Beyond Hill and Valley: A Sociogeographic Analysis of Argument Coding Complexity in the Eastern Himalayan Region

# Introduction

Recently, many studies have investigated the link between the complexity of a language and the scale (Nettle, 2012), physical environment (Urban and Moran, 2021) or isolation (Bromham et al., 2025) of the societies that speak them. Particularly, whether a language is spoken by a large population, made up of heterogenous groups and loose social networks: so-called exoteric societies; or by a smaller, homogenous population with tight social networks: so–called esoteric societies (Wray and Grace, 2007) occupies a central position in these debates (Shcherbakova et al. 2023). Language contact is theorized to play a crucial role in this dynamic (Nichols, 2009): when languages are acquired by adult second-language learners in contact situations, they are theorized to undergo simplification of morphological paradigms. Conversely, languages with minimal contact in esoteric societies may retain or even develop more complex morphological systems, due to the greater shared knowledge their close-knit social networks allow (Wray and Grace, 2007).

One region where this link between societal structure and linguistic complexity has been explicitly discussed is the Eastern Himalayan Region (henceforth: EHR) (DeLancey, 2015). The EHR is home to considerable diversity ecologically (Basnet et al., 2019) and linguistically (Hua et al., 2019:7), making it an appropriate testing ground for exploring hypotheses that link linguistic complexity to the environment. The EHR stretches from Eastern Nepal to the Indo-Burma Patkai ranges, containing four distinct language families, and the epicenter of Trans-Himalayan[[1]](#footnote-1) phylogenetic diversity (Post and Burling, 2017). Here, DeLancey argues that societal structure reflects ecological setting, producing ‘Hill’ and ‘Valley’ societies, named after their respective ecological conditions.

This paper investigates complexity in how core arguments are coded in Trans-Himalayan, the largest (in number of named languages) language family in the region. Proto-Trans-Himalayan is claimed to have had a complex system of verbal agreement that has been lost in many languages of this region, linked to the patterns of migration and changing settlement histories of linguistic groups (DeLancey, 2013). On the other hand, languages of this area are claimed to be ‘Indospheric’ (Genetti, 2009:03)[[2]](#footnote-2), case marking being a characteristic trait, suggesting areal dynamics at play. I look at if complexity in core argument marking, whether nominal (case marking) or verbal (agreement) expression, is indeed linked to Hill or Valley societal setting. Furthermore, I test whether sociocultural traits linked to Hill or Valley societies can individually predict complexity.

I find that The Hill/Valley framework overall does not effectively explain the complexity of core argument coding in the languages of this region. However, the component sociocultural traits that give rise to Hill/Valley societies, when analyzed individually, show consistent negative (but non-significant) direction as predicted, and one – whether a language is spoken as an L2, is a statistically significant predictor of morphological simplification. I also find that the sociocultural classification of a society as Hill/Valley is not actually predicted by the presumed ecological determinants, suggesting a more complex account of the dynamics between environment, sociocultural traits, and language complexity. Instead, considering specific language histories, the spatial dynamics of convergence, and the diachronic processes involved in complexification/simplification illuminates a better understanding of how ecological environment, societal setting and linguistic complexity are linked.

# Background

## What is core argument coding?

Core argument is a widely used term in the linguistics literature (Shcherbakova et al., 2024, Haspelmath, 2000 et al., Comrie, 1979) but lacks a universally accepted definition. Generally, it refers to arguments (non-omissible noun phrases to a certain predicate) that express grammatical relations. However, identifying grammatical relations in a particular language can be difficult, such as in Manipuri, a language of the EHR, where core arguments are marked according to semantic role, rather than grammaticalized ‘subject’ or ‘object’ contexts (Chelliah, 1997:109).

Thus, core arguments are often identified with some reference to semantic macroroles (Haspelmath, 2011:543, Foley and van Valin, 1984) in addition to grammatical structure. Minimally, this gives the set of “S”, “A” and “P” as labels for core arguments that are often used for cross-linguistic comparison. “S” refers to the sole argument of the intransitive verb, “A” the ‘actor-like’ argument in a transitive verb, and “P” the ‘less-actor-like’ argument of a transitive verb (Bickel, 2010).

Identifying core arguments in a particular language is not straightforward, given that the ‘S/A/P’ labels invoke syntax via the valence of the verb. Generally, there are no crosslinguistically applicable syntactic criteria for establishing transitivity (Haspelmath, 2011:544). One solution to this is defining A and P as the actor-like and less-actor-like arguments only in the major construction that expresses a prototypical action (typically physical effect verbs, such as kill/break). Lazard (2002:153) notes that this construction/valency frame generally forms the canonical mode of distinguishing ‘who is doing what to whom’. Similarly, S can be defined as the noun in an intransitive sentence that receives the morphosyntactic treatment ‘typical’ to a single argument of a one-argument predicate (Andrews, 2007:68).[[3]](#footnote-3)

Core argument coding (the morphological marking of these arguments that signals grammatical relations) is a good choice for understanding hypotheses about language complexity in this region for three reasons. The first is that all languages must distinguish grammatical relations in some way, and it is amongst the best described parts of grammar. This enables comparison, especially in an area where documentation is sparse and emerging. Secondly, it is an established linguistic subsystem for which claims about complexity have been studied, such as how the number of forms used in core argument coding may direct reflecting an increase/loss of complexity (Bentz and Winter, 2013) or how different strategies of encoding core arguments (case marking, verbal agreement or word order) may trade-off complexity (Sinnemäki 2008).

The third and most compelling reason for the focus on core argument coding is that it is explicitly the subject of claims surrounding the changes in complexity in Trans-Himalayan languages as a response to the social environment (DeLancey, 2014). Proto-Trans-Himalayan is suggested to have had a complex system of verbal agreement (DeLancey, 2023) that has been lost, retained, or even innovated in different subgroups of Trans-Himalayan. DeLancey suggests the archaic verbal agreement system has only been retained in some languages due to the influence of the mountain environment, which has made many of these languages inaccessible (and hence isolated) until recently (2015:63).

This is not a direct effect of the environment, of course. DeLancey (2014:78) argues that ecological setting creates a cultural distinction in the region: the Hill/Valley distinction (Scott, 2009). This distinction then gives rise to the differing profiles of linguistic complexity we find in this region. Hill languages are spoken by few people, in esoteric communities, allowing them to retain their archaic morphology (or re-innovate something complex, as in the case of Lai, DeLancey 2014:65). Valley languages are languages of empire, spoken by not just the native community but also as a second language, and have undergone erosion of the proposed complex Proto-Trans-Himalayan verbal system. In the original Hill/Valley framework proposed by Scott (2009), Hill and Valley societies are associated with a number of sociocultural traits that arise from their environments, that we can use to categorize them.

## The Hill/Valley Distinction

Scott’s (2009) defines *Zomia*, or Upland Southeast Asia – which he describes as the last remaining non-state enclosure in the world. Between India, China, Bangladesh and Mainland Southeast Asia, it ‘lies at a great distance from the main centers of economic activity; it bestrides a contact zone between eight nation-states and several religious traditions and cosmologies’ (p. 14), where ‘variety rather than uniformity is its trademark’ (p. 16). DeLancey (2014) applies this framework to Northeast India, which is considered part (and central) to the Eastern Himalayan linguistic region (Konnerth et al., 2020), the focus of this paper.



Figure 1: The Eastern Himalayan Region (black box). Original image from Hazarika 2016.

The cultural heterogeneity of *Zomia* can be understood when one views it as arising from groups adapting to the ‘agro-economic possibilities of the region’ , leading to cultural differences with neighbouring groups that occupy different ecological niches, but similarities with groups that lie more than a thousand kilometers away, but at a similar altitude (Scott, 2009:18-19). More importantly, Scott argues that *Zomia* must be understood via the “dialectic of state-formation in the valleys, and statelessness/resistance to statehood in the hills” (p. 19).[[4]](#footnote-4) The altitude and difficulty of traversing terrain determines the ease by which states may be formed or resisted, which in turn correlate with cultural traits that define Hill and Valley societies. This includes (but is not limited to) their style of agriculture, their level of political organization, their total speaker population, and whether their language may be spoken by other groups.[[5]](#footnote-5) These variables are identified and the reasoning behind them is explored in S3.2.

DeLancey (2014) applies this framework to the languages of the Brahmaputra drainage region, the major river system in the EHR. He identifies case marking systems, like those found in Bodo-Garo languages as the creoloid type, due to their reconstructed past as the vehicular language for a large Valley state (DeLancey, 2012). Conversely, complex verbal paradigms such as in Kuki-Chin are associated with their hill settlement, and smaller esoteric communities (called the archaic type) (DeLancey, 2014:61). However, DeLancey admits that not all subgroups fall into this pattern, as some hill communities display ‘simple’ case marking profiles (such as the Tani and Naga languages), and archaic languages do not necessarily inherit the complexity of verbal agreement from Proto-Trans-Himalayan as their name may suggest, instead sometimes re-innovating it, as in the case of Kuki-Chin.

# Study Design

I now turn to the Hill/Valley framework, or rather its associated sociogeographic factors, which may account for argument coding complexity patterns in the Eastern Himalayas. To do this, I designed a representative sample of speech communities (S3.1), identified the sociocultural variables linked to Hill/Valley cultures (S3.2), and the ecological variables that underpin them (S3.3), and lastly, operationalized of the measures of argument coding complexity (S3.4).

## Sampling and Data Collection

The original claim by DeLancey (2014) on Hill and Valley societies and complexity focused on ‘Northeast India’ and referenced the Brahmaputra drainage (p. 61). However, many of the scholars working in this region acknowledge that ‘Northeast India’ refers to a modern-day political division, while the region of linguistic interest is better defined as the Eastern Himalayan Region, which contains adjacent regions in Bangladesh, Bhutan, Nepal, China and Myanmar (Konnerth et al., 2020). I define the EHR then, with reference to Post et al. (2022) who say it extends from Eastern Nepal to the Patkai range in Myanmar. Only subgroups of Trans-Himalayan and their members inside this region were considered. Thus, many subgroups with numerous members (such as Bodic) are only represented by a few members, as most of the daughter languages lie outside the range of the EHR.

This study looks at 32 languages from all major subgroups of Trans-Himalayan that are present in the EHR. The languages were selected by convenience of finding linguistic description and ethnographic description that would enable full coding of all features. However, I aimed for maximal genealogical distance when sampling within a subgroup, following the genealogy in Glottolog (Hammarstrom et al., 2024) as a consensus view amid divergent opinions on Trans-Himalayan phylogeny. Additionally, to control for the imbalances in sampling, subgroup is used as a random effect in all statistical analyses. The full list of languages with their subfamily is given below in Table 1.

Table 1: Table of languages sampled. All details are taken from Glottolog (Hammarstrom et al., 2024). Common names are used in some cases for subfamily, where the name on Glottolog is given in brackets.

|  |  |  |  |
| --- | --- | --- | --- |
| Language | glottocode | Subfamily (node below top-level) | Node below subfamily |
| Rabha | [rabh1238](https://glottolog.org/resource/languoid/id/rabh1238) | Sal (Brahmaputran) | Bodo-Garo |
| Bodo-Mech | [bodo1269](https://glottolog.org/resource/languoid/id/bodo1269) | Sal (Brahmaputran) | Bodo-Garo |
| Garo | [garo1247](https://glottolog.org/resource/languoid/id/garo1247) | Sal (Brahmaputran) | Bodo-Garo |
| Tiwa | [tiwa1253](https://glottolog.org/resource/languoid/id/tiwa1253) | Sal (Brahmaputran) | Bodo-Garo |
| Turung | [sing1264](https://glottolog.org/resource/languoid/id/sing1264) | Sal (Brahmaputran) | Jingpho-Asakian |
| Khiamniungan | [khia1236](https://glottolog.org/resource/languoid/id/khia1236) | Sal (Brahmaputran) | Patkaian |
| Hakhun Tangsa | [noct1238](https://glottolog.org/resource/languoid/id/noct1238) | Sal (Brahmaputran) | Patkaian |
| Hills Karbi | [karb1241](https://glottolog.org/resource/languoid/id/karb1241) | South-Central (Kuki-Chin-Naga) | Karbic |
| Manipuri | [mani1292](https://glottolog.org/resource/languoid/id/mani1292) | South-Central (Kuki-Chin-Naga) | Manipuri |
| Ao Naga | [mong1332](https://glottolog.org/resource/languoid/id/mong1332) | South-Central (Kuki-Chin-Naga) | Angami-Ao |
| Angami | [anga1288](https://glottolog.org/resource/languoid/id/anga1288) | South-Central (Kuki-Chin-Naga) | Angami-Ao |
| Suansu | [suan1234](https://glottolog.org/resource/languoid/id/suan1234) | South-Central (Kuki-Chin-Naga) | Tangkhul-Maring |
| Zeme | [zeme1240](https://glottolog.org/resource/languoid/id/zeme1240) | South-Central (Kuki-Chin-Naga) | Zemeic |
| Mizo | [lush1249](https://glottolog.org/resource/languoid/id/lush1249) | South-Central (Kuki-Chin-Naga) | Kuki-Chin |
| Thado Chin | [thad1238](https://glottolog.org/resource/languoid/id/thad1238) | South-Central (Kuki-Chin-Naga) | Kuki-Chin |
| Lamkang | [lamk1238](https://glottolog.org/resource/languoid/id/lamk1238) | South-Central (Kuki-Chin-Naga) | Kuki-Chin |
| Daai Chin | [daai1236](https://glottolog.org/resource/languoid/id/daai1236) | South-Central (Kuki-Chin-Naga) | Kuki-Chin |
| Bori-Karko | [bori1243](https://glottolog.org/resource/languoid/id/bori1243) | Macro-Tani | Eastern Tani |
| Milang | [mila1245](https://glottolog.org/resource/languoid/id/mila1245) | Macro-Tani | Koro-Holon |
| Mising | [miny1240](https://glottolog.org/resource/languoid/id/miny1240) | Macro-Tani | Mising-Padam-Miri-Minyong |
| Galo | [galo1242](https://glottolog.org/resource/languoid/id/galo1242) | Macro-Tani | Subansiri |
| Kera'a | [idum1241](https://glottolog.org/resource/languoid/id/idum1241) | Digarish | Digarish |
| Hrusso Aka | [hrus1242](https://glottolog.org/resource/languoid/id/hrus1242) | Hruso | Unclear |
| Tibetan | [tibe1272](https://glottolog.org/resource/languoid/id/tibe1272) | Bodic | Central Tibetan |
| Brokpa/Brokpake | [brok1248](https://glottolog.org/resource/languoid/id/brok1248) | Bodic | Southern Tibetic |
| Bjokapakha | [bjok1234](https://glottolog.org/resource/languoid/id/bjok1234) | Bodic | Tshanglic |
| Tamangic | [tama1367](https://glottolog.org/resource/languoid/id/tama1367) | Bodic | Ghale-Tamangic |
| Limbu | [limb1266](https://glottolog.org/resource/languoid/id/limb1266) | Himalayish | Kiranti |
| Lepcha | [lepc1244](https://glottolog.org/resource/languoid/id/lepc1244) | Himalayish | Himalayish |
| Puroik | [puro1234](https://glottolog.org/resource/languoid/id/puro1234) | Kho-Bwa | Kho-Bwa |
| Miji | [miji1239](https://glottolog.org/resource/languoid/id/miji1239) | Miji | Miji |
| Trung/Drung | [drun1238](https://glottolog.org/resource/languoid/id/drun1238) | Nungish | Gunong |

For each language, I consulted research papers, grammars, and ethnographies. Examples of basic intransitive and transitive clauses were extracted, primarily from sections on "argument indexation," "case marking," "agreement," or "person marking”. I also collected the morphosyntactic alignment of the language, following Bickel and Nichols, 2008.

Information on social, cultural, and geographic aspects either came from descriptions in the linguistic source reference, or when inadequate, ethnographies and research papers on the speech community. For two languages (Khiamniungan and Kera’a), I reached out to researchers working on the language to fill out a questionnaire. The coding of this data into variables is described below, but for further details refer to Appendix A, alongside the full questionnaire and full list of references.

## Sociocultural Variables

Based on Scott (2009) and DeLancey (2014), I tracked four key sociocultural variables as indicators of 'Hill’-ness or 'Valley'-ness. These are: intensity of agriculture, speaker population, level of political organization, and whether the language is spoken as an L2. As DeLancey (2014:79) writes: “Valley peoples are wet-rice farmers, with cities, kings, armies, writing, and institutionalized religion... Valley cultures are imperialistic... Thus, Valley languages are necessarily exoteric. In contrast... Hill peoples practice swidden agriculture, and live in small communities with no higher-level political organization... Thus, their languages tend to be esoteric".

This framework connects Valley cultures with exoteric language use, wet-rice cultivation, complex political organization, and a large speaker population. Hill cultures are then associated by opposition with esoteric language use, shifting/swidden (jhum) agriculture, and small, locally organized speech communities. These traits were operationalized as the variables L2\_status, agricultural\_intensity, political\_organization, and population\_size, which were then used to give a Hill and Valley score to languages. These scores were summed to classify a language as Hill or Valley, based on which score was higher (if both scores were equal, the language was classified as Split).

Whether a language is spoken as an L2 is identified by DeLancey (2014, 2015) as a primary factor for morphological simplification. In fact, he strongly suggests that large-scale contact may be the only circumstance in which (radical) simplification occurs. It is not clarified what level/kind of simplification is radical, but he strongly argues that this is what happened in the history of Bodo-Garo languages (p. 78) – which shows a complete loss of verbal indexing paradigms (and a shift to agglutinating morphology). I coded L2\_status as a binary variable based on information in grammars (both from the speech community and others), ethnographies, and articles on multilingualism in sub-areas inside the EHR.

Considering DeLancey’s claim that ‘large-scale contact’ is what leads to simplification (2014:76), I assumed if there is no mention of another language being spoken by another community in any of the sources I consulted, it is not widespread enough to deserve mention, and hence not the kind of intense contact we think leads to simplification. Languages spoken as an L2 were scored as Valley, while languages that were not spoken as an L2 were scored as Hill. Due to lack of data, this variable does not capture what proportion of the population may be L2 speakers, and thus obscures over whether a language may be spoken as a lingua franca or in a situation of stable bilingualism which may have important differences (Kauhanen et al., 2023). However, future studies can explore the differing impacts of these types of multilingualism as more data becomes available from this region.

Scott (2009) gives paramount importance to agriculture in the formation of Hill and Valley societies. He writes: “The permanent association of the state and sedentary agriculture is at the center of this story” (p. 9). In the valleys, people can practice wet-rice agriculture, which both supports a large population, and binds them to an area, crucial for state-making (p. 41). On the other hand, shifting agriculture, which involves dispersed populations, mixed cropping, and less labor (hence binding a smaller population to its upkeep), is the anathema to all state-makers (p. 77). For comparability, I first coded agricultural intensity referencing the Ethnographic Atlas (Murdock et al., 1999) feature 28 (Agriculture: Intensity). However, it soon became clear that in the EHR, only three of the five codes from the Ethnographic Atlas variable can be observed: horticulturalists, shifting agriculturalists, and intensive agriculturalists. Furthermore, some groups seem to practice both shifting and intensive (wet-rice) agriculture (such as Tiwa, Doloi et al. 2024:4). So, I revised this variable to have three values: non-intensive (including shifting agriculture and horticulture), mixed (when they practice both), and intensive (wet-rice agriculture). I used information from grammars, ethnographies, research papers on agricultural practices, and where available, I used the existing code from the Ethnographic Atlas feature.

Table 2: Mapping of Ethnographic Atlas feature 28 (Agriculture: Intensity) codes to agricultural\_intensity variable and associated Hill/Valley scoring in this study.

|  |  |  |
| --- | --- | --- |
| EA028 (Agriculture: Intensity) | Revised coding | Hill/Valley scoring |
| Horticulture | Non-intensive | Hill |
| Shifting agriculture | Non-intensive | Hill |
| Intensive | Intensive | Valley |
| Intensive irrigated | Intensive | Valley |
| None | Mixed | None |

For Scott (2009), the term Valley society is synonymous with state: and Hill societies are definitionally stateless – or rather, deliberately avoiding statehood, and so, they are ‘nonstate’ (p. 9). Scott considers that there have been state-projects in the hills, but few have succeeded, and is satisfied to consider them exceptions (p. 19). Scott relates a scale of ‘social units’ with state-incorporation, or their resistance – pointing out the most resistant kind of social structure, prevalent in the hill are acephalous (‘headless’) small aggregates of households (p. 208). On the other hand, Valley states consisted of agglomerations of petty chiefdoms, villages, and hamlets (p. 36). Similarly, I coded this variable with reference to the Ethnographic Atlas (Murdock et al., 1999) feature 33 (jurisdictional hierarchy beyond the local community) and adapted it to fit the binary Hill/Valley distinction. The resulting variable is coded “1” for states, and “0” for acephalous or small collections of villages.

Table 3: Mapping of Ethnographic Atlas feature 33 (jurisdictional hierarchy beyond local community) codes to political\_organization variable and associated Hill/Valley scoring in this study.

|  |  |  |
| --- | --- | --- |
| EA033 | Revised coding | Hill/Valley scoring |
| 1 (Acephalous) | 0 (Acephalous, petty chiefdoms) | Hill |
| 2 (One level, petty states/paramount chiefdoms) | 0 (Acephalous, petty chiefdoms) | Hill |
| 3 (Two levels, petty states/paramount chiefdoms) | 0 (Acephalous, petty chiefdoms) | Hill |
| 4 (Three levels, larger states) | 1 (States) | Valley |
| 5 (Four levels, larger states) | 1 (States) | Valley |

Scott holds that manpower is a key resource for state-formation, at least in Zomia (2009:33). Manpower was necessary for the extraction of grain or the defense of trade routes, depending on the revenue of the state (p. 33). In a way, the Hill and Valley divide can be characterized as one of ‘manpower-poor’ and ‘manpower-rich’ political systems, respectively (p. 64). Thus, I code ‘speaker population’ as a variable that measures the size of the speech community. I used information from academic sources primarily, resorting to census data if not available. Larger speaker populations are hypothesized to be subject to leveling pressures on complexity, leading to 'simpler' linguistic systems (Koplenig, 2019:9). Languages with a speaker population above the median were scored as Valley, and below were scored as Hill. The median was used as a more reliable measure of central tendency given some very large values of population size that skewed the average upwards.

Table 4: Summary of sociocultural variables used for Hill/Valley classification

|  |  |
| --- | --- |
| Variable name | Hill/Valley Scoring |
| L2 status | Yes = Valley, No = Hill |
| Agricultural intensity | Intensive = Valley, Non-intensive = Hill, Mixed = None |
| Political organization | 1 = Valley, 0 = Hill |
| Speaker population | Above median = Valley, Below median = Hill |

## Ecological Variables

The Hill/Valley distinction, while about societal organization, inherently and intentionally evokes geographical distinctions. Specifically, Scott points out that elevation and terrain roughness are primary determinants of how societies organize in Zomia. So much so, that certain groups that seem scattered/fragmented on a map seem cohesive if one were looking at the map through lateral slices — explaining the distribution of groups like the Akha (Scott, 2009:19). Two variables define how Hill and Valley societies come to form: elevation, and terrain roughness.

Elevation is a determining factor mainly because of the availability of agricultural options: for wet-rice agriculture, one needs to be settled in alluvial floodplains. This in turn, supports the larger population and concentration of state power that gives rise to Valley states. On the other hand, the slopes of hills favored mixed, shifting agriculture: particularly tubers such as yam and taro (Scott, 2009:99). This method, which requires a larger area of land[[6]](#footnote-6), and clearing and shifting of settlements is much more common in hills (p. 18). More importantly, it requires far less labor, and cannot be easily captured by other raiding groups, making it suitable for the ‘nonstate’ project in Hill societies (p. 199).

It is not just elevation though, which serves as a barrier to state-formation: rather, it is the hindrance of navigability that it comes with. Terrain roughness (which Scott calls "friction of terrain") is an important factor because it is hard to concentrate state power without access to relatively unencumbered transportways (p.45). This is why many of the Valley states were based around rivers (p. 50), which extends the distance over which supplies can be exchanged, and hills, swamps and mountains served as the natural limits of these states (p. 45).

I obtained data for these environmental variables from the NASA Shuttle Radar Topography Mission (Farr and Kobrick, 2000), a publicly available digital elevation model at one arc-second resolution. I collected coordinate information on the languages by selecting the largest settlement where the speaker group is dominant, then finding the coordinates for that settlement from publicly available sources (described in Appendix A), if not described in the grammar. To set the range over which we can assume the language is spoken (minimally), I created a buffer corresponding to the minimum distance between languages in my sample. This method underestimates language range sizes, but crucially does not overestimate, and is large enough to observe sharp drops in elevation. I calculated elevation as the average value in this buffer, and terrain roughness as the standard deviation of slope in this buffer, following Grohmann et al. (2011).

## Operationalizing Complexity

Commonly, complexity is measured as the number of parts or rules in a system: the absolute approach. This can be, for example, a list of the number of parts of the system: such as the number of phonemes in an inventory, but a more comprehensive approach also considers their interactions: making absolute complexity about 'the length of description' a phenomenon requires (Dahl, 2004: 21-24). As is common in cross-linguistic studies, I take an absolute approach to complexity (Miestamo, 2008).

This results in three measures of core argument coding complexity in my study: case marking, cell, and form complexity. Case marking complexity measures the number of rules that determine the expression of nouns that encode core arguments. Cell and form complexity measure complexity in the verbal form as it encodes core arguments, particularly the size of the paradigm and the number of forms in the system respectively. Measuring core argument coding in these three distinct measures is important, given claims about the loss/gain of complexity in Trans-Himalayan are specific to the verbal domain, and these different measures reflect and hence help us assess differing claims about complexity in core argument marking in Trans-Himalayan.

### Case-Marking Complexity

I measure case marking complexity as the count of distinct core case markers plus the number of optionality rules governing their expression. This is an absolute measure of complexity, but is expected to correlate to learning complexity as well, since the more distinct case markers a language has, and the more rules that dictate their insertion, the more a learner must “learn”. Bentz and Winter (2013) find that languages with more second language speakers tend to have smaller case inventories, lending credence to the idea. This is a measure restricted to morphosyntactic complexity, so it excludes allomorphy.

For example, Bodo (bodo1269) has two primary case markers for core arguments: a nominative marker (for S and A) -a (with four allomorphs) and an object marker, -khwu (Haokip and Brahma, 2018:3). It has two additional markers for P: -lo is used when P is in contrastive focus, and -nao marks weakly affected objects (homophonous with the locative marker). Bodo's case marking complexity is therefore 4 (four distinct core case markers: -a, -khwu, -lo, -nao) + 2 (two optionality rules: contrastive focus on P, weakly affected P) = 6, reflecting both the inventory of forms and the conditional rules governing their expression.

### Verbal Complexity

Verbal argument indexes usually show a lot more distinctions than case markers, as they track person and number. These complex verb forms can be understood as resulting from paradigms that track dimensions of variations, or as the combination of the verb with distinct morphemes that index core arguments. I operationalize this as cell complexity, which measures the number of cells in a paradigm where the verb is inflected for a transitive clause, and form complexity, which records the number of unique forms that can be used on the verbal unit to index S, A, and P. For both, I only considered declarative, affirmative clause types, and only considered one tense type (typically the morphologically unmarked tense, and if both were marked, the one with fewer distinctions).

For an example of why this dual approach is justified, consider Karbi (karb1241). Karbi employs a "non-subject speech act participant marker" whenever the P argument is first or second person (Konnerth, 2014:224). This means Karbi utilizes 4 out of 9 slots in its verbal paradigm (cell complexity = 4) but employs only a single distinct form (form complexity = 1). Karbi’s resultant cell complexity is 4, but its form complexity is only 1.

Table 5: Karbi ‘Non-Subject SAP marker’ (Konnerth 2014:224). In the original table, ‘P’ is represented as 'non-A'. Rows and columns depict person combinations. + indicates presence, - indicates absence.

|  |  |  |  |
| --- | --- | --- | --- |
| A \ P | 1 | 2 | 3 |
| 1 | - | + | - |
| 2 | + | - | - |
| 3 | + | + | - |

Form complexity also includes forms such as inverse markers (that distinguish a set of A/P combinations) and portmanteau morphemes (that encode a specific A/P combination). However, form complexity is exceeded by cell complexity for every language in my sample. Both measures are included, as they capture different measures of complexity that represent different obstacles to language learners — learning when to use a marker in the possible space of A/P combinations, or learning specific, unique forms that encode an A/P combination. All measures of complexity were then centred and scaled (represented as standard deviations from the mean).

# Analysis

I used mixed effects models to test relationships between the Hill/Valley classification, or its individual sociocultural predictors, and core argument coding complexity. Additionally, I also tested whether Hill/Valley classification itself is associated with its suggested ecological conditions. With N=31 languages after excluding an influential outlier (discussed in S4.2), this study is an exploratory rather than confirmatory analysis. Due to the small sample size, I included only one fixed effect at a time in mixed effects models to avoid overfitting.

Despite these constraints, the sample represents all major Trans-Himalayan subgroups present in the EHR and provides sufficient variation to assess substantial effects. I control for genealogical non-independence by using subfamily (following Glottolog classification) as a random effect in all models. Where effects approach or achieve statistical significance, I assess their robustness through sensitivity analyses, while I interpret non-significant results with consistent directional patterns as suggestive findings warranting investigation with larger samples.

All quantitative analyses were conducted in R (version 4.3.0; R Core Team, 2023). Mixed-effects models were carried out using lme4 (Bates et al., 2015), with p-values calculated using lmerTest (Kuznetsova et al., 2017). Data and R code for all analyses are available at https://github.com/patrickdas/ehr-argument-coding or by request from the author. Spatial analyses (for calculating ecological variables, described in S3.3) and maps were created in QGIS 3.28.1 (QGIS Development Team, 2022).

Quantitative analysis (S4.2, Table 6) shows that the absolute complexity of argument coding in Trans-Himalayan languages of the EHR is not explained by Hill or Valley classification, in either domain (case-marking or verbal agreement marking). In fact, Hill/Valley classification itself is not predicted by the ecological conditions said to give rise to it, questioning the causal claims made in this framework (S4.1). Instead, the individual sociocultural traits linked to Hill/Valley societies show the most consistent effect of all predictors explored in this study, with one factor (whether a language is spoken as an L2) being significant (at p < 0.05) (S4.3). Furthermore, the languages of this region show interesting spatial patterns and semantic features, that I explore in S4.4.

## Ecological variables do not predict Hill/Valley cultural type

I find that neither ecological variable: elevation or terrain roughness; in isolation or combination; could accurately predict Hill or Valley categorization (regardless of whether Split was included in Hill or Valley, see Appendix B.3.1). This may be due to the small sample size, or that there is some better way to characterize the ecological distinction between Hill and Valley societies. However, at the very least, Hill and Valley cultural traits may not be that closely linked to ecological variables, at least those identified by Scott (2009). As Fig. 4 below shows, Hill, Split and Valley languages occur at many different elevations and levels of terrain roughness. This is likely due to the presence of high-altitude Valley like societies, like Limbu and Tibetan.

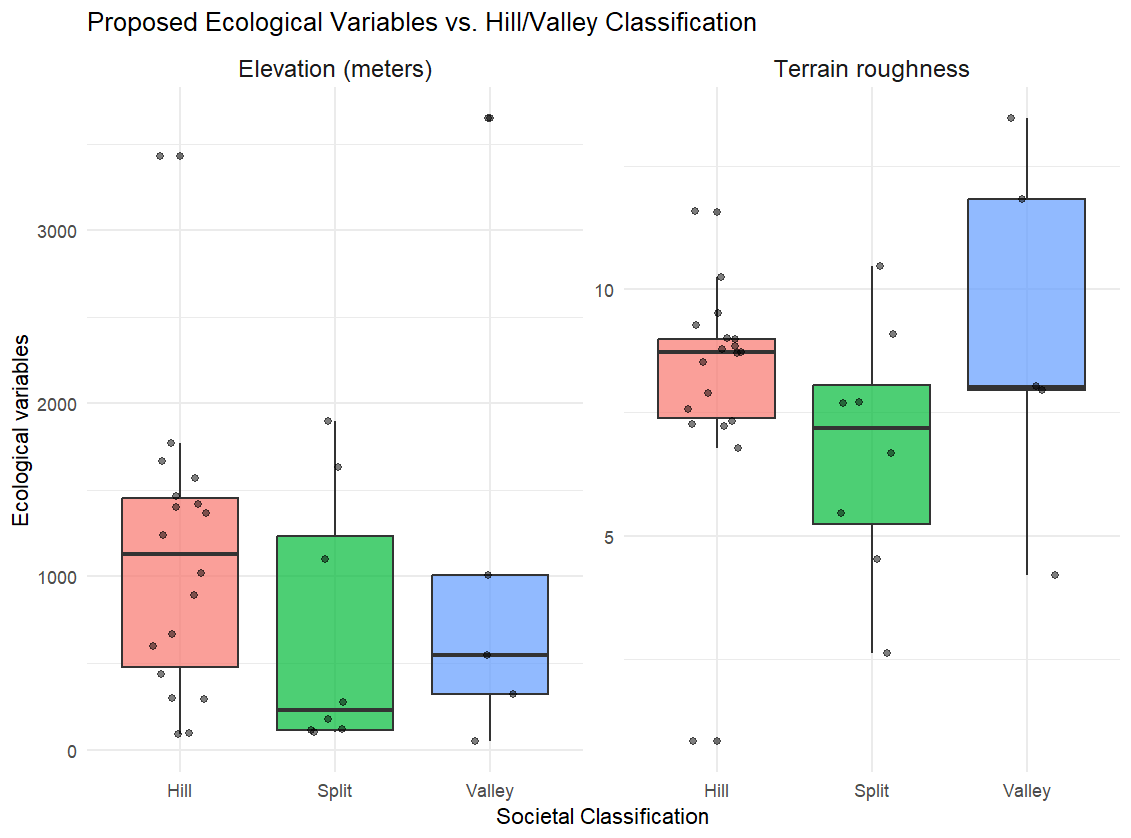


Figure 2: Ecological variables (Elevation in meters above sea level, terrain roughness in standard deviations) versus Hill, Spit and Valley categorization

Studies that link the physical environment to linguistic complexity, particularly altitude, point out that the terrain is not itself the causal factor (Nichols, 2013:38), but indirect, being mediated by its sociolinguistic consequences, such as isolation (Urban and Moran, 2021:6-7). The high terrain roughness for Valley groups is surprising – this may be an artifact of the scale of the digital elevation model, or it may be the case that Valley cultural traits are found even in Hill environments in the EHR. Either way, there is not a clear distinction between languages that are socioculturally Hill/Valley according to the topography, as suggested in the literature – atleast for the Trans-Himalayan languages in the EHR.

## Sociocultural ‘Hill’ and ‘Valley’ do not effectively explain complexity

Furthermore, the classification of a speech community as Hill or Valley according to its sociocultural traits does not in turn predict the complexity of the argument coding strategy its language uses. Visual inspection of the complexity scores in a box plot (in Fig. 5 below) makes it apparent that our results may be heavily skewed due to outliers.

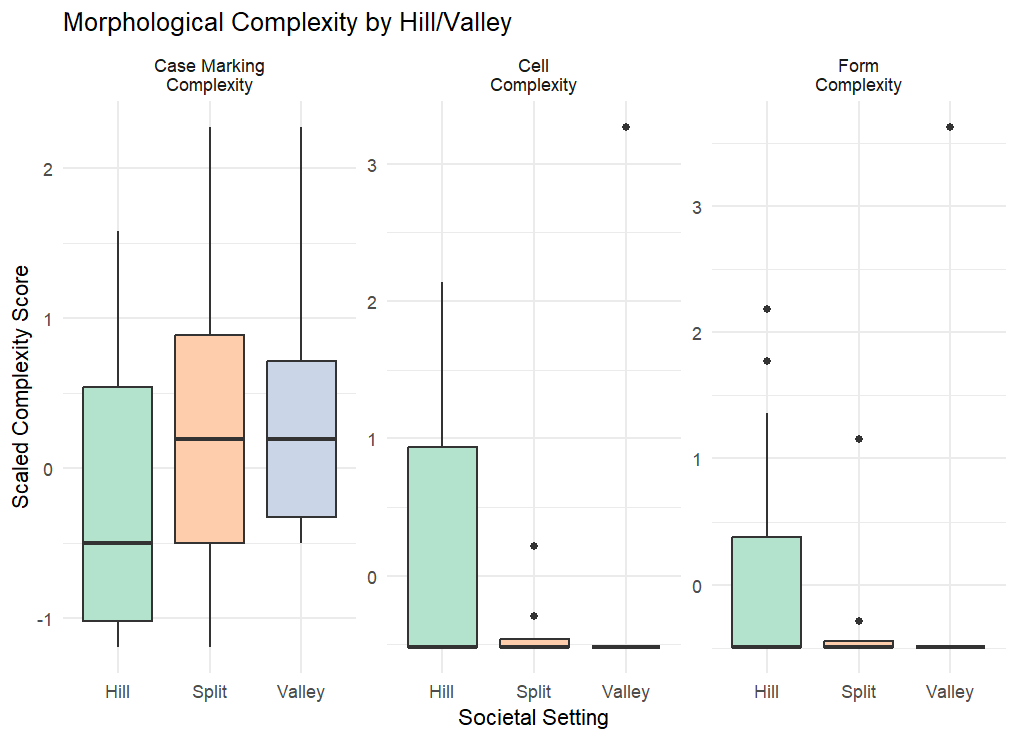


Figure 3: Box-plot of scaled complexity scores (case marking, cell, and form) versus societal setting (Hill, Split or Valley). Colored boxes indicate interquartile range (middle 50% of the data), and black bars indicate the median.

However, this also allows us to notice some patterns: languages with simple verbal indexation systems are just as likely to be Hill (culturally), as they are to be Valley. On the other hand, verbally complex languages tend to be Hill, rather than Valley. This provides support for a unidirectional version of the Hill/Valley framework: where Valley languages differ from Hill languages, they tend to use simpler indexation. When it comes to case marking, the effect is reversed. Hill languages tend to have simpler case marking systems, as their median case marking complexity lies outside the interquartile range of Valley languages’ case marking complexity (Fig 5, above).

We can also see some outliers for Split and Valley languages when it comes to verbal complexity measures. To quantify their influence on a model, I calculated Cook’s distance, a diagnostic that identifies observations which disproportionately affect model estimates. The most influential observation for both cell and form complexity was Limbu, a Valley language (glottocode: limb1266) (Cook’s D for form: 1.202295, for cell: 1.02319). For case marking complexity, the most influential outlier was Bori-Karko (bori1243), which was classified as Split (Cook’s D = 0.3037258). No other outliers went past the conventional threshold of 4/n (0.125) (Bollen and Jackman, 1985). However, even after removing an influential outlier like Limbu, the data fails to meet some assumptions, such as normality (all p-values < 0.05 for Shapiro-Wilk test) and homoscedasticity.

It turns out there is good reason to exclude Limbu. First, Limbu lies outside where Scott would say the traditional ranges of Zomia fall (2009:16), so the Hill/Valley distinction – and hence the hypothesized state-expanding mechanisms – may have played out differently, not bringing to bear with them linguistic simplification. While it is still firmly within the geographical bounds of the EHR as construed by Post et al. (2022), its typology and cultural traits cast into doubt whether the her, defined in a way where it contains Limbu, is a single cultural zone. DeLancey’s (2014) paper which first applies the Hill/Valley distinction and its sociocultural traits as potential determinants of complexity focused on the ‘Brahmaputra drainage’, a much narrower zone inside the EHR. Bori-Karko however, having a much smaller value of Cook’s D and occupying a central position in the EHR, has been kept in the present sample.

The original models with Limbu were not significant and poorly fitted – with low R-squared values (Appendix B.2). After excluding Limbu, the model aligns with theoretical predictions, but with no significant predictors. The results of each model (after excluding Limbu) are given below.

Table 6: Coefficient estimates, standard error, degrees of freedom (df), t-values, p-values, and R-squared value for the mixed effects model predicting cell complexity from hill/valley classification with subfamily as a random effect (10 levels). N = 31.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Parameter | Estimate | SE | df | t | p | Conditional R2 | Marginal R2 |
| Intercept (Hill) | 0.098 | 0.207 | 12.98 | 0.47 | 0.643 | 0.276 | 0.144 |
| Split | –0.582 | 0.315 | 25.17 | –1.85 | 0.076 |
| Valley | –0.730 | 0.377 | 26.19 | –1.94 | 0.064 |

Table 7: Coefficient estimates, standard error, degrees of freedom (df), t-values, p-values, and R-squared value for the mixed effects model predicting form complexity from hill/valley classification with subfamily as a random effect (10 levels). N = 31.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Parameter | Estimate | SE | df | t | p | Conditional R2 | Marginal R2 |
| Intercept (Hill) | –0.010 | 0.197 | 15.82 | –0.05 | 0.959 | 0.100 | 0.242 |
| Split | –0.371 | 0.299 | 26.03 | –1.24 | 0.226 |
| Valley | –0.631 | 0.358 | 26.76 | –1.76 | 0.089 |

Table 8: Coefficient estimates, standard error, degrees of freedom (df), t-values, p-values, and R-squared value for the mixed effects model predicting form complexity from hill/valley classification with subfamily as a random effect (10 levels). N = 31.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parameter | | Estimate | SE | | Df | | T | P | | Conditional R2 | | Marginal R2 | | |
| Intercept (Hill) | | –0.262 | 0.243 | | 11.69 | | –1.08 | 0.302 | | 0.124 | | 0.093 | | |
| Split | | 0.542 | 0.440 | | 25.66 | | 1.29 | 0.209 | |
| Valley | | 0.759 | 0.503 | | 26.13 | | 1.52 | 0.142 | |
|  |  | | |  | |  | | |  | |  | |  |

Between the case marking and verbal (cell and form complexity) models, the difference between Conditional and Marginal R2 increases substantially. Marginal R² represents variance explained by the fixed effects alone (Hill/Valley classification), while Conditional R² includes both fixed and random effects (subfamily). The larger difference in verbal models suggests that subfamily explains more variance in verbal argument coding strategies than in nominal (case marking) strategies. Despite this, the component sociocultural traits of the Hill/Valley framework show consistent directions of effect, and one of the factors – whether a given language is spoken as an L2 by other speaker groups – is statistically significant (S4.3), robust to sampling alternatives (Appendix B.4.2).

## L2 use may impact argument coding complexity, while sociocultural traits show consistent effect

While the Hill/Valley framework does not seem to effectively predict the diversity in argument coding strategies, it does invoke sociocultural variables that are commonly thought to pattern together in relation to linguistic structure. All the sociocultural variables show consistent direction of effect with complexity measures, even though their effects are generally not significant. Notably, whether a language is used as a second language by other speaker groups is significantly associated with verbal (cell) complexity (β = -0.58, p = 0.047). The full results of this model is given below in Table 9.

Table 9: Coefficient estimates, standard error, degrees of freedom (df), t-values, p-values, and R-squared value for L2\_status as a predictor of cell complexity in a mixed effects model with subfamily as a random effect (10 levels). N = 31.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Parameter | Estimate | Std. Error | df | t | **p** | Conditional R2 | Marginal R2 |
| Intercept | 0.109 | 0.20 | 13.30 | 0.55 | 0.589 | 0.22 | 0.14 |
| L2\_status | –0.617 | 0.27 | 28.22 | –2.26 | 0.032 \* |

Given the instability of prior results in response to changes in sampled languages, this result warrants further scrutinity. To assess its robustness, I conducted a sensitivity analysis where I ran the model on modified dataset, where one of the remaining languages was removed in each iteration from the Limbu-excluded dataset (see Appendix B.4.2). The resulting set of model coefficients show that the L2 coefficient remained consistently negative across all iterations (from -0.676 to -0.504) and maintained statistical significance in 87% of these tests, with a median p-value of 0.038.

While L2 status was the only sociocultural predictor for which a significant effect on argument marking complexity was detected, the other sociocultural traits associated with Hill and Valley societies show directions of effect consistent with what the framework predicts. Population size (β = -0.16, p = 0.286), agricultural intensity (β = -0.53, p = 0.173), and political organization (β = -0.49, p = 0.260) all showed negative associations with verbal cell complexity. Form complexity followed a similar pattern, and case marking complexity, as before, shows a positive (though non-significant) effect. The same pattern of negative associations was observed for form complexity, with estimates for L2 status (β = –0.43, p = 0.107), population size (β = –0.07, p = 0.598), agricultural intensity (β = –0.51, p = 0.159), and political organization (β = –0.48, p = 0.230) all in the predicted direction. By contrast, models predicting case marking complexity returned positive (but non-significant) estimates for the same sociocultural variables, including L2 status (β = 0.56, p = 0.133), agricultural intensity (β = 0.25, p = 0.626), and political organization (β = 0.05, p = 0.930). Full details of these models can be found in Appendix B.4.

This consistent pattern, that Hill-type, esoteric sociocultural traits show negative effect on verbal, but positive effect on case marking complexity, merits further exploration than can be achieved through quantitative modelling of a relatively small sample of languages. If real, this could reflect a trade-off, where languages develop transparent case marking inventories after losing verbal indexing, or vice versa. The other explanation is that the factors that lead to the loss of verbal indexing complexity (such as second language learning) actually promote complexification of case marking systems, via borrowing.

## Further patterns and qualitative analysis

While the Hill/Valley distinction does not seem to explain argument coding complexity in the aforementioned models and is not found to correlate in this set of languages with the ecological conditions theorized to give rise to it, we do see some clear patterns in the data. I find that all the Trans-Himalayan languages in my sample have case marking and SOV word order. The former is noted as an ‘Indospheric’ trait (Genetti, 2009:03), but this trait stretches far beyond the typical reaches of Matisoff’s (1991) ‘Indosphere’ into Mainland Southeast Asia (as far as south Myanmar in Daai Chin, or western China in Drung).

Of these, more than half show optional marking of A or P argument, and a large majority show Nominative-Accusative alignment, which is centred around the Brahmaputra region (roughly, the region surrounding Tiwa). The shared form and meaning in case marking systems leads me to consider geographic convergence as a mechanism for the increase in case marking complexity – discussed further in S5.3.

A map of asia with different countries/regions

AI-generated content may be incorrect.

Figure 4: Map of alignment types in the EHR. Red diamonds indicate Nominative/Accusative alignment, green diamonds indicate Ergative/Absolutive alignment, blue diamonds indicate Split-S (Active-Stative) alignment, white diamonds indicate Tripartite alignment.

When it comes to alignment, it needs to be mentioned that several languages described in the literature as having Ergative-Absolutive alignment are coded here as Nominative-Accusative, following a strict definition of alignment (Bickel and Nichols, 2008). These languages have what is described in the literature as an agentive marker (LaPolla, 1995) : a marker that appears mainly on A arguments, but can appear on highly marked S arguments (typically when the S is highly volitional, animate, definite, or satisfies some other condition of agentivity). Sometimes, this marker appears on the S argument due to an aspect split (such as in Tibetan, DeLancey 2003:274).

Even though this marker appears mainly on A arguments, because it can appear on S arguments, this reflects an accusative alignment (where S and A are neutralized under some conditions). Since this trait has been discussed extensively in the literature, it needs no further analysis here, other than to explain why some languages may seem differently coded here than in prior descriptions.



Figure 5: Map of verbal indexing. Color indicates verbal indexing (red is presence, blue is absence), shape indicates subfamily (diamond is Kuki-Chin-Naga, triangle is Sal (Brahmaputran), square is Macro-tani, Pentagon is Bodic, and circles indicate all other subfamilies).

Verbal indexing is rarer, appearing in only 10 languages out of 32. Spatially, verbal indexing seems to appear at the periphery, while languages that lack verbal indexing form a central cluster in the EHR. Genealogically, the presence of the trait seems to appear in specific subfamilies (such as Kuki-Chin, where all members have verbal indexing), but is not entirely explained by it either – consider the Patkaian subgroup, which includes Hakhun Tangsa, one of the most verbally complex languages in the sample, while its nearest relative in the sample Khiamniungan completely lacks indexing. The development of verbal indexing in Trans-Himalayan is a matter of rich debate, but its presence/absence does not seem to be about just inheritance or exotericity, requiring follow up.

# Discussion

The prior section suggests that the Hill/Valley framework cannot be supported in the data. One possible explanation for this is that complexity in core argument coding is just not that responsive to culture and demographics, or is responding to many more factors than just the sociolinguistic ones suggested to align with Hill and Valley classification in DeLancey (2014). While status as a second language for another community was a significant predictor, others such as agricultural intensity may be interacting in more complex ways with other sociocultural factors to produce certain linguistic outcomes. In 5.1, I discuss how teasing apart L2 learning as a simplification process from other diachronic processes may better explain the variety in core argument coding complexity in these languages.

Additionally, Hill and Valley society classifications are themselves not well predicted by their actual ecological environments. In 5.2, I discuss how the linguistic structure of modern-day Hill and Valley societies in the EHR might reflect their pasts, giving us a potential insight to the history of these speech communities in the EHR.

Lastly, I consider the geographic patterns found in alignment and verbal indexing (5.3): these may reflect past periods of convergence and diffusion, which could indicate the kind of intense contact that would result in simplification. Separating genealogical from areal influences and understanding the convergence patterns in this region is important if we want to understand the sociolinguistic pasts of these speech communities, where not much literary evidence exists.

## A three process model

If the Hill/Valley distinction fails to capture the differences in argument coding complexity, and hence the diachronic processes that occur when a large language family such as Trans-Himalayan diversifies over a varied ecological environment like the Eastern Himalayas, what might better explain the diachronic processes and distributional facts of core argument coding complexity? Here, I discuss an alternative classification to the archaic/creoloid predicted typology of the Hill/Valley framework, the diachronic processes that it suggests, and how complexity trade-offs may help model these processes.

While the Hill/Valley model predicts a bipartite division of simple vs. complex languages, there are actually three diachronic processes through which one may arrive at a point in between those poles: (i) loss of complexity, (ii) retention of complexity, and (iii) gain of complexity. The question of which factors underlying the Hill/Valley distinction promote which of these processes remains unclear, but it seems that learning a language as a second language may play a significant role in the loss of complexity.

The gain of complexity (iii) is exemplified in languages like those in the Kuki-Chin subfamily, where complex verbal agreement systems have developed that are not cognate with other Trans-Himalayan languages (DeLancey, 2014). This process of complexification represents neither retention of archaic features nor simplification under contact, but rather a third trajectory that the binary Hill/Valley distinction fails to capture. This three-process model helps explain why verbal complexity appears in some languages classified as 'Hill' (like Daai Chin and Hakhun Tangsa) but not others (like Hrusso Aka or Miji). The latter may represent cases where simplification occurred despite hill settings, or where complexity never re-developed after an initial loss. Understanding what drives these as distinct processes may prove fruitful in understanding the dynamics of core argument marking systems as a whole.

Previous cross-linguistic studies have not found robust support for complexity trade-offs (Sinnemaki, 2008, Shcherbakova et al., 2023a). However, the opposing directional pattern observed in the data -- negative effects on verbal complexity and positive effects in case marking, specifically in Valley languages – aligns with a trade-off account. In particular, this contrasts with Shcherbakova et al.’s (2023a) finding of a positive correlation between nominal and verbal complexity in Trans-Himalayan languages.

Two factors can explain this difference. First, Shcherbakova et al. focus on all grammatical domains (such as gender/noun class, tense, negation), while the focus on core argument coding complexity here may detect signals specific to shifting functional load in a subsystem that is obscured in a broader measure. Secondly, the detection of these trade-offs depends on how you operationalize complexity, given the sociolinguistic context under which these changes might occur. For example, L2 learning may only lead to simplification of the verbal paradigm in languages that also start marking cases more explicitly for ease of L2 learners, increasing (absolute) case marking complexity while reducing the relative complexity of the case marking system for an adult learner (via more transparent morphology).

## Valley pasts obscure Hill presents

The Hill/Valley framework, as originated by Scott (2009) and adapted to make predictions about linguistic complexity by DeLancey (2014) is made up of two causal links. The first is that ecological setting gives rise to social structure: rugged, hilly terrains promote small communities and shifting agriculture, fertile valleys give rise to large farming states. The second is that this social structure gives rise to linguistic structure, where these small hill communities tend to be esoteric and hence retain/innovate complexity, and the valley states assimilate new populations and sustain a larger population, making them exoteric, and hence susceptible to loss of complexity. However, both causal links can be critiqued.

We find that in the EHR, Hill and Valley societies are not straightforward results of their ecological environments. This discrepancy is resolved, once one considers Wouter’s (2012) critique of Scott’s framework in its applicability to the EHR. Wouters (p. 49-50) explains that Scott's account misses the formation of secondary states in the hills. These are smaller states that exist in the shadow of powerful valley states. This has important linguistic consequences, since the presence of state formation in the Hills, even if limited, could lead to other Valley-type cultural traits. The Angamis, generally considered a hill tribe, who reside in a largely hilly terrain, regularly engaged in other state-like behaviour, such as terraced wet-rice cultivation, and levying tribute from neighbouring tribes (Hutton, 1921:11). These adaptations may enable the Angamis to support a larger population, possibly making them susceptible to the same simplifying pressures languages with larger populations may face.

In other cases, it seems the Hill societies incorporated, assimilated or were otherwise in intense contact with Valley states. The Singpho, a hill tribe, engaged in conquest and/or enslavement of local populations that formerly spoke other languages (DeLancey, 2015:64). Turung, the variety of Singpho sampled here, underwent intense contact with Tai languages, and Singpho in the EHR lost agreement particles that are characteristic of the Jingphaw family elsewhere (DeLancey, 2011). The Akas too, who are coded as a hill tribe in my sample, regularly demanded resources, including manpower, from the Ahom Valley state (Wouters, 2012:49). These situations would lead to an influx of adult L2 learners, which would promote simplification, the causal pathway I have found support for. This is what we find in Aka and Turung, as they show case marking profiles and a lack of verbal indexing, which are typically associated with Valley languages.

Another way we can try to understand why this classification isn’t supported by analysis of this dataset is by understanding which languages would be 'miscategorized' if we were to try to predict if they were Hill or Valley based on their argument coding complexity. In Figure 8 below, I visualize which languages are mis-classified based on whether their mean cell complexity was above the mean for Valley languages, or below it for Hill languages.

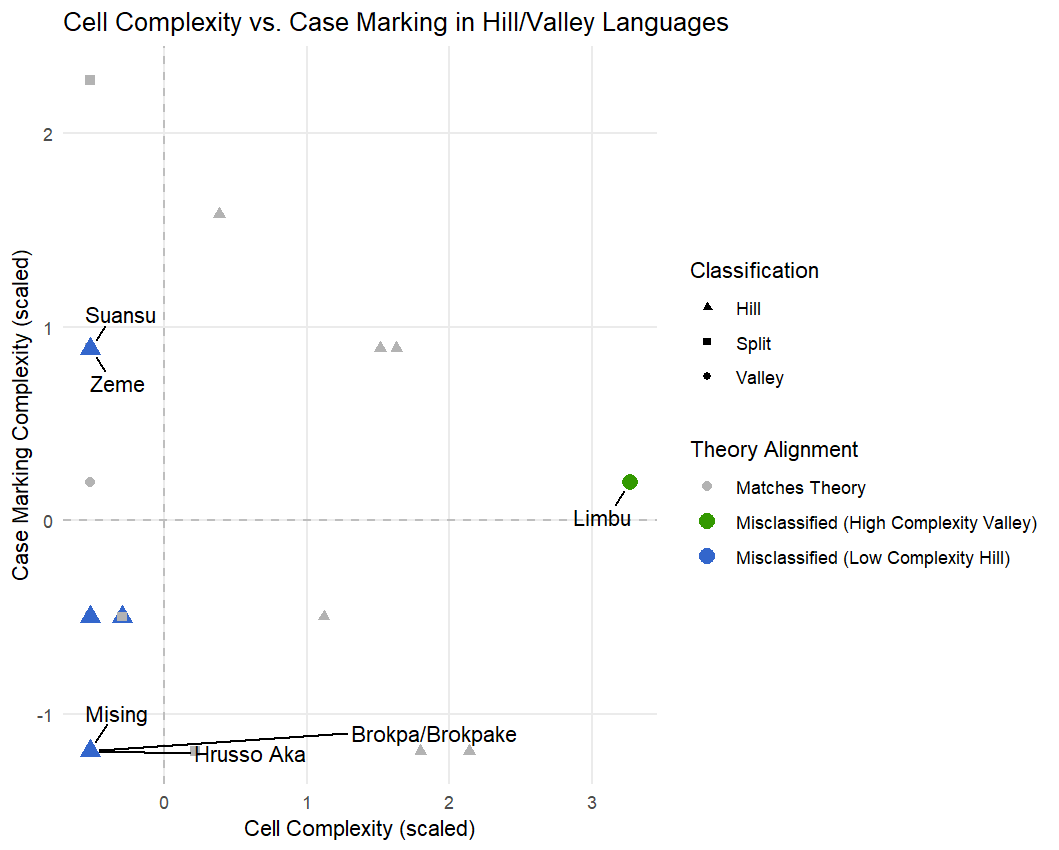


Figure 6: Scatter plot of case marking complexity vs. cell complexity. Shape indicates classification (Hill = Triangle, Square = Split, Circle = Valley), color indicates prediction (green = high complexity Valley, blue = low complexity Hill, gray = correctly classified).

In the figure above, two things are clear: there are more ‘misclassified’ Hill languages, and there is only one misclassified Valley language (Limbu). This means perhaps the Hill/Valley classification is partly true, given the theoretically justified exclusion of Limbu. If Limbu is the only ‘misclassification’ for Valley languages, it means the generalization that Valley languages have simpler structures is largely true.

However, the opposite claim – that Hill languages retain/innovate complexity – is less robust. I propose this is due to two reasons. First, Hill languages may engage in 'secondary state' activities, which leads to an influx of adult L2 learners, reducing their complexity. Secondly, there may be a time-depth problem. Hill languages today may be descended from Valley languages of the past, which means they lost complexity at an earlier stage, and never regained it. While sociocultural data was coded with reference to the earliest possible time-date, these languages might have lost complexity at a pre-historical stage for which we lack data. This may be true for Proto-Bodo-Garo, which is suggested to be the lingua franca of a former valley state (DeLancey, 2012). However, historical accounts place descendant groups like Bodo solidly as ‘Hill’ cultures, with acephalous political structure and shifting agriculture, that had to trade with Valley cultures for resources like rice (Saikia, 2012). Their language then, reflects a 'Valley' creoloid structure that has simplified, but did not subsequently innovate complexity. This further motivates examination of the distinct processes that lead to the loss, retention, and gain of complexity, and which factors promote them.

## Exploring geographic patterns in alignment

Another finding, which should be interpreted with caution given the lack of certainty in Trans-Himalayan subgrouping, is that subfamily[[7]](#footnote-7) explains more variation in cell complexity compared to case marking complexity. This fits with my observation that verbal indexing is largely predictable by genealogy in the Trans-Himalayan languages of the EHR: it only occurs in certain subfamilies, such as Kuki-Chin, or Tangsa-Nocte. If verbal indexing is a-priori more complex than nominal case-marking, it may be the case that more complex systems are harder to borrow than simpler ones. On the other hand, the differences in case marking complexity, on the other hand, show that despite case marking being present in every language of the EHR, the languages differ greatly within subfamilies regarding the extent of what and how they mark.

This is reflected when we map languages not just by alignment, but also whether A/P marking is optional. The map below compared to Fig. 6 shows a less coherent core/periphery pattern. Rabha, Tiwa and Hills Karbi share the optionality pattern of only differentially marking P, despite Karbi belonging to a different subgroup. This pattern extends northeast into the Tani languages (Milang, Galo) in Arunachal Pradesh.

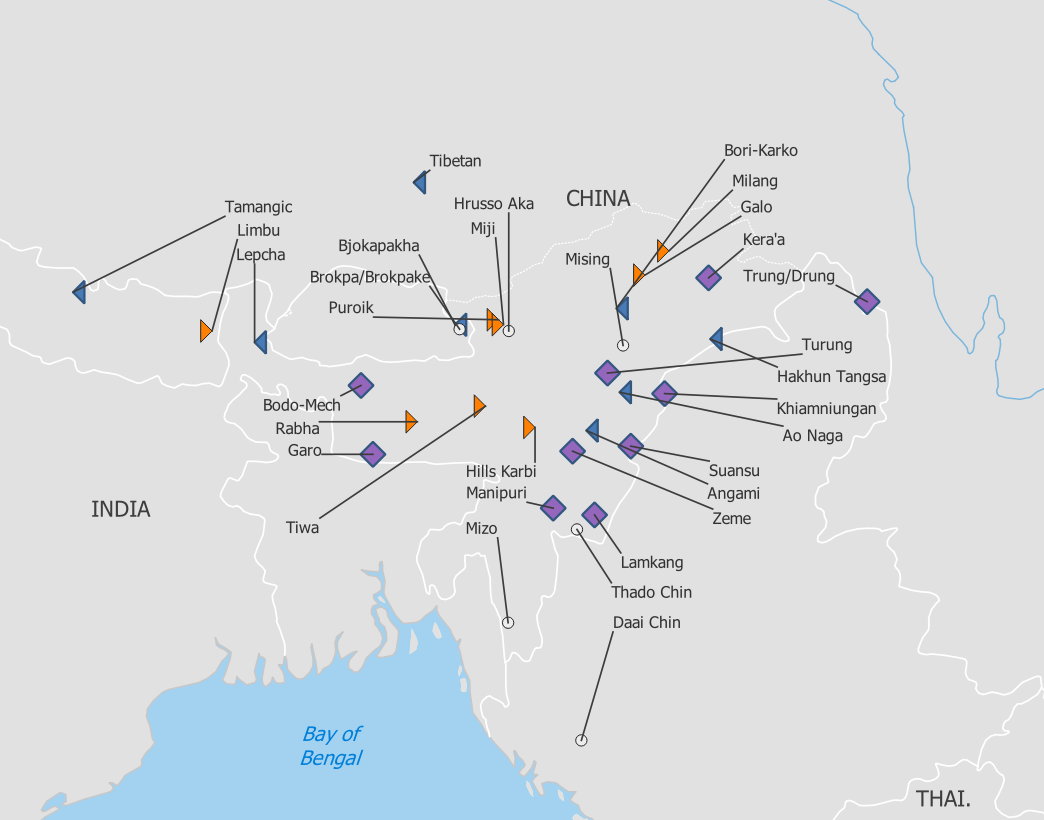


Figure 7: Map of optional case marking. Left-facing blue triangles indicate Optional A marking. Right facing orange triangles indicate optional P marking. Purple diamonds indicate that both A and P markers are optional. White circles indicate non-optional marking.

Elsewhere, from Manipuri via the Naga languages to Drung, both markers can be optional. This is reflected in a few different alignment types: Manipuri is Nominative/Accusative, Drung is Active-Stative. In fact, Manipuri and the ‘Naga’ languages share an optional A marker that has been called “agentive” elsewhere in the literature (LaPolla 1994, 1995). This marker shares form (-\*nə) and meaning across languages that are not usually reconstructed to the same subfamily (and absent in close relatives), prompting us to consider the contact events that may have led to the spread of this marker.

Earlier, I noted similarity in alignment types between neighbouring languages (S4.4, Figure 6). This similarity is perhaps explainable by looking at the diachronic changes underway in languages like Turung. Turung, despite being a variety that has undergone intense contact, has retained two distinct case-markers, both of which are optional. One is an agentive marker that can appear on S and A arguments, and the other is an “anti-agentive” that can appear on S and P arguments (Morey, 2010:334-367). The conditions for appearing are complex, but the agentive is reported to be on the verge of disappearing in modern day Turung (Morey, 2010:336). Turung is hence coded as a Split-S language.

From this Split-S system, one can see easily that the loss of either marker (either the agentive or the anti-agentive) could give rise to Nominative/Accusative alignment (if the anti-agentive is lost) or Ergative-Absolutive alignment (if the agentive is lost). This is also visible in some Tripartite systems in this sample, like Khiamniungan and Lamkang, where distinct A and P markers are present, but if they were to be lost, would result in the more common nominative-accusative and ergative-absolutive profiles here.

We see more cohesive patterns of optional case marking and nominative-accusative alignment in the center than one would expect if case marking systems were a result of random chance. Here, I suggest that the past status of Proto-Bodo-Garo as a lingua franca may have resulted in a levelling effect and spread this alignment more than would have happened via random chance. This process is not suggested for every language in the sample, but might explain a large proportion of the alignment systems.

Furthermore, the languages of this region show far-reaching similarities in conditions for differential case marking. Miji and Puroik show differential object marking based on specificity (Weedall, 2021:403, Lieberherr, 2017:208) while belonging to different subgroups. Rabha, Garo and Tiwa (all Bodo-Garo languages) and Limbu (Kiranti) use the presence of the accusative marker for indicating definiteness (Joseph, 358; Burling, 2003:188; Dawson, 2020:28; van Driem, 1987:34).

Tiwa also uses it to mark animacy, as does Turung (Morey, 2010), Manipuri (Chelliah, 1977:135), Galo (Post. 2007:721), Milang (Modi, 2017:412) and Kera’a (Peck and Reinohl, p.c.). Each of these factors could be explainable as emerging from functional needs, such as managing expectation sensitivity (Haspelmath, 2025). However, this similarity in semantic factors is also compatible with Proto-Bodo-Garo expansion account. Further research needs to disentangle family-internal contact (as suggested above), external contact (via Indo-Aryan languages, as suggested by the Indospheric hypothesis), and functional factors to better illustrate the dynamics of change in Trans-Himalayan case marking inventories.

# Conclusion

There is a complex relationship between core argument coding complexity and the environment in Trans-Himalayan languages of the EHR. I find that the The Hill/Valley framework, while useful for understanding the dynamics between social groups and how their sociocultural traits may have arisen from ecological conditions, is weakly supported by its ecological foundations in the EHR. When applied to language complexity, it shows limited predictive power for explaining the complexity of core argument coding in this sample. This may be due to limits on the theory’s applicability to the EHR (Wouters, 2012), or due to the limited effect of esoteric/exoteric cultural traits (as operationalized by the Hill/Valley distinction) on language complexity (Shcherbakova et al., 2023).

However, the factors aligned with Hill/Valley societies show consistent patterns. Particularly, whether a language is used as an L2 is supported even in quantitative models as a predictor of argument marking complexity. This provides support for the hypothesis that widespread L2 learning facilitated morphological simplification in Trans-Himalayan, at least when it comes to core argument indexation on verbs. This finding aligns with theoretical predictions regarding the effect of L2 learning on complexity (Wray and Grace, 2007, Lupyan and Dale, 2010, Trudgill, 2011). On the other hand, L2 learning seems to have a positive, but non-significant relationship with case marking complexity. This needs follow-up, as larger case marking inventories in languages learnt as a second language may reflect adding transparent case marking, aiding L2 learners (Kusters, 2003:26).

Linguistic outcomes reflect language-specific histories, and argument coding complexity might not be the most sensitive element of language to sociocultural divisions like Hill and Valley. However, the divide between "Hill" and "Valley" societies may not be that neat to begin with. Investigating their past reveals regular assimilation, migration and dispersal between the two types, and state-like mechanisms and intense multilingualism are potentially more common than expected in Hill societies. This predicts simplification for Hill languages, but does not predict complex Valley languages, consistent with what we find! The question then, is why and when do Hill languages gain, retain and lose complexity – and what conditions promote these processes?

A clearer understanding of the processes of loss, retention or gain of complexity can help us to understand the dynamics of how linguistic systems change, especially if these can be paired to language-specific histories. This study shows that even in linguistically complex and under-documented regions, systematic cataloguing and analysis of linguistic features, paired with careful sociocultural and geographic analysis can illustrate how pressures such as L2 learning interact with specific domains, such as the complexity of verbal paradigms. Future research can link these pressures to specific mechanisms such as the loss, retention, or innovation of complexity, improving our understanding of language change while incorporating lesser-studied parts of the world.

# References

Andrews, Avery D. 2007. The major functions of the noun phrase. In Timothy Shopen (ed.), Language Typology and Syntactic Description, 132–223. 2nd edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511619427.003.

Bakker, Dik. 2010. Language Sampling. In Jae Jung Song (ed.), The Oxford Handbook of Linguistic Typology. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199281251.013.0007.

Basnet, Deepa, Pratikshya Kandel, Nakul Chettri, Yongping Yang, Mahendra Singh Lodhi, Naing Zaw Htun, Kabir Uddin & Eklabya Sharma. 2019. Biodiversity Research Trends and Gaps from the Confluence of Three Global Biodiversity Hotspots in the Far-Eastern Himalaya. International Journal of Ecology 2019(1). 1323419. https://doi.org/10.1155/2019/1323419.

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Gabor Grothendieck, Peter Green, John Fox, Alexander Bauer, Pavel N. Krivitsky, Emi Tanaka, Mikael Jagan, Ross D. Boylan. 2025. lme4: Linear Mixed-Effects Models using “Eigen” and S4. https://cran.r-project.org/web/packages/lme4/index.html. (12 May, 2025).

Bentz, Christian & Bodo Winter. 2013. Languages with More Second Language Learners Tend to Lose Nominal Case. Language Dynamics and Change. Brill 3(1). 1–27. https://doi.org/10.1163/22105832-13030105.

Bickel, Balthasar. 2010. Grammatical Relations Typology. In Jae Jung Song (ed.), The Oxford Handbook of Linguistic Typology. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199281251.013.0020.

Bickel, Balthasar & Johanna Nichols. 2008. Case Marking and Alignment. In Andrej L. Malchukov & Andrew Spencer (eds.), The Oxford Handbook of Case. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199206476.013.0021.

Bollen, Kenneth A. & Robert W. Jackman. 1985. Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. Sociological Methods & Research 13(4). 510–542. https://doi.org/10.1177/0049124185013004004.

Bromham, Lindell, Russell Dinnage, Hedvig Skirgård, Andrew Ritchie, Marcel Cardillo, Felicity Meakins, Simon Greenhill & Xia Hua. 2022. Global predictors of language endangerment and the future of linguistic diversity. Nature Ecology & Evolution. Nature Publishing Group 6(2). 163–173. https://doi.org/10.1038/s41559-021-01604-y.

Bromham, Lindell, Keaghan Yaxley, Oscar Wilson & Xia Hua. 2025. Macroevolutionary analysis of polysynthesis shows that language complexity is more likely to evolve in small, isolated populations. Proceedings of the National Academy of Sciences 122(24). e2504483122. https://doi.org/10.1073/pnas.2504483122.

Burling, Robbins. 2003. Language of the Modhupur Mandi (Garo), Volume 1. Michigan Publishing, University of Michigan Library. https://doi.org/10.3998/spobooks.bbv9808.0001.001.

Chelliah, Shobhana Lakshmi. 1997. A Grammar of Meithei. Berlin: Walter de Gruyter.

Chen, Sihan, David Gil, Sergey Gaponov, Jana Reifegerste, Tessa Yuditha, Tatiana Tatarinova, Ljiljana Progovac & Antonio Benítez-Burraco. 2024. Linguistic correlates of societal variation: A quantitative analysis. PLOS ONE. Public Library of Science 19(4). e0300838. https://doi.org/10.1371/journal.pone.0300838.

Dahl, Östen. 2004. The Growth and Maintenance of Linguistic Complexity. Amsterdam: John Benjamins Publishing Company 1–343.

Dawson, Virginia Ellen. 2020. Existential Quantification in Tiwa: Disjunction and Indefinites. United States -- California: University of California, Berkeley Ph.D. https://www.proquest.com/docview/2462470108/abstract/2337A142CB424FB7PQ/1. (2 September, 2024).

DeLancey, Scott. 2003. Lhasa Tibetan. In Graham Thurgood & Randy J. LaPolla (eds.), The Sino-Tibetan languages (Routledge Language Family Series), 270–288. 1. publ. in paperback. London: Routledge.

DeLancey, Scott. 2010. Towards a history of verb agreement in Tibeto-Burman. Himalayan Linguistics 9(1). https://doi.org/10.5070/H99123042.

DeLancey, Scott. 2012. On the origins of Bodo-Garo. In Gwendolyn Hyslop, Stephen Morey & Mark W. Post (eds.), North East Indian Linguistics. Volume 4. New Delhi: Foundation Books. https://doi.org/10.1017/9789382264521.

DeLancey, Scott. 2013. Creolization in the Divergence of the Tibeto-Burman Languages. In Thomas Owen-Smith & Nathan Hill (eds.), Trans-Himalayan Linguistics: Historical and Descriptive Linguistics of the Himalayan Area, 41–70. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110310832.41.

DeLancey, Scott. 2014. Sociolinguistic typology in North East India: A tale of two branches. Journal of South Asian Languages and Linguistics. De Gruyter Mouton 1(1). 59–82. https://doi.org/10.1515/jsall-2014-0004.

DeLancey, Scott. 2015. The Historical Dynamics of Morphological Complexity in Trans-Himalayan. Linguistic Discovery 13(2). https://doi.org/10.1349/PS1.1537-0852.A.463.

DeLancey, Scott. 2023. The antiquity of verb agreement in Trans-Himalayan (Sino-Tibetan). Bulletin of SOAS 86(1). 101–119. https://doi.org/10.1017/S0041977X23000162.

Doloi, Gargi. 2024. The Tiwa Community of Assam: A Historical and Cultural Study. The Review of Contemporary Scientific and Academic Studies 4(8). https://doi.org/10.55454/rcsas.4.08.2024.005.

Driem, George L. van. 2013. Trans-Himalayan. In Thomas Owen-Smith & Nathan Hill (eds.), Trans-Himalayan Linguistics, 11–40. Berlin: De Gruyter. https://doi.org/10.1515/9783110310832.11.

Driem, George van. 2011. A Grammar of Limbu. Berlin: Walter de Gruyter.

Genetti, Carol. 2009. A Grammar of Dolakha Newar. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110198812

Gil, David & Yeshayahu Shen. 2019. How Grammar Introduces Asymmetry Into Cognitive Structures: Compositional Semantics, Metaphors, and Schematological Hybrids. Frontiers in Psychology. Frontiers 10. https://doi.org/10.3389/fpsyg.2019.02275.

Grohmann, Carlos Henrique, Mike J. Smith & Claudio Riccomini. 2011. Multiscale Analysis of Topographic Surface Roughness in the Midland Valley, Scotland. IEEE Transactions on Geoscience and Remote Sensing 49(4). 1200–1213. https://doi.org/10.1109/TGRS.2010.2053546.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2024. Glottolog 5.1. Leipzig. https://doi.org/10.5281/zenodo.14006617.

Haokip, Pauthang & Daimalu Brahma. 2018. Differential Case Marking in Bodo. Journal of the Southeast Asian Linguistics Society 11(1). 1–15.

Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. De Gruyter Mouton 15(3). 535–567. https://doi.org/10.1515/LITY.2011.035.

Haspelmath, Martin. 2025. Ambiguity avoidance vs. expectation sensitivity as functional factors in grammatical change. LingBuzz. https://ling.auf.net/lingbuzz/008859. (14 April, 2025).

Hua, Xia, Simon J. Greenhill, Marcel Cardillo, Hilde Schneemann & Lindell Bromham. 2019. The ecological drivers of variation in global language diversity. Nature Communications 10(1). 2047. https://doi.org/10.1038/s41467-019-09842-2.

J H Hutton M A. 1921. The Angami Nagas. London: Macmillan and Company, Limited.

Joseph, U. V. 2007. Rabha (Brill’s Tibetan Studies Library, Languages of the Greater Himalayan Region v. 5/1). Leiden; Boston: Brill.

Kauhanen, Henri, Sarah Einhaus & George Walkden. 2023. Language structure is influenced by the proportion of non-native speakers: A reply to Koplenig (2019). Journal of Language Evolution 8(1). 90–101. https://doi.org/10.1093/jole/lzad005.

Konnerth, Linda Anna. 2014. A grammar of Karbi. United States -- Oregon: University of Oregon Ph.D. https://www.proquest.com/docview/1535090744/abstract/EB991E2644FB4E91PQ/1. (12 May, 2025).

Konnerth, Linda, Stephen Morey, Mijke Mulder, Mark W. Post & Kellen Parker van Dam. 2020. Languages and Peoples of the Eastern Himalayan Region and the North East Indian Linguistics Society: Taking stock. Himalayan Linguistics 19(2). https://doi.org/10.5070/H91151002.

Koplenig, Alexander. 2019. Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. Royal Society Open Science. Royal Society 6(2). 181274. https://doi.org/10.1098/rsos.181274.

Kusters, Christiaan Wouter. 2003. Linguistic complexity: the influence of social change on verbal inflection (LOT 77). Utrecht: LOT.

Kuznetsova, Alexandra, Per Bruun Brockhoff, Rune Haubo Bojesen Christensen & Sofie Pødenphant Jensen. 2020. lmerTest: Tests in Linear Mixed Effects Models. https://cran.r-project.org/web/packages/lmerTest/index.html. (12 May, 2025).

LaPolla, Randy J. 1994. Parallel grammaticalizations in Tibeto-Burman languages: evidence of Sapir’s “drift.” https://doi.org/10.32655/LTBA.17.1.02.

LaPolla Randy J. 1995. ‘Ergative’ Marking in Tibeto-Burman. Senri Ethnological Studies. 国立民族学博物館 41. 189–228.

Lazard, Gilbert. 2002. Transitivity Revisited as an Example of a More Strict Approach in Typological Research. De Gruyter Mouton 36(3–4). 141–190. https://doi.org/10.1515/flin.2002.36.3-4.141.

Matisoff, James A. 1991. Sino-Tibetan Linguistics: Present State and Future Prospects. Annual Review of Anthropology. Annual Reviews 20. 469–504.

Miestamo, Matti. 2008. Grammatical complexity in cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), Language Complexity: Typology, contact, change, 23–41. Amsterdam: John Benjamins Publishing Company.

Modi, Yankee. 2017. The Milang Language: Grammar and Texts. Universität Bern PhD Thesis.

Morey, Stephen. 2010. Turung: a variety of Singpho language spoken in Assam (Pacific Linguistics). Vol. 614. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Murdock, G. P., R. Textor, H. Barry, III, D. R. White, J. P. Gray, and W. T. Divale. 1999. Ethnographic Atlas. World Cultures 10:24-136

Murdock, G. P., R. Textor, H. Barry, I., D. R. White, J. P. Gray, & W. T. Divale. (2024). D-PLACE dataset derived from Murdock et al. 1999 'Ethnographic Atlas' (v3.1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13319191

Nettle, Daniel. 2012. Social scale and structural complexity in human languages. Philosophical Transactions of the Royal Society B: Biological Sciences. Royal Society 367(1597). 1829–1836. <https://doi.org/10.1098/rstb.2011.0216>.

Nichols, J. (2009). Linguistic complexity: A comprehensive definition and survey. In G. Sampson, D. Gil, & P. Trudgill (Eds.), Language complexity as an evolving variable (pp. 110–125). Oxford, UK: Oxford University Press.

Nichols, Johanna. 2013. The vertical archipelago: Adding the third dimension to linguistic geography. In Peter Auer, Martin Hilpert, Anja Stukenbrock & Benedikt Szmrecsanyi (eds.), Space in Language and Linguistics, 38–60. Berlin: De Gruyter. https://doi.org/10.1515/9783110312027.38.

Farr, T. G., and M. Kobrick, 2000, Shuttle Radar Topography Mission produces a wealth of data. Eos Trans. AGU, 81:583-583.

Post, Mark W., Stephen Morey & Toni Huber (eds.). 2022. Ethnolinguistic Prehistory of the Eastern Himalaya. Brill. https://brill.com/edcollbook/title/61878. (28 August, 2024).

Post, Mark William. 2007. A Grammar of Galo. PhD Dissertation. La Trobe University.

Post, Mark W. & Robbins Burling. 2017. The Tibeto-Burman Languages of Northeast India. In The Sino-Tibetan Languages, 213–242. 2nd edn. Second edition. | Milton Park, Abingdon, Oxon; New York, NY: Routledge. https://doi.org/10.4324/9781315399508.

QGIS Development Team, 2023. QGIS Geographic Information System 3.28 Firenze. QGIS Association. https://www.qgis.org.

R Core Team. 2023. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

Saikia, Arupjyoti. 2012. The Historical Geography of the Assam Violence. Economic and Political Weekly. Economic and Political Weekly 47(41). 15–18.

Scott, James C. 2009. The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia. New Haven: Yale University Press.

Shcherbakova, Olena, Damián E. Blasi, Volker Gast, Hedvig Skirgård, Russell D. Gray & Simon J. Greenhill. 2024. The evolutionary dynamics of how languages signal who does what to whom. Scientific Reports 14(1). 7259. https://doi.org/10.1038/s41598-024-51542-5.

Shcherbakova, Olena, Volker Gast, Damián E. Blasi, Hedvig Skirgård, Russell D. Gray & Simon J. Greenhill. 2023. A quantitative global test of the complexity trade-off hypothesis: the case of nominal and verbal grammatical marking. Linguistics Vanguard. De Gruyter Mouton 9(s1). 155–167. https://doi.org/10.1515/lingvan-2021-0011.

Shcherbakova, Olena, Susanne Maria Michaelis, Hannah J. Haynie, Sam Passmore, Volker Gast, Russell D. Gray, Simon J. Greenhill, Damián E. Blasi & Hedvig Skirgård. 2023. Societies of strangers do not speak less complex languages. Science Advances. American Association for the Advancement of Science 9(33). eadf7704. https://doi.org/10.1126/sciadv.adf7704.

Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), Language complexity: typology, contact, change (Studies in Language Companion Series 0165-7763 v. 94). Amsterdam Philadelphia: John Benjamins Pub. Co.

Trudgill, Peter. 2011. Sociolinguistic Typology: Social Determinants of Linguistic Complexity. Oxford University Press Oxford.

Urban, Matthias & Steven Moran. 2021. Altitude and the distributional typology of language structure: Ejectives and beyond. PLOS ONE. Public Library of Science 16(2). e0245522. https://doi.org/10.1371/journal.pone.0245522.

Foley, William A. & Robert D. Van Valin. 2009. Functional syntax and universal grammar (Cambridge Studies in Linguistics 38). Re-issued. Cambridge: Cambridge University Press.

Walkden, George & Anne Breitbarth. 2019. Complexity as L2-difficulty: Implications for syntactic change. Theoretical Linguistics. De Gruyter Mouton 45(3–4). 183–209. https://doi.org/10.1515/tl-2019-0012.

Weedall, Christopher. 2021. A grammar of West Kameng Sajolang (Miji): (Upper Dzang and Khellong villages).

Wouters, Jelle JP. 2012. Keeping the Hill Tribes at Bay: A critique from India’s Northeast of James C. Scott’s paradigm of state evasion. European Bulletin of Himalayan Research 39. 41–65.

Wray, Alison & George W. Grace. 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. Lingua 117(3). 543–578. https://doi.org/10.1016/j.lingua.2005.05.005.

1. Trans-Himalayan is the language family also known as ‘Sino-Tibetan’ and ‘Tibeto-Burman’, with the glottocode sino1245. The name Trans-Himalayan is preferred and chosen here for multiple reasons, which can be surveyed in van Driem (2013). [↑](#footnote-ref-1)
2. However, the Sinosphere/Indosphere divide is a contested notion: see Post (2011). [↑](#footnote-ref-2)
3. While this ignores language-internal variation in verb types that may utilize different syntactic frames to mark core arguments differently, this approach holds the most promise for cross-linguistic comparison as we seek to understand how variation in coding of these arguments might reflect ‘more complex’ or ‘less complex’ linguistic structure. [↑](#footnote-ref-3)
4. 4 However, this claim has been contested: see Karlsson (2013) and S5.2 for a critique particular to the EHR. [↑](#footnote-ref-4)
5. Scott (2009) and DeLancey (2015) do not make a distinction between stable bilingualism and vehicular bilingualism, but these may have differing consequences for language contact and resulting linguistic complexity. [↑](#footnote-ref-5)
6. It is not clear whether Scott means a larger area of land per caloric yield, or plot size [↑](#footnote-ref-6)
7. According to Glottolog, and Trans-Himalayan subgrouping is still evolving which effects the applicability of these results. [↑](#footnote-ref-7)