail.com

Independent https://orcid.org/0000-0001-5700-2179

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11820404/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Copula and Agent Marking in Nepali:

Testing Areal Convergence and Substrate Explanations.

Nawa Raj Subba

Independent Researcher

ORCID: https://orcid.org/0000-0001-5700-2179

Corresponding author: dr.subba2015@gmail.com

Article type: Research article **Subject:** Linguistics; Indo-European; South Asian studies

Word count: [about 9,200 words, including tables and references]

Preprint notice. This is a preprint and has not been peer reviewed. An earlier version of this article is archived on Zenodo: https://doi.org/10.5281/zenodo.24415241. The present version supersedes it.

Copula and Agent Marking in Nepali:

Testing Areal Convergence and Substrate Explanations.

Abstract

The lack of early records makes it impossible to apply standard criteria for contact-induced change directly, since the recipient language's pre-contact state is unknown. Rather than relying on unavailable historical evidence, this study uses a series of predetermined comparative proxies applied in a fixed order. The method is tested on Nepali, an Indo-Aryan language of the Himalayas, by comparing it with a range of languages, including Indo-European languages from outside South Asia, Sanskrit, Plains and Pahari Indo-Aryan languages, Assamese, Shina, and four Tibeto-Burman languages spoken in Nepal. Reduplication (as a calibration case) is part of the older South Asian stratum. The opposition between the copulas 'ho' and 'cha' is shared with Tibeto-Burman but remains unresolved, because the same opposition also arises independently within Indo-European. Agent marking with -le beyond the perfective (used by every speaker in Lindemann's 2019 corpus sample) meets all the criteria: within Indo-Aryan, it occurs only in languages in contact with Tibeto-Burman; it is not attested in Nepali's closest Pahari relatives; and it is the norm in the Tibeto-Burman languages examined. The comparative evidence therefore indicates that the Tibeto-Burman languages of the Himalaya have shaped Nepali grammar in this domain: an inherited Indo-Aryan marker has been reorganised on a Tibeto-Burman pattern. The evidence cannot yet establish the exact mechanism: whether this influence entered Nepali via substrate transfer from shifting speakers or through borrowing during long-term bilingualism.

Keywords: language contact; pattern replication; substrate; agent marking; Nepali; Tibeto-Burman

1 Introduction

Nepali is an Indo-Aryan language with a large proportion of its formal vocabulary drawn from Sanskrit. It developed in the Himalayas and, for a long time, served as the language of Tibeto-Burman-speaking peoples (Hutt 1988; Noonan 2003). That in itself raises a question that has not been properly addressed: are any of the basic grammatical categories of Nepali, in fact, more like those of the Tibeto-Burman languages than those of Sanskrit and the other Indo-Aryan languages of the plains?

The question is easy to state but difficult to answer with any degree of certainty. Any features that Nepali shares with Tibeto-Burman languages may have been inherited from Indo-European, may be part of the older South Asian linguistic area (which also influenced Sanskrit) (Emeneau 1956; Masica 1976), may result from independent evolution, or may be due to contact. A single point of similarity between Nepali and a neighbouring language is not enough to establish the origin. A comparative method is therefore necessary, one that examines all possible sources in turn and pre-emptively rules out any outcome inconsistent with a contact hypothesis.

This paper describes such a procedure and clearly distinguishes between two contact situations that the literature on Nepali frequently merges. The substrate situation is one in which speakers of a Tibeto-Burman language take up Khas-Nepali and transfer features from their first language. The convergence situation is one in which features are acquired through long-term bilingualism, with no assumption about which language the speakers of the other language have shifted to (Thomason, 2001). Either situation could lead to Nepali sharing features with Tibeto-Burman languages. The presence of a shared feature therefore does not, in itself, tell us whether we are dealing with substrate, convergence, or inheritance. The question is therefore answered in the literal sense, and the conclusions are restricted to what the comparison can show.

The paper's main contribution is methodological. It sets out a procedure for identifying, in contact zones with no documentation of the language's pre-contact state, which shared grammatical features can be attributed to contact. The procedure replaces the historical evidence required by the standard criteria (Thomason 2001; Poplack & Levey 2010) with a fixed set of comparative proxies applied in a fixed order and specifies which results would rule out a contact explanation. Nepali is an ideal test case, with well-documented Indo-European/Indo-Aryan ancestry and very little documentation of contact with Tibeto-Burman languages. The three features under consideration are reduplication (known to be an areal feature), the opposition between the copulas 'ho' and 'cha', and agent marking with '-le' beyond the perfective.

The results show that one feature is attributable to the old South Asian stratum, one remains unresolved, and one passes the contact test, pointing to Tibeto-Burman influence on Nepali grammar. The procedure is not specific to Nepali and can be used in any situation where you have a language with a well-documented ancestor that developed in contact with a language from a different family, and where the history of that contact is not recorded. The scope of the claim is deliberately restricted. The tense, case, and bulk of Nepali verb morphology are Indo-European, and none of the discussion that follows suggests that Nepali grammar as a whole is Tibeto-Burman. The issue is whether some distinctions may, in fact, be of Tibeto-Burman origin within an Indo-Aryan framework.

Section 2 is background. Section 3 is the conceptual framework (the processes of contact-induced change, the diagnostics proposed, the comparative proxies that we use in place of historical evidence). Section 4 is the research questions and hypotheses. Section 5 is the data and methods. Section 6 is the results. Section 7 covers alternative explanations, one feature's failure to meet the contact criteria, the substrate question, and the study's limitations. Section 8 is the conclusion.

2 Background

2.1 An older South Asian layer in Sanskrit

Sanskrit's core grammar is, of course, Indo-European, and its genealogy is not in dispute here. The question is how much Old Indo-Aryan had already picked up features from the languages it came into contact with in South Asia. Emeneau (1956) showed that the various languages of the subcontinent share several structural features that their relatives

outside the subcontinent do not. Masica (1976) has produced a very detailed map of this linguistic area. Kuiper (1991) and Witzel (1999) have identified loanwords and structural features of pre-Indo-Aryan origin in the Rigveda itself, and Southworth (2005) goes on to show the extent of the Dravidian and other substrata. Retroflexes and productive reduplication are among the features generally ascribed to this layer (Abbi 1992).

This has a methodological implication. A feature that Nepali shares with the Tibeto-Burman languages and that is also present in Sanskrit cannot be the result of Himalayan contact simply because it is absent from Indo-European languages outside South Asia. Any test for a Himalayan stratum therefore has to look at Sanskrit as well as what came before and around it.

2.2 Khas-Nepali in the Himalaya

Khas-Nepali developed in the western Himalaya in the Karnali basin and spread eastwards with the Khas polities, the Gorkha state, and later with the nation's expansion (Hutt 1988). Throughout its history, it has been the language of speakers of Tibeto-Burman languages in the west and west centre (where Kham and Magar are spoken) and in the centre and east (where it has come into contact with Gurung, Tamang, Newar, Rai and Limbu). It has acted as a lingua franca for many of these groups, and Noonan (2003) has documented recent cases of grammatical borrowing between Nepali and Tibeto-Burman languages in both directions.

The study design resulted in part from two considerations. One concerned the age of Tibeto-Burman influence on Nepali grammar. If that influence is old, the most likely source would be among the western languages in which Khas-Nepali initially developed, rather than the eastern languages that came in later. The other consideration concerned Austro-Asiatic languages. These are presently restricted to the eastern Tarai, but indirect evidence suggests an earlier presence in the hills (van Driem 2013:383-385). For this reason, the comparative work focuses on Tibeto-Burman.

2.3 Previous work

The descriptive grammars of Nepali (e.g. Acharya 1991) and the comparative work on Indo-Aryan (Masica 1991; Verbeke 2013a) all describe the features we're looking at here, though very rarely do they ask which of these are regional. Verbeke (2013a) compares alignment across New Indo-Aryan and carefully points out that Hindi-Urdu should not be taken as the paradigm for all Indo-Aryan languages. In her chapter on Nepali, she suggests (following Genetti 2007) that extending ergative marking to imperfective clauses may reflect frequent bilingualism in Nepali and Newar (Phillips 2014: 105). The typological background for section 6.3 comes from work on optional and semantically conditioned agent marking in Tibeto-Burman (LaPolla 1995; Chelliah & Hyslop 2011; DeLancey 2011). What has been missing is a procedure which brings these literatures together and tests the features of interest against all the relevant comparison groups at the same time.

3 Contact processes and how to diagnose them

3.1 Two routes to shared structure

Since Weinreich (1953) contact-induced change has been categorised into two broad types. Thomason & Kaufman (1988:37) make the distinction between borrowing (where speakers take features of another language into their own) and interference via shift (where a group shifting to a target language brings features of their first language with them). Van Coetsem (2000) reworks the distinction in terms of agentivity. In borrowing, the agents dominate the recipient language and import material from the source (recipient-language agentivity); in imposition, the agents dominate the source language and impose their patterns on the language they are acquiring (source-language agentivity). Winford (2005) contends that this is a more useful way of categorising the phenomena, as it removes the confusion between processes and outcomes that terms such as 'substratum influence' and 'interference' have generated, and the same speakers can be borrowers in one situation and imposers in another.

The two routes yield different predictions. On the Thomason and Kaufman (1988) account, borrowing normally starts with vocabulary and reaches structure only under conditions of very intense and prolonged contact. The alternative is interference via shift, which can impact phonology and syntax without necessarily leaving much of a trace in the lexicon. For Nepali, the two scenarios outlined in Section 1 map onto these two routes. A substrate account is an account of imposition (Tibeto-Burman speakers acquiring Khas imposed their categories on it). A convergence account fits either route or a combination of both, as long-term bilingualism allows each community to model one language on the other. This has a very real bearing on the evidence, which would in fact determine the issue. The substrate account would predict structural Tibeto-Burman features in Nepali, with only a small amount of Tibeto-Burman vocabulary in the core lexicon. The borrowing account would predict that structural replication occurred both before and during the lexical borrowing period. The lexical aspect of that prediction is outside the remit of this paper, but it does provide an independent test of the two accounts (Section 7.3).

3.2 Pattern replication

A second distinction concerns what is actually being transferred. Matras and Sakel (2007) distinguish between the replication of matter (i.e., phonological forms) and the replication of patterns, in which the speaker reproduces the construction's patterning in the model language using material from the replica language. In their account, bilinguals identify a key feature of the model construction, match it to an existing construction in the replica language, and then apply it to a new function. Heine and Kuteva (2003, 2005) refer to the same process as contact-induced grammaticalisation: the speaker replicates a grammatical category from the model language by grammaticalising native material along paths which are attested in the absence of contact. Ross (2007) takes the opposite approach and shows that you can have very extensive grammatical calquing without any metatypy at all (i.e. without the wholesale restructuring of one language on the model of another).

Together, these constitute what a contact-induced feature of Nepali would be like. It would have an Indo-Aryan base; it would differ from Sanskrit and the plains Indo-Aryan pattern in its organisation rather than its material; and it would leave the rest of Nepali's grammar Indo-Aryan. Both features discussed in section 6 fit this profile, and Nepali as a whole is

precisely what Ross has in mind as grammatical calquing without metatyping: its inflectional morphology, tense system, and case system are all Indo-Aryan. The profile, of course, is not in itself diagnostic of contact. If contact-induced grammaticalisation follows the same routes as internal grammaticalisation (which Heine and Kuteva (2003) argue it does), the outcome is indistinguishable whether or not a model is involved. The matter-pattern distinction therefore tells us what to look for but not whether or not contact has occurred.

3.3 Reading the process from the outcome

Ross (2013) asks whether bilingually induced and shift-induced change produce different results and concludes they do, because they occur at different points in the speaker's life. In his view, grammatical calquing and restructuring are the domain of those who acquire both languages prior to adolescence, whereas adults who shift to a new language as adults are more likely to transfer phonology and particular constructions; generational shift (where the children of one community grow up speaking the language of another) tends to leave very little more than a set of specialist vocabulary (Ross 2013: 23–30, 36–37). Ross did not intend these criteria to function as universal tests. Rather, they were drawn from patterns observed in Melanesian contact settings. For that reason, they cannot be applied to the Himalayan context without careful consideration. Even so, they remain valuable because they allow the substrate hypothesis to be evaluated against linguistic evidence. If substrate influence and long-term bilingual contact have played different roles, they should leave different structural traces in the language.

For Nepali, the implication is that the substrate/convergence decision can be based on other features in addition to those under discussion. Whether Nepali itself has Tibeto-Burman-type phonological features is essential: on Ross's diagnostics, phonological transfer combined with the calquing of specific constructions would indicate adult shift, whereas calquing the grammar without any phonological transfer would indicate long-term bilingualism. The present paper doesn't look at phonology, but it sets out the grammatical results in such a way that such evidence could be brought to bear on them (Section 7.3).

3.4 Areas, proxies and the burden of proof

South Asia is the textbook example of a linguistic area (Emeneau 1956; Masica 1976), and there's a good deal of evidence for convergence between Indo-Aryan and its neighbours at several levels. We have evidence of grammatical features spreading from Munda into Indo-Aryan in particular areas (e.g., Jharkhand) (Peterson 2010) and evidence of a structural split between the eastern and western parts of Indo-Aryan as a whole (Ivani, Paudyal & Peterson 2021). Campbell (2006) argues that linguistic areas are not entities with defining characteristics but simply the result of multiple individual diffusions, each of which would need to be explained on its own. This paper takes that view. For each feature, it asks whether contact can account for it rather than whether Nepali belongs to a Himalayan linguistic area.

The most rigorous standard for such claims comes from the variationist tradition. Poplack and Levey (2010) state that claims of contact-induced change must be evaluated against comparable varieties that have not had the contact, using methods that can rule out internal

explanations, and that the onus of proof is on the claim of contact. Thomason (2001) has similar criteria: you need an identified source, evidence that the feature in question was present in the source and absent from the recipient before contact, and you must rule out internal and universal explanations.

In a contact zone with no early records, neither standard can be met at first, as we have no evidence of the recipient's pre-contact state. The procedure adopted here substitutes the missing historical evidence with a series of comparative proxies (Table 1). The Old Indo-European languages outside of South Asia stand in for the inherited state; Sanskrit stands in for the pre-Himalayan state of Indo-Aryan; the plains Indo-Aryan languages stand in for the non-contact varieties that the variationist work requires; the other Indo-Aryan languages in contact with Tibeto-Burman, as well as the Central Pahari languages closest to Nepali, provide further tests as to whether a feature is contact-induced or is actually a feature of Nepali itself. Later Indo-European languages outside South Asia provide the control for independent development. Each proxy is imperfect, and Table 1 makes clear what each one can and can't tell us.

Table 1. *Diagnostic criteria and the comparative proxies used for them*

Criterion	Proxy in this study	Limitation
Feature not inherited (Thomason 2001)	Latin, Ancient Greek, Avestan	Absence may reflect later loss rather than original absence
Feature absent from the recipient before contact (Thomason 2001; Poplack and Levey 2010)	Vedic and Classical Sanskrit	Sanskrit is a learned standard, not the direct ancestor of Khas
Comparison with non-contact varieties (Poplack and Levey 2010)	Hindi-Urdu, Punjabi, Marathi, Bengali	Plains varieties have contact histories of their own
Feature tied to contact rather than to Nepali (Campbell 2006)	Assamese, Shina; Kumauni, Garhwali	Few languages, coded for agent marking only
Independent development excluded (Thomason 2001)	Spanish, Irish	Shows that a development is possible, not how likely it is
Identifiable source (Thomason 2001)	Kham, Magar, Newar, Limbu	Small sample; contact with any one source cannot be dated

The comparative proxies help identify contact-related patterns, but they have clear limits. They can show that a feature is shared by Nepali and neighbouring Tibeto-Burman languages while being absent from Sanskrit and other relevant Indo-Aryan comparison points. What they cannot show is how the feature spread or when the transfer took place. For this reason, the present analysis focuses on convergence rather than on historical reconstruction. Features that pass all comparative controls may reasonably be treated as candidates for contact influence, yet the choice between substrate transfer and borrowing must ultimately rest on independent historical evidence.

4 Research questions and hypotheses

The study addresses three questions:

1. RQ1. The first question concerns inheritance. Are any of the selected grammatical features found in Nepali missing from pre-South Asian Indo-European languages? If so, their origin may require an explanation other than straightforward inheritance.
2. RQ2. Which of these were already present in Old Indo-Aryan (the old regional layer), and which are absent from Sanskrit and the Indo-Aryan languages but present in the Tibeto-Burman languages (the contact-layer candidates)?
3. RQ3. For each contact-layer candidate, does it pass the controls for independent development, and does the evidence support substrate transfer, borrowing, or simply convergence?

Each hypothesis states in advance the result which would count against it (Table 2).

Table 2. *Hypotheses and disconfirming results*

Hypothesis	Prediction	Counts against it
H1	The feature, as defined in Table 2, is absent from the old Indo-European languages outside South Asia (Latin, Ancient Greek, Avestan)	The feature, as defined, occurs in these languages
H2	Some features are already present in Sanskrit	None of the features occurs in Sanskrit
H3	Some features are absent from Sanskrit and from plains Indo-Aryan but present in Tibeto-Burman languages of Nepal	A plains Indo-Aryan language shows the same feature
H4	Where a contact-layer candidate occurs in Indo-Aryan outside Nepal, it occurs in languages in contact with Tibeto-Burman	The feature occurs in Indo-Aryan languages without Tibeto-Burman contact

Later Indo-European languages outside South Asia (Spanish, Irish) serve a different function: a feature that arose in one of those branches is not inherited, but its presence in that language shows the feature can arise independently. That control is the basis of step 4 of the procedure (Section 5.4).

Where a hypothesis fails for a particular feature, the feature is reassigned, and the failure is noted. Two further distinctions are made throughout. The first is between matter and pattern (Matras & Sakel, 2007): a feature may be composed of inherited Indo-Aryan forms arranged according to a pattern found in the surrounding languages, without any actual forms having been borrowed. The second is between convergence and borrowing: convergence occurs when a pattern is shared, whereas borrowing requires evidence of direction and timing (Section 5.5).

5. Data and method

5.1 Features

The three features were defined prior to the comparison (Table 3). Reduplication is the calibration case because its areal status in South Asia is well established (Abbi 1992), so

the procedure should classify it as an old regional feature. If it didn't, then the procedure is suspect.

Table 3. *Features and definitions used for comparison*

Feature	Definition used for comparison
F1 Reduplication	Productive full repetition of a word to express distribution, intensity, manner or continuation (*bistāri bistāri* 'gradually')
F2 Copula opposition	An overt identity copula (*ho*) lexically distinct from an existential-locative copula (*cha*)
F3 Agent marking beyond the perfective	An agent marker that is not restricted to perfective clauses, and that in non-past clauses may occur depending on semantic or discourse factors

The definition of F2 is more restricted than "two copulas". The feature in question is the specific opposition found in Nepali, where one overt copula expresses identity and class membership and a second one expresses existence and location. Languages which have no overt copula of identity in the present tense (e.g. Bengali) lack

the feature. The negative existential form is noted but not used as a criterion, as in Nepali (*chaina*) and in most of the comparison languages it's simply the negated existential.

5.2 Languages compared

Table 4. *Comparison groups*

Group	Languages	Role in the comparison
Old Indo-European outside South Asia	Latin, Ancient Greek, Avestan	Shows what is inherited from Indo-European
Later Indo-European outside South Asia	Spanish, Irish	Shows what can arise independently within Indo-European
Old Indo-Aryan	Vedic and Classical Sanskrit	Shows what was present before Nepali
Plains Indo-Aryan	Hindi-Urdu, Punjabi, Marathi, Bengali	Shows what is general to South Asian Indo-Aryan
Central Pahari	Kumauni, Garhwali	The Indo-Aryan languages closest to Nepali outside Nepal (F3 only, from published descriptions)
Indo-Aryan in other Tibeto-Burman contact zones	Assamese, Shina	Tests whether a feature is specific to Nepali or associated with Tibeto-Burman contact more generally (F3)
Tibeto-Burman, west and west-centre	Kham, Magar	Candidate sources in the area where Khas-Nepali developed
Tibeto-Burman, centre and east	Newar, Limbu	Candidate sources in the area of later contact

5.3 Sources

Forms are drawn from published reference grammars and descriptions: Nepali from Acharya (1991); Hindi-Urdu from Kachru (2006); Bengali from Thompson (2006, 2012); Marathi from Dhongde and Wali (2009); general Indo-Aryan agent marking from Masica

(1991), Verbeke (2013a) and Khokhlova (2013); frequencies of agent marking in Nepali in spontaneous speech from Lindemann (2019); Kumauni and Garhwali from Stroński (2010); Kham from Watters (2002, 2004, 2008); Magar from Grunow-Hårsta (2007, 2011) (her examples are from her grammar of the Syangja (western) and Tanahu (eastern) dialects, Grunow-Hårsta 2008); Newar from Hale and Shrestha (2006) and Genetti (2007); Limbu from van Driem (1987). The Nepali examples without sources are standard forms. The native speaker provided the Limbu examples: the copula forms match those described in van Driem (1987: 55-68). Examples follow the Leipzig Glossing Rules, and examples from other sources retain the original author's morpheme segmentation. Nepali is transliterated according to ISO 15919 (so 'cha' for छ and 'ca' for च).

5.4 Classifying a feature

Each feature is assigned to a layer by five tests, applied in a fixed order:
Step 1. If the feature as defined occurs in the old Indo-European languages outside of South Asia, then it is inherited (it gives us no evidence of contact with South Asia). That it occurs only in the later branches of the language family does not mean it is inherited; that would be evidence that it arose independently.

Step 2. If it doesn't, but it does occur in Sanskrit, it's part of the old regional stratum.

Step 3: If it occurs in neither Sanskrit nor the plains Indo-Aryan languages but does occur in the Tibeto-Burman languages of Nepal, then it is a contact-layer candidate, and Step 4 applies.

Step 4. A feature passes the contact test only if (a) it is present in at least one Tibeto-Burman language of the area in which Khas-Nepali originally developed, and (b) where it is found in Indo-Aryan languages outside Nepal, it is only found in those languages which are in contact with Tibeto-Burman (H4), and it is not attested as an independent development in the later Indo-European control languages. A feature that fails the test is classified as unresolved: on the available evidence, we cannot tell whether it is contact or independent development.

Step 5. If the feature is not present in any comparison group, treat it as an internal development of Nepali.

As the order is fixed, the same comparison will always yield the same classification, and any reader can re-run it with the additional languages. Step 4 makes explicit a criterion applied informally throughout: similarity to a neighbouring language is evidence of contact only if the inheritance and independent development controls have been satisfied.

Section 3 details the criteria for assigning a feature to contact and the proxies used here. If a candidate passes step 4, the result is stated as convergence unless historical evidence of the direction and timing of the transfer in question is available, which would be necessary to establish borrowing or substrate transfer (Thomason, 2001). A candidate that fails step 4 is not thereby shown to be an internal development; it is simply classified as unresolved, i.e., the comparison does not allow us to distinguish between contact and independent development.

6 Results

6.1 Reduplication: the old regional layer

Nepali uses full reduplication productively for distribution, intensity and manner:

(1)	bistāri	bistāri	hiṭ-nu-hos
	slowly	slowly	walk-INF-IMP.HON

'Walk slowly.'

(2)	ṭhūlā	ṭhūlā	ghar
	big	big	house

'big houses (each of them big)'
Hindi has the same pattern (*dhīre dhīre* 'slowly') and so does Sanskrit. Pāṇini's rule *nityavīpsayoḥ* (Aṣṭādhyāyī 8.1.4) covers reduplication for constant/repeated action and distribution, with *pacati pacati* 'he cooks again and again' and *grāme grāme* 'in every village' as examples. Manner reduplication of the Nepali type is also attested (śanaiḥ śanaiḥ 'gradually' in Bhagavadgītā 6.25).

Old IE languages outside of South Asia have lexicalised reduplication (e.g. Latin *quisquis* 'whoever'), but productive full reduplication for distribution and manner is a feature of South Asia rather than of IE as a whole (Abbi 1992). Under the procedure of Section 5.4, reduplication is assigned to the old regional stratum (H1 and H2 are supported). The calibration case therefore behaves as expected, and Nepali reduplication provides no evidence of contact in the Himalaya.

6.2 The copula opposition: Indo-Aryan matter, Tibeto-Burman pattern?

Nepali makes a lexical distinction between identity and existential-location predicates through two separate copulas. The negative counterpart of the existential copula is *chaina*.

(3)	yo	ḍhuṅgā	ho
	this	stone	COP.ID.3SG

'This is a stone.'

(4)	ḍhuṅgā	tyahā˜	cha
	stone	there	COP.EXIST.3SG

'The stone is there.'

(5)	ghar-mā	kohī	chaina
	house-LOC	anyone	COP.EXIST.NEG.3SG

'There is nobody in the house.'

The choice is not stylistic at all. Ho is a classifier/identifier of the subject; cha is a locative describing the subject's state, and you can't substitute one for the other. The two forms are Indo-Aryan. Ho is from the Hunu 'to be, to become' paradigm (Turner 1931 s.v. hunu),

from Sanskrit bhavati. *'Cha'* is probably cognate with Bengali *āche*. Whatever is regional about F2 is in the pattern, not in the forms.

The pre-IE languages outside of South Asia have a single copula for both uses: Latin *esse*, Ancient Greek *eimí* and Avestan *ah- all* preserve the one PIE copula *h_1*es-*. F2 is therefore not inherited, and H1 is supported. Later branches, however, show that the opposition can arise independently. Spanish has ser (used for identity and class membership) and estar (used for location and temporary states), with estar descending from Latin *stare* 'stand'. Irish has the copula *'is'*, which expresses essential or inherent qualities, and the substantive verb *bí* (present *tá*), which expresses temporary qualities, existence, and location (Stenson 1981:94). All of this helps undermine the argument that F2 must be due to contact (Section 7.1).

Within South Asia, the comparison is summarised in Table 5.

Table 5. *Identity, existential and negative existential copulas*

Language	Identity	Existence/ location	Negative existence	Source
Nepali	ho	cha	chaina	Acharya 1991: 85–87
Hindi-Urdu	hai	hai	nahī ̃hai	Kachru 2006: 145–146
Bengali	(no overt copula; neg. *noy*)	āche	nei	Thompson 2012: 141–142, 193; Thompson 2006
Marathi	āhe	āhe	nāhī	Dhongde and Wali 2009: 236
Sanskrit	*asti* or zero	asti	nāsti	Whitney 1889: 636
Newar	*khaḥ* (neg. *makhu*)	du	madu	Hale and Shrestha 2006: 62
Limbu	ro	yuŋ	hop	van Driem 1987: 55–68
Kham	(no overt copula; neg. *ma:hkə*)	le (Takale), 'li (Taka)	'male	Watters 2004: 24, 28; Watters 2008: 32–33; Grunow-Hårsta 2007: 176
Magar	ale	le	ma-le	Grunow-Hårsta 2007: 160, 164, 167; Grunow-Hårsta 2011: 105, 112

Hindi-Urdu and Marathi make use of a single copula for both identity and existence. Bengali has an existential *āche* and a suppletive negative existential *nei*, but no overt copula for identity in the present tense. The existential is therefore common to the eastern Indo-Aryan languages and can't be said to be a feature of the Himalayan region. Bengali lacks an overt copula for identity, unlike the existential. Bengali does distinguish between negative identity *(noy)* and negative existential *(nei)* (Thompson, 2006), though in the affirmative present this opposition is not overtly marked.

Newar, Limbu and Magar have an overt identity copula which is distinct from the existential. Newar has *khaḥ* and *du* as its identity and existential copulas respectively (with negatives *makhu* and *madu*); Limbu has *ro* and *yuŋ*; and Magar has *ale* and *le* (Table 5). In Magar, the equative copula *is* the locative existential copula, which also functions as the marker of the imperfective aspect in the verb (Grunow-Hårsta 2011: 105 n.17, 112 n.26). The two dialects differ: in Syangja (western) Magar, nominal and adjectival predicates take

'ale', whereas in Tanahu (eastern) Magar, *'ale'* is being replaced by *'le'* (Grunow-Hårsta 2011: 112, 120, 122). The Limbu sentences below are parallel to (3)-(5):

(6)	en	luŋ	ro
	this	stone	COP.ID

'This is a stone.'

(7)	luŋ	na	yuŋ
	stone	there	EXIST

'The stone is there.'

(8)	him-mo	mantha	hop
	house-LOC	person	EXIST.NEG

'There's nobody in the house.'

In Limbu, *'ro'* expresses identity, *yuŋ expresses* existence and location, and 'hop' expresses negative existence (i.e. the opposite of existence) and is the functional equivalent of Nepali 'ho', 'cha' and 'chaina'.

Kham fits the picture only in part. Its existential is 'li' in the Takale dialect (Watters 2002: 289, cited in Grunow-Hårsta 2007: 176) and *'li'* in the Taka dialect as in *'li-zya* 'he is' (Watters 2004 p.28) and its negative existential is *'male'* (Watters 2004: 24). In affirmative equatives Kham normally has no copula so that 'ao mi:' can mean either 'this person' or 'this is a person' (Watters 2008: 33). In the negative however it does distinguish the negative equative copula 'ma:hkə' from the negative existential *'male'* (Watters 2004: 24; Watters 2008: 32-33) and the emphatic marker 'zə' is grammaticalising as an affirmative equative copula (Watters 2008 pp.32-33). Kham therefore makes the identity/existence distinction in its negative paradigm and may be developing an overt affirmative copula, but it has no overt identity copula, which is what defines F2. In this respect, it is similar to Bengali, which also has no overt identity copula in the present but does have an opposition between negative 'noy' and negative existential 'nei' (Thompson 2006). For the Western languages, among which Khas-

Nepali first developed; the outcome is mixed: Magar has the opposition, and Kham has it only in the negative. This does not support a specifically Western source for F2. It is compatible with three possibilities: convergence after Nepali had spread into central and eastern Nepal; an opposition that Kham is only now extending to the affirmative; or independent development reinforced by contact. The present evidence cannot decide between these three.

The opposition is a contact-layer candidate at the first three stages of the procedure (Section 5.4), as it's missing from Sanskrit and the plains Indo-Aryan languages examined, and present in three of the four Tibeto-Burman languages. It fails the contact test in stage 4, though. The Spanish and Irish copulas show that the same opposition arises within Indo-European without any South Asian contact at all; the Indo-Aryan languages of other contact zones have not been shown to have it, so H4 cannot be established; and Kham (one

of the two western Tibeto-Burman languages) has it only in the negative. F2 is therefore left unresolved: on the evidence currently available, we cannot tell whether it resulted from contact or developed independently (Section 7.2).

6.3 Agent marking with -le: an inherited system with an extension

In perfective transitive clauses the agent takes *-le*, and in ordinary non-past clauses it is usually unmarked:

(9)	mai-le	bhāt	khā-ẽ
	1SG-ERG	rice	eat-PST.1SG

'I ate rice.'

(10)	ma	bhāt	khā-n-chu
	1SG	rice	eat-NPST-1SG

'I eat rice.'

This is not evidence of contact. The ergative constructions in New Indo-Aryan generally derive from Old Indo-Aryan participial constructions with an instrumental agent, and Nepali shares the perfective core with Hindi-Urdu, Punjabi and Marathi (Masica 1991: 344; Verbeke 2013a). In Nepali, -le is, in fact, a new instrumental marker that has taken over the agentive function (Masica 1991: 344). The core system is inherited.

Nepali also allows *'-le'* in non-past clauses:

(11)	rām-le	nepālī	rāmro	bol-cha
	Ram-ERG	Nepali	well	speak-NPST.3SG

'Ram speaks Nepali well.' (a characteristic of Ram)

Descriptive work agrees that these uses are systematic; they differ only in the conditions that govern them. Li (2007) and Poudel (2017) link the non-past-*le* to individual-level or characteristic predication; R.L. Singh (2023) links it to habitual clauses with an individual-level reading; and Verbeke (2013b) links it to focus and a sense of completeness/determination. F3 is defined only as agent marking beyond the perfective and is conditioned in non-past clauses by semantic or discourse factors (the exact nature of the conditioning is not specified), and no conclusion is reached as to which of the above analyses is correct.

The pattern is by no means marginal or regional. Published corpus data demonstrate this. From a sample of the Nepali National Spoken Corpus (four interviews, 67 minutes, 14 speakers, 2,845 clauses), Lindemann (2019: 43, 57) reports ergative-marked subjects in transitive imperfective clauses with overt subjects in the speech of all speakers, at rates ranging from 29% to 67% (mean 48%). In an acceptability survey carried out in Kathmandu, ergative forms in simple present clauses received a mean rating of around 4.5 on a 5-point scale (Lindemann 2019 p. 57). The corpus speakers were from six of the seven provinces of Nepal (Lindemann 2019: 55). These figures are frequencies from spontaneous speech, added to the categorical judgements on which the comparison is otherwise based; no comparable figures exist for the Tibeto-Burman languages.

In the Indo-Aryan languages examined in the plains, ergativity appears only in the perfective (Masica 1991; Bjorkman 2018), and Bengali has no ergativity whatsoever (Phillips 2014: 104-105). There is one construction-specific exception. In Punjabi and in sub-standard Hindi-Urdu, the agent of an obligational construction with an infinitive and a copula can take the ergative marker (*rām ne śādī karnī hai* 'Ram is to marry' (Khokhlova 2013: 94)). This is a construction-specific extension and not something that occurs more generally across non-past clauses on the basis of semantic or discourse factors and so doesn't satisfy F3 as defined. The plain Indo-Aryan evidence therefore supports H3 for F3.

Outside the plains, two Indo-Aryan languages extend agent marking into the imperfective. Assamese (the only member of the eastern branch to mark agents at all) does so in any aspect, and Shina has a separate agent marker.

-s(a) in the imperfective (Verbeke 2013a; Phillips 2014: 104-105). Assamese is spoken in the Brahmaputra Valley alongside Tibeto-Burman languages and Shina alongside Tibetan languages. So, within Indo-Aryan, agent marking extends beyond the perfective in Nepali, Assamese, and Shina (all spoken in contact with Tibeto-Burman languages) (H4 is supported).

The Indo-Aryan languages most closely related to Nepali outside Nepal further support this. In Kumauni and Garhwali (the Central Pahari languages spoken to the west of the Kali River), the agent marker continues as *'le'* (Kumauni *-le*, *-el*, *-l*; Garhwali *-n* or *-l*), and Stroński (2010: 87–89) describes it in the perfective. He describes agent marking in the perfective as being part of the stable split ergative pattern that has persisted from the earliest records. Outside the perfective, he describes agent marking only in an obligational construction with the infinitive, which in Garhwali tends to be a simple future (Stroński, 2010, pp. 92–93). This is simply an extension of the kind of construction found in Punjabi (Khokhlova, 2013), not agent marking that functions as a result of the clause's semantics or discourse in the non-past. If these descriptions are exhaustive, then F3 does not extend to the nearest Indo-Aryan relatives of Nepali as a contact-based account centred on Nepal would predict, and there is no evidence of a shared Pahari inheritance. Speakers should confirm this point, as absence from a description is considerably weaker evidence than a negative judgement (Section 7.5).

The Tibeto-Burman languages of Nepal all point in the same direction. In Newar, the ergative is independent of aspect (Genetti 2007; Phillips 2014: 105). In Kham the split in agent marking is a function of person rather than aspect, with 1st and 2nd person high and 3rd person low on the nominal hierarchy (Watters 2002: 2). In western Magar the ergative is consistently ergative; in eastern Magar it is only ergative in the perfective (analogous to the Nepali pattern) (Grunow-Hårsta 2011: 101 n. 6). More generally the dominant Tibeto-Burman pattern is a 'pragmatic' ergative whose presence is a function of agentivity, perfectivity and contrast (DeLancey 2011; Chelliah & Hyslop 2011). Limbu exhibits a pattern very similar to (11):

(12)	rām-re	nepālī	nuba	pāːp
	Ram-ERG	Nepali	good	speak.NPST

'Ram speaks Nepali well.' (a characteristic of Ram)

(13)	hat-re	chadu?	rām-re	chadu
	who-ERG	eat.PST	Ram-ERG	eat.PST

'Who ate it?' 'Ram ate it.'

In (12), *-re* marks the agent of a non-past verb with a characteristic reading (as -le does in (11)); in (13) it marks the agent in a focused sense.

Syangja (western) Magar is the same in this respect. In (14), from a narrative, the agent of an imperfective verb takes the ergative marker *-e*:

(14)	rokotyak-e	hi	jya-ke	yaɦ-le	ma-de-haŋ
	frog-ERG	what	eat-NMLZ	give-IMPF	NEG-say-COND

‘What if the frog does not give (you) anything to eat?’ (Syangja Magar; Grunow-Hårsta 2007: 167)

The verb has the imperfective marker 'le'; the agent, however, is marked with '-e', as it would not be in Eastern Magar or Hindi-Urdu.

In none of Newar, Kham, western Magar, or Limbu is agent marking restricted to the perfective. Eastern Magar (which is restricted to the perfective) is described as following the Nepali pattern, implying that influence in this area may have been bidirectional, as Noonan (2003) found for other features. Under the procedure in Section 5.4, the core perfective system of *-le* is inherited, and its extension beyond the perfective is a contact-layer candidate that passes the contact test (H3 and H4 are supported).

6.4 Summary

Table 6. *Classification of the three features*

Feature	Old IE outside South Asia	Sanskrit	Plains IA	IA in other TB contact zones	Tibeto-Burman	Classification
F1 Reduplication	absent (productive type)	present	present	not coded	present	Old regional layer
F2 Copula opposition	absent; present in later Spanish, Irish	absent	absent	not coded	present (Newar, Limbu, Magar); negative only (Kham)	Unresolved: fails contact test
F3 Agent marking beyond the perfective	absent	absent	absent (except obligational constructions); not described for Kumauni, Garhwali	present (Assamese, Shina)	present (Newar, Kham, western Magar, Limbu)	Contact-layer candidate; passes contact test

The procedure splits the three features. The calibration case is in the old regional layer, as predicted. F3 passes the contact test: it occurs only in Indo-Aryan languages in contact

with Tibeto-Burman, is not attested in the nearest Indo-Aryan relatives of Nepali outside Nepal, and is the normal state in the Tibeto-Burman languages examined. It is therefore the clearest evidence in this study of Tibeto-Burman influence on Nepali grammar. F2 fails step 4 and remains unresolved: the pattern is found in the Tibeto-Burman languages of Nepal but also arises independently within Indo-European, so the comparison cannot separate contact from independent development.

7 Discussion

7.1 Alternative explanations for F3

For F3, the feature which passes the contact test has four possible explanations: inheritance, independent development, convergence and borrowing/substrate transfer. Inheritance is excluded in the form as defined. The -le marker is Indo-Aryan, but its extension beyond the perfective is not found in Sanskrit, any of the Plains Indo-Aryan languages examined, or, as far as published descriptions go, Kumauni or Garhwali.

Independent development is unlikely (though of course not impossible). It would be a curious coincidence if Nepali had extended -le beyond the perfective through an internal change and if the only other Indo-Aryan languages to have done so (Assamese and Shina) are also in contact with Tibeto-Burman, if in Shina the very perfective marker seems to be of Tibetan origin and if the nearest Pahari languages to Nepali have not made the change.

Contact with Tibeto-Burman is the explanation the data support, and in Section 3, it takes the form of convergence through pattern replication. The pattern appears in both Nepali and the Tibeto-Burman languages examined; Nepali has it via inherited Indo-Aryan material, and Indo-Aryan languages with no Tibeto-Burman contact do not. That is exactly what you would expect to see in a contact situation (Matras & Sakel, 2007; Heine & Kuteva, 2005). It is consistent with Verbeke's suggestion that bilingualism in Nepali and Newar may have been a factor (Phillips, 2014, p.105). The evidence here is more general, though, and points to the region's Tibeto-Burman languages rather than Newar alone. For F3, then, the answer to the question posed in Section 1 is yes: this part of Nepali grammar patterns with its Tibeto-Burman neighbours rather than with Sanskrit and the plains Indo-Aryan languages, and contact is the best explanation for why.

What remains undemonstrated is the mechanism. Neither substrate transfer nor borrowing in Thomason's (2001) sense is demonstrated. You would need evidence that the pattern was absent in earlier stages of Khas and then appeared through contact with a particular source. That kind of historical and dialectal evidence is not yet available.

7.2 What the failure of F2 shows

A procedure that would classify every single Tibeto-Burman-like feature of Nepali as contact-induced would be of little use, as it would just restate what it was trying to test. F2 shows that it can, in fact, give a negative. The copula opposition looks like a pretty good case at first glance. Nepali and three of the four Tibeto-Burman languages under consideration have a copula opposition (i.e., distinct copulas for identity and existence). In contrast, Sanskrit and the plains Indo-Aryan languages lack this opposition. It fails, however, at the controls that the procedure adds to that comparison: the opposition is

present in Indo-European without any South Asian contact (as shown by Spanish and Irish); it is not uniform across the Tibeto-Burman languages of the region where Khas-Nepali arose, and no evidence suggests it occurs in any other Indo-Aryan language in contact with Tibeto-Burman.

The failure gives us information in a second way. In Spanish, 'estar' is the verb 'to stand' (a Latin verb) which has been taken over to express location and temporary states. In Nepali, 'ho' is from the hunu 'be, become' paradigm (Turner 1931 s.v. *hunu*). Kham shows yet another way in which the identity/existence distinction is first established negatively and then later grammatised as an affirmative equative copula from an emphatic marker (Watters, 2008, pp. 32-33). Different paths in completely unrelated languages lead to the same oppositions, as we would expect for such a readily available feature. Contact is notoriously difficult to demonstrate in such cases, as contact-induced and internal grammaticalisation will produce the same outcome (Heine & Kuteva, 2003). F2 may still be the result of contact in some way (e.g. as an internal development which a Tibeto-Burman model has reinforced). That would, however, require the dialectal and historical evidence discussed in Section 7.5.

7.3 Substrate or convergence?

The title asks whether a feature that passes the test reflects areal convergence or a Tibeto-Burman substrate. Within the framework of Section 3, a substrate account is one of the impositions by speakers shifting to Khas. That account predicts features beyond those we have looked at here (phonological features of Tibeto-Burman type and comparatively little Tibeto-Burman vocabulary in the core lexicon (Thomason & Kaufman, 1988; Van Coetsem, 2000)). On Ross's (2013) diagnostic, grammatical calquing of the type we see in F3 is more the sort of thing you would expect from speakers who are bilingual from childhood. By contrast, phonological transfer and the calquing of specific constructions are more what you would expect from adult shift.

Earlier fieldwork also informs what matters for the social distribution of F3. Lindemann (2019: 55–56) notes that both the teachers and a Nepali linguist he consulted viewed the imperfective *'-le'* as prescriptively wrong and as a feature of Nepali spoken by speakers of other languages (in particular the Magar languages) in the area west of Kathmandu; Grierson had already ascribed the use of the form to Tibeto-Burman influence (Lindemann 2019: 8, 80 n. 10). Lindemann, however, found the pattern in the speech of every consultant and every corpus speaker, including Brahmin and Chhetri speakers from the east (2019: 56). The common perception is thus that F3 is a feature of second language Nepali, as one would expect under a shift account, while the current distribution of the feature would seem to indicate that it is in fact a feature of all the speakers in the sample. The two observations are fully compatible with either account, and they illustrate why the current distribution of a feature cannot, by itself, tell us anything about its origin.

The current evidence does not settle the matter. F3 is a case of pattern replication with inherited material that both accounts can produce, and the Magar evidence from the east suggests that influence in this area has also gone from Nepali into Tibeto-Burman (Grunow-Hårsta 2011: 101 n.6), as mutual bilingualism would predict (Noonan, 2003).

The framework provides a series of independent tests. If all the expected features of a Tibeto-Burman phonology, a lexical profile with very little Tibeto-Burman core vocabulary and early attestation of F3 in the Khas documents, were found, then the case for a substrate would be much stronger. If, on the other hand, F3 were found to be late and the Nepali phonology showed no evidence of a Tibeto-Burman pattern, then convergence via bilingualism would be the better explanation. In either case, the question becomes one of fact rather than interpretation.

7.4 Limits

Three features simply cannot tell us the full extent of a contact layer. The Tibeto-Burman sample is small, the F2 results for the two western languages differ markedly, and the two Magar dialects themselves differ on both features (Sections 6.2 and 6.3). We coded the Indo-Aryan languages from the other contact areas and the Central Pahari languages only for F3, and we have published descriptions for Kumauni and Garhwali. So speakers have to establish the absence of F3. The Limbu data come from a single speaker, and the Tibeto-Burman agent marking data are based on published descriptions rather than corpus counts; for Nepali, the frequencies come from a single corpus sample (Lindemann 2019). The comparison is synchronic, of course: it can only show how Nepali patterns relate to those of its neighbours, not when or how they came to be so.

7.5 Further work

The three lines of work proceed directly. The first is dialectal in nature and could be done with very modest resources. Elicited matched data from speakers of Doteli and Kumauni (using the equivalents of 'This is a stone', 'The stone is there' and 'Ram speaks Nepali well') would allow us to test whether these varieties have the copula opposition and the extension of agent marking, and thus whether the features stop at the Kali River or extend across the Pahari continuum. Lindemann's (2019) corpus data show that counting non-past *-le* in spontaneous speech is feasible; extending that count to the Far-Western varieties and to the Tibeto-Burman model languages would let us compare the conditioning of -le with that of the models, as variationist work (Poplack & Levey, 2010). The second line of work is historical: going back to the copulas and agent marking in the Early Khas inscriptions and early Nepali documents would provide the time depth that the procedure itself cannot. The third line of work is to apply the same procedure to other candidates such as the mirative *-echa*, reduced gender agreement, honorific grammar, ideophones and so on, as well as to the phonological features which have such a direct bearing on the substrate question (Section 7.3).

8 Conclusion

Nepali is of Indo-European origin and inherits its inflectional morphology, tense system, and much of its core vocabulary. This article looked at a more specific question: do any of its key grammatical categories align more closely with its Tibeto-Burman neighbours than with Sanskrit and the Indo-Aryan languages of the plains?

For one of the three features, the procedure answers yes. Reduplication is not evidence of contact in the Himalayas: it is missing from pre-Himalayan IE languages outside South

Asia but is present in Sanskrit and part of an older stratum of South Asian languages that Nepali shares with Hindi. Agent marking beyond the perfective passes the contact test: it occurs only in Indo-Aryan languages in contact with Tibeto-Burman, is not attested in the Pahari languages closest to Nepali, and is the norm in the Tibeto-Burman languages we have looked at. On the present evidence, then, the Tibeto-Burman languages of the Himalaya have left a demonstrable mark on Nepali grammar in agent marking: an inherited Indo-Aryan marker, -le, now follows a Tibeto-Burman pattern of use. The copula opposition (despite its similarity to Tibeto-Burman) remains unresolved, because the same opposition arises independently within Indo-European (Spanish, Irish) and Kham has it only in the negative.

The evidence does not yet show how this influence entered Nepali: whether F3 was borrowed under long-term bilingualism or carried into Nepali by speakers shifting from Tibeto-Burman languages. Either scenario would require evidence that the feature was absent from earlier Khas and was acquired through contact with a particular language. The steps outlined are therefore historical and dialectal.

The key result is methodological. Without documentation of a language's pre-contact state, the criteria for contact-induced change cannot be applied directly. They can, however, be approximated using a set of comparative proxies selected in advance and applied in a fixed order. Applied to Nepali, the procedure confirms a well-known areal feature, leaves a plausible contact candidate unresolved, and identifies one feature, agent marking beyond the perfective, that passes every criterion and shows Tibeto-Burman influence on Nepali grammar. It thus turns the question in the title into one which further dialectal and historical data can answer. Describing Nepali in these terms, rather than in terms of Sanskrit alone, gives a more accurate account of the language and its history.

Data availability

All forms other than the Nepali and Limbu examples are drawn from the published sources cited (with page references where available). The Nepali and Limbu examples are given in full in the text.

Abbreviations

1. first person; 2. second person; 3. third person; 4. COND conditional; 5. COP copula; 6. ERG ergative; 7. EXIST existential; 8. HON honorific; 9. IA Indo-Aryan; 10. ID identity; 11. IE Indo-European; 12. IMP imperative; 13. IMPF imperfective; 14. INF infinitive; 15. LOC locative; 16. NEG negative; 17. NMLZ nominaliser; 18. NPST non-past; 19. PST past; 20. SG singular; 21. TB Tibeto-Burman

References

1. Abbi, A. (1992). Reduplication in South Asian languages: An areal, typological and historical study. New Delhi: Allied Publishers. https://wals.info/refdb/record/Abbi-1992
2. Acharya, J. (1991). A descriptive grammar of Nepali and an analysed corpus. Washington, DC: Georgetown University Press. https://wals.info/refdb/record/Acharya-1991
3. Bjorkman, B. M. (2018). Ergative as perfective oblique. Syntax 21(4). 321–361. https://doi.org/10.1111/synt.12162
4. Campbell, L. (2006). Areal linguistics: A closer scrutiny. In Yaron Matras, April McMahon & Nigel Vincent (eds.), Linguistic areas: Convergence in historical and typological perspective, 1–31. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230287617_1
5. Chelliah, S. L. & Gwendolyn Hyslop. (2011). Introduction to a special issue on optional case marking in Tibeto-Burman. Linguistics of the Tibeto-Burman Area 34(2). 1–7. https://doi.org/10.32655/LTBA.34.2.01
6. Van Coetsem, Frans. 2000. A general and unified theory of the transmission process in language contact. Heidelberg: Winter. https://search.worldcat.org/title/general-and-unified-theory-of-the-transmission-process-in-language-contact/oclc/477268034
7. DeLancey, S. (2011). “Optional” “ergativity” in Tibeto-Burman languages. Linguistics of the Tibeto-Burman Area 34(2). 9–20. https://doi.org/10.32655/LTBA.34.2.02
8. Dhongde, R. V. & Kashi Wali. (2009). Marathi (London Oriental and African Language Library 13). Amsterdam: John Benjamins. https://doi.org/10.1075/loall.13
9. Emeneau, M. B. (1956). India as a linguistic area. Language 32(1). 3–16. https://doi.org/10.2307/410649
10. Genetti, C. (2007). A grammar of Dolakha Newar (Mouton Grammar Library 40). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110198812
11. Grunow-Hårsta, K. (2007). Evidentiality and mirativity in Magar. Linguistics of the Tibeto-Burman Area 30(2). 151–194. https://theswissbay.ch/pdf/Books/Linguistics/Mega%20linguistics%20pack/Sino-Tibetan/Magar%20Dialects%20of%20Nepal, %20A%20Descriptive%20Grammar%20of%20Two%20%28Grunow-H%C3%A5rsta%29.pdf
12. Grunow-Hårsta, K. A. (2008). A descriptive grammar of two Magar dialects of Nepal: Tanahu and Syangja Magar. Milwaukee, WI: University of Wisconsin-Milwaukee dissertation.
13. Grunow-Hårsta, K. (2011). Adjectives and adjectivals in Magar. Himalayan Linguistics 10(1). 101–126. https://doi.org/10.5070/H910123570
14. Hale, A. & Kedar P. Shrestha. (2006). Newār (Nepāl Bhāsā) (Languages of the World/Materials 256). Munich: LINCOM Europa. https://glottolog.org/resource/reference/id/313608
15. Heine, B. & Tania Kuteva. (2003). On contact-induced grammaticalization. Studies in Language 27(3). 529–572. https://doi.org/10.1075/sl.27.3.04hei
16. Heine, B. & Tania Kuteva. (2005). Language contact and grammatical change. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511614132
17. Hutt, M. (1988). Nepali: A National Language and Its Literature. New Delhi: Sterling. https://soas-repository.worktribe.com/output/432269/nepali-a-national-language-and-its-literature

18. Ivani, J. K., Netra Paudyal & John Peterson. (2021). Indo-Aryan – 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). 287–326. https://doi.org/10.1515/jsall-2021-2029
19. Kachru, Yamuna. 2006. Hindi (London Oriental and African Language Library 12). Amsterdam: John Benjamins. https://doi.org/10.1075/loall.12
20. Khokhlova, L. V. (2013). Obligational constructions in New Indo-Aryan languages of Western India. Lingua Posnaniensis 55(2). 91–107. https://doi.org/10.2478/linpo-2013-0016
21. Kuiper, F. B. J. (1991). Aryans in the Rigveda. Amsterdam: Rodopi.
22. LaPolla, R. J. (1995). 'Ergative' marking in Tibeto-Burman. In Yoshio Nishi, James A. Matisoff & Yasuhiko Nagano (eds.), New horizons in Tibeto-Burman morphosyntax (Senri Ethnological Studies 41), 189–228. Osaka: National Museum of Ethnology. https://doi.org/10.15021/00002996
23. Li, C. (2007). Split ergativity and split intransitivity in Nepali. Lingua 117(8). 1462–1482. https://doi.org/10.1016/j.lingua.2006.09.002
24. Lindemann, L. (2019). A jewel inlaid: Ergativity and markedness in Nepali. New Haven, CT: Yale University dissertation. https://elischolar.library.yale.edu/ling_graduate/4
25. Masica, C. P. (1976). Defining a linguistic area: South Asia. Chicago: University of Chicago Press. https://archive.org/details/defininglinguist0000masi
26. Masica, C. P. (1991). The Indo-Aryan languages. Cambridge: Cambridge University Press. https://books.google.com/books/about/The_Indo_Aryan_Languages.html?id=J3RSHWePhXwC
27. Matras, Y., & Jeanette Sakel. (2007). Investigating the mechanisms of pattern replication in language convergence. Studies in Language, 31(4). 829–865. https://doi.org/10.1075/sl.31.4.05mat
28. Noonan, M. (2003). Recent language contact in the Nepal Himalaya. In David Bradley, Randy LaPolla, Boyd Michailovsky & Graham Thurgood (eds.), Language variation: Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff (Pacific Linguistics 555), 65–88. Canberra: Pacific Linguistics. https://doi.org/10.11588/xarep.00000189
29. Peterson, J. (2010). Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India. Himalayan Linguistics 9(2). 56–86. https://doi.org/10.5070/H99223479
30. Phillips, M. (2014). Review of Alignment and Ergativity in New Indo-Aryan Languages, by Saartje Verbeke. Journal of South Asian Languages and Linguistics 1(1). 103–110. https://doi.org/10.1515/jsall-2014-0006
31. Poplack, S., & Stephen Levey. (2010). Contact-induced grammatical change: A warning tale. In Peter Auer & Jürgen Erich Schmidt (eds.), Language and space: An international handbook of linguistic variation, vol. 1, 391–418. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110220278.391
32. Poudel, Tikaram. 2017. The semantics of the ergative in Nepali. Gipan 3(2). 61–86. https://doi.org/10.3126/gipan.v3i2.48900
33. Ross, M. (2007). Calquing and metatyping. Journal of Language Contact, 1(1). 116–143. https://doi.org/10.1163/000000007792548341
34. Ross, M. (2013). Diagnosing contact processes from their outcomes: The importance of life stages. Journal of Language Contact, 6(1). 5–47. https://doi.org/10.1163/19552629-006001002

35. Singh, R. L. (2023). Habituality, perfectivity and ergativity in Nepali. Journal of Productive Discourse, 1(1). 6–14. https://doi.org/10.5281/zenodo.7880363
36. Southworth, F. C. (2005). Linguistic archaeology of South Asia. London: Routledge Curzon. https://doi.org/10.4324/9780203412916
37. Stenson, N. (1981). Studies in Irish syntax. Tübingen: Gunter Narr. https://catalog.hathitrust.org/Record/00055957
38. Stroński, K. (2010). Non-nominative subjects in Rajasthani and Central Pahari: The status of the ergative and obligatory constructions. Lingua Posnaniensis 52(1). 81–97. https://doi.org/10.2478/v10122-010-0007-9
39. Thomason, S. G. (2001). Language contact: An introduction. Edinburgh: Edinburgh University Press. https://doi.org/10.1515/9781474473125
40. Thomason, S. G., & Terrence Kaufman. (1988). Language contact, creolisation, and genetic linguistics. Berkeley: University of California Press. https://doi.org/10.1525/9780520912793
41. Thompson, H.-R. (2006). Negation patterns in Bengali. Bulletin of the School of Oriental and African Studies, 69(2). 243–265. https://doi.org/10.1017/S0041977X06000115
42. Thompson, H.-R. (2012). Bengali (London Oriental and African Language Library 18). Amsterdam: John Benjamins. https://doi.org/10.1075/loall.18
43. Turner, R. L. (1931). A Comparative and Etymological Dictionary of the Nepali Language. London: Kegan Paul, Trench, Trübner. https://dsal.uchicago.edu/dictionaries/turner/
44. van Driem, George. (1987). A grammar of Limbu (Mouton Grammar Library 4). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110846812
45. van Driem, George. (2013). The ancestry of Tibetans. In Grey Tuttle, Kunsang Gya, Karma Dare & Jonathan Wilber (eds.), The Third International Conference on Tibetan Language, vol. 1, 363–397. New York: Trace Foundation. https://researchers.une.edu.au/en/publications/the-ancestry-of-tibetan-5/
46. Verbeke, Saartje. 2013a. Alignment and ergativity in New Indo-Aryan languages (Empirical Approaches to Language Typology 51). Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110292671
47. Verbeke, S. (2013b). Differential subject marking in Nepali: The agent marker 'le' in imperfective constructions. Linguistics 51(3). 585–610. https://doi.org/10.1515/ling-2013-0021
48. Watters, D. E. (2002). A grammar of Kham. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486883
49. Watters, D. E. (2004). A dictionary of Kham: Taka dialect (a Tibeto-Burman language of Nepal). Kirtipur: Central Department of Linguistics, Tribhuvan University. https://glottolog.org/resource/reference/id/164640
50. Watters, D. E. (2008). Nominalisation in the Kiranti and Central Himalayan languages of Nepal. Linguistics of the Tibeto-Burman Area 31(2). 1–43. https://doi.org/10.15144/LTBA-31.2.1
51. Weinreich, U. (1953). Languages in contact: Findings and problems. New York: Linguistic Circle of New York. https://api.pageplace.de/preview/DT0400.9783110802177_A19970832/preview-9783110802177_A19970832.pdf
52. Whitney, W. D. (1889). Sanskrit grammar, 2nd edn. Cambridge, MA: Harvard University Press. http://gretil.sub.uni-goettingen.de/gretil_elib/Whi889__Whitney_SanskritGrammar.pdf

53. Winford, D. (2005). Contact-induced changes: Classification and processes. Diachronica 22(2). 373–427. https://doi.org/10.1075/dia.22.2.05win
54. Witzel, M. (1999). Substrate languages in Old Indo-Aryan (Ṛgvedic, Middle and Late Vedic). Electronic Journal of Vedic Studies 5(1). 1–67. https://doi.org/10.11588/ejvs.1999.1.828