

The Classification Dilemma of Banai: Indo-Aryan Core or Tibeto-Burman Influence?

Dr. Gayatri Das

University of Science and Technology
Meghalaya
gatz09@gmail.com

Abstract

Banai (also referred to as Dasgaya) is spoken in the Indian districts of Dhemaji, Goalpara, and Karbi Anglong in Assam, as well as in the border areas of the Garo Hills in Meghalaya and parts of Bangladesh. The Banai community is considered a sub-group of the Koch, who have traditionally been associated with the Tibeto-Burman (TB) language family. However, from a linguistic standpoint, Banai shows strong similarities with Hajong, an Indo-Aryan (IA) language. Reliable demographic information on the Banai population is lacking due to the absence of census data.

This study adopts a descriptive-comparative methodology, drawing primarily on field data collected from Dhemaji, Goalpara, and the South Garo Hills district. It investigates the linguistic affiliation of Banai by examining features shaped by language contact and bilingualism. While Banai continues to serve as the primary medium of intra-community communication, Assamese in Assam, and Garo and Bengali in Meghalaya, act as lingua francae, adding complexity to its linguistic profile. The lack of an indigenous script, along with the use of Assamese and Bengali scripts by speakers in different regions, further highlights the language's adaptive and contact-influenced nature. Through systematic analysis of Banai's structural characteristics and external influences, this paper argues that Banai is best classified as a primarily Indo-Aryan language with a significant Tibeto-Burman substratum.

Keywords: Banai, Assamese, Koch, Assam, Dhemaji, Goalpara, Meghalaya, language contact, Indo-Aryan, Tibeto-Burman

1. Introduction

The Banai community is distributed across the districts of Dhemaji, Goalpara, and Karbi Anglong in Assam, with smaller populations in the West Garo Hills of Meghalaya and in parts of Bangladesh. As a relatively small and under-documented group, their exact population remains uncertain due to the absence of reliable census data. Estimates provided by the North-East India Banai Welfare Association (NEIBWA) suggest approximately 12,000 Banai speakers in Assam, concentrated mainly in Goalpara, Dhemaji, and Karbi Anglong. In Meghalaya, they inhabit areas along the Bangladesh border in the West Garo Hills, including settlements such as Ganga Nagar, Pataghat, and Borringku. The community's literacy rate is notably low, ranging between 14% and 18%.

In his *Linguistic Survey of India* (1903), Grierson identified a small group referred to as Koch or Pani Koch residing in regions such as Dacca, the Garo Hills, and Goalpara. He further classified the Koches of the Garo Hills into six divisions: Harigaya, Satparia, Dasgaya (or Banai), Chapra, Wanang, and Tintekiya, concluding that their speech represented a blend of Garo with Bengali or Assamese. Later, Allen (1960) recognized three divisions of the Koches in the Garo Hills—Dasgaya, Harigaya, and Anang—while Koch (1984) advocated for their inclusion in the Scheduled Tribe list of Meghalaya. Banai (1998) documented the community's clan structure, noting the existence of forty-five clans known as *Nikini*. Bareh (2001) pointed out the absence of clear historical records regarding the origin and migration of the Banai people.

Kondakov (2013), in his dialect survey, described Banai as a small, distinct group primarily located in a single village in the South Garo Hills, alongside other communities such as the Koch and Hajong in the wider region. Although their lifestyle and language closely resemble those of the Hajong, the Banai tend to identify themselves as Koch, often referring to themselves as Banai Koch—a classification also noted by Majumder (1984), as cited in Kondakov (2013).

Banai is used as the primary language for intra-community communication, while Assamese or Hindi is employed for interaction with outsiders. The Banais are predominantly an agrarian society with a distinct cultural identity. Linguistically, the language exhibits feature characteristic of Indo-Aryan languages such as Assamese and Bangla, alongside elements associated with Tibeto-Burman languages. Assamese script is used in Assam, whereas Bangla script is employed in Meghalaya, reflecting regional variation and external influence.

2. Methodology and Criteria for Linguistic Classification

The classification of a language within a genetic family requires distinguishing inherited features — those traceable to a common proto-language — from borrowed or contact-induced features. This study applies three complementary criteria. First, lexical comparison is used to identify potential cognates with Proto-Tibeto-Burman (PTB) reconstructions (sourced from the Sino-Tibetan Etymological Dictionary and Thesaurus, STEDT) and with Indo-Aryan reflexes in Assamese, Bengali, and Hajong. Cognates are identified on the basis of regular phonological correspondences and semantic similarity, not mere surface resemblance. Second, morphological analysis examines case markers and inflectional affixes to determine whether they are inherited or borrowed, with particular attention to directionality — that is, whether borrowing flows from Banai into contact languages or from those languages into Banai. Third, typological profiling compares syntactic features such as word order, agreement, and classifier placement across Indo-Aryan (Assamese) and Tibeto-Burman (Garo).

A critical methodological distinction is maintained throughout: typological similarity alone does not establish genetic relatedness, as languages in prolonged contact may converge typologically without sharing common ancestry (Masica 1991; Campbell 2004). Accordingly, the classification of Banai rests primarily on the weight of lexical and morphological evidence for shared inherited features, with typological comparisons serving as supporting rather than determining criteria. Borrowed features are identified when phonological or morphological integration patterns suggest secondary adoption rather than inheritance, and when the sociolinguistic context (e.g., sustained contact with Assamese and Bengali communities) independently motivates the transfer.

3. Affinity of Banai with other Languages

Languages are classified into families based on shared features such as phonology, vocabulary, and grammar, all of which point to descent from a common ancestral proto-language. Typological classification focuses on structural similarities across languages, whereas genetic classification groups languages according to historical relatedness (Campbell 2004). Within these families, subgrouping identifies closer relationships among sister languages and maps their internal structures.

Evidence from the data collected by the researcher indicates that Banai has undergone significant lexical, morphological, and syntactic borrowing from Assamese and Bengali, and to a lesser extent from Hajong. This extensive borrowing is attributable to several sociolinguistic factors: (1) Banai is spoken across both India and Bangladesh; (2) migration of Banai speakers to Assam and Meghalaya following the East Pakistan riots of 1964 led to sustained contact with Koch, Garo, Hajong, Assamese, and Bengali communities; and (3) the language exists in an environment dominated by two Indo-Aryan lingua francas — Assamese and Bengali. Table 1 presents an overview of the principal languages spoken in Assam and Meghalaya.

	Meghalaya	Assam
Official Language	English	Assamese, Bodo and Bengali
Regional/Community Languages	Khasi, Garo. Koch, Garo and Bengali in the South Garo Hills (border areas with Bangladesh)	Assamese, Bodo, Bengali, etc. In Dhemaji: Assamese. In Goalpara: Assamese, Bengali and Garo.

Table 1: Principal Languages Spoken in Assam and Meghalaya

A close reading of Table 1 reveals that Banai speakers in Meghalaya and Bangladesh maintain contact primarily with Garo and Bengali communities. In Assam, speakers in Goalpara interact with Assamese, Bengali, and Garo speakers, while those in Dhemaji are largely in contact with Assamese-speaking populations. This distribution accounts for Banai's exposure to both Indo-Aryan and Tibeto-Burman linguistic environments and is the key sociolinguistic driver of its mixed lexical profile. The following sections compare Banai with two sets of languages: Eastern Indo-Aryan languages (Assamese, Bengali, and Hajong) and Tibeto-Burman languages (Koch and Garo).

Table 2 illustrates the lexical word comparison among IA languages: Assamese, Bengali and Hajong with Banai.

Banai	Indo-Aryan			Gloss
	Hajong	Bengali	Assamese	
ujju, ma	mau	ma	ma,ai, amma	‘mother’
baba	baba	baba	deuta,bap,abba, pita	‘father’
magu	magu	bou	g ^h oini	‘wife’
b ^h atar	b ^h atar	džamai	giri,poi	‘husband’
džontu	džontu	džontu	ziwo-zontu	‘animal’
mas	mas	mach	mas	‘fish’
bisi	bisi	bici	guti	‘seed’
tsan	tsan	cād,condro	zün	‘moon’
gaŋ	nodi	nôdi	noi, nodi	‘river’
bosor	bosor	bôchôr	bosor	‘year’
moi	moi	ami	moi	‘I’
toi	apni/toi	tui(Informal) tumi(Casual), apni(Formal)	toi (Informal) , tumi (Casual), apuni(Formal)	‘you’ (Singular)
oi	oi	o	xi,i	‘s/he’
ami	ami	amra	ami	‘we’

tora	apnagilu/ tura	tora,tumra, tomra,apnara	tohôt (Casual),tumaluk (Informal),apunaluk (Hon.)	'you' (plural)
ora	ora	ora	xihôt, ihôt	'they'
iduu	uduu	eta	ei	'this'
uduu	hiduu	ota	xei	'that'
idete	iduni	ekhane	iat	'here'
udini	uduni	okhane	tat	'there'

Table 2: Lexical comparison of Banai with Indo-Aryan Languages (Assamese, Bengali and Hajong).

Table 2 reveals that the overwhelming majority of basic Banai vocabulary — including pronouns, kinship terms, and common nouns — aligns closely with Indo-Aryan forms, particularly Hajong and Assamese. The first-person singular pronoun *moi* is shared across Banai, Hajong, and Assamese; the first-person plural *ami* appears in Banai, Hajong, and Assamese; second-person singular *toi* occurs in Banai, Hajong, and Assamese. These pronominal parallels are especially significant because pronouns are among the least susceptible categories to borrowing (Campbell 2004), suggesting that the IA affiliation of Banai's pronominal system is more likely inherited than contact-induced. By contrast, Banai forms such as *gay* 'river' and *iduu/uduu* 'this/that' diverge from their IA equivalents and align with Tibeto-Burman forms (see Table 3), indicating selective retention of TB elements alongside pervasive IA influence.

Banai	Tibeto-Burman		Gloss
	Koch	Garô	
ujuu,ma	amai	ma·gipa	'mother'
g ^h ormun/bhatar	mia	segipa	'husband'
magu	mitʃik	jikgipa	'wife'
masaŋ	kan	bitin	'meat'
haraŋ	kereŋ	greŋ	'bone'
ninjur	dimi	ki·me	'tail'
hikru,hima	tʃater	ja·dil	'root'
meg	raŋ	mikka	'rain'
kouluu	apʻ	aram	'cloud'
gaŋ	gaŋ	c ^h ibima	'river'
bak ^h ar	biblə	baŋ·a	'many'
lebra	debra	jakkasi	'left'
k ^h atak	k ^h atakʻ	kan·dika,bakkandika	'short'
hakle	sakale	on·a	'early'
ora	oroŋ	bisimaŋ	'they'
iduu	jaʔa	ia	'this'
uduu	waʔa	ua	'that'
idete	i	iano	'here'
udini	uʻ	uano	'there'

Table 3: Banai Words with Probable Tibeto-Burman Provenance

Table 3 identifies a set of Banai lexical items that lack clear Indo-Aryan etymologies and show formal similarity to Koch and Garo. Two patterns are particularly significant. First, the demonstrative pronouns *iduu/uduu* ('this'/'that') deploy the vowels /i/ and /u/ to distinguish proximal from distal reference — a strategy also employed in Koch (*jaʔa/waʔa*) and Garo (*ia/ua*). While direct borrowing cannot be confirmed, the structural parallel across adjacent languages suggests at minimum contact reinforcement of an originally TB-derived system. Second, the phoneme /u/ — the unrounded back vowel — occurs in multiple Banai forms (*ujju* 'mother', *g^hormun* 'husband', *hikruu* 'root') and has no counterpart in Assamese or Bengali. Its presence in Tibeto-Burman languages such as Garo and Bodo, and its systematic occurrence in Banai, points strongly to a TB substratum. These items are therefore treated as probable retentions from an earlier TB-affiliated stage of the language rather than as recent borrowings from contemporary Koch or Garo.

4. Banai Lexical Words and PTB Roots

Banai words from Table 3 are compared with the PTB words which are shown in Table 4. The PTB words are taken from the Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT).

Banai	PTB	Gloss
ujju, ma	*pwi(y)-n, *ma, *k/m-na, *n(y)u, *(y/?)ay, *neyꜜni(y), *ʔay	‘mother’
g ^h or mun/bhatar	*b/m-laŋ, *p ^{wa} , *wa	‘husband’
magu	*s-nam	‘wife’
masaŋ	*sya-n, *s-nya-kꜜhyak, *mey, *ka-y, *m-dwaŋ, *kan, *m(y/w)anꜜmin, *ta(m)-, *s/g-la, *s-na, *m/s-kam, *syi, *ni, *krukꜜs-gyuk, *(ŋ/k)wa, *bu(r)	‘meat’
haraŋ	*s/m/g-rus, *s-rus, *g-rus, *gr(w/y)a(ŋ/k), *g-ra-t, g-du(ŋ), *m-	‘bone’
	bwaŋ	
niŋkur	*m/b-gyu(ŋ/k), *mu(n/ŋ), *swap, *du, *r-mayꜜm-ray, *r-mey, *r-ŋa, *k-lyak, *r-mi, *r-may, *k(w)a(ŋ/k), *d/s-pəy, *(d/t)am, *m-daŋ, *(p/b)waŋ	‘tail’
hikruu, hima	*(b/p)ul	‘root’
meg	*r/s/g-wa, *m/s-raŋ, *tshyar,	‘rain’
kouluu	*s-dim, *mway	‘cloud’
gaŋ	*m-t(w)əy-nꜜm-ti-s, *laŋ, *kl(y)u(ŋ/k), *lwi(y), *rap	‘river’
bak ^h ar	*m(r/y)a	‘many’

lebra	*b(w)ay, *kyay	‘left’
k ^h atak	*phə ^B , *t(y)u(ŋ/n), *dəw, *s-ŋ(i/u)ŋ, *m(y)ap	‘short’
hakle	*m-na(k/p)	‘early’
ora	norecordsfound	‘they’
iduu	*m-dayꜛm-di, *(h)iꜛ(h)ay	‘this’
uduu	*wa-n, *m-dayꜛm-di	‘that’
idete	*(h)iꜛ(h)ay	‘here’
udini	*m-dzyaŋ	‘there’

Table 4: Banai Words Compared with PTB Reconstructions (STEDT)

Table 4 shows that several Banai forms have PTB etymological antecedents. The demonstrative forms *iduu* ‘this’ and *idete* ‘here’ may be related to PTB **(h)iꜛ(h)ay*, while *uduu* ‘that’ can be compared with **wa-n*. The term for ‘root’ (*hikru/hima*) is tentatively relatable to PTB **(b/p)ul*. However, caution is warranted: many of the PTB reconstructions for individual glosses list multiple competing roots, and no systematic sound correspondences between Banai and PTB have yet been established. The comparison is therefore indicative rather than conclusive. A rigorous proto-language comparison, including regular correspondence sets across a broader lexical sample, is needed to confirm or disconfirm genetic relatedness at the PTB level and remains a priority for future research.

4.1 Case Inflections

In New Indo-Aryan (NIA) languages, case functions at both the syntactic and morphological levels (Masica 1991: 230). Assamese marks case through the following suffixes: $\emptyset$ (nominative), *-e* (ergative), *-[a]k* (accusative and dative), *-[a]r* (genitive), *-[o]t* (locative), *-pora* (ablative), *-oloi* (allative), and *-re/-dara* (instrumental). In Tibeto-Burman languages like Garo, the case marker is the final constituent of the noun phrase, sometimes followed by a postposition; Garo case markers include $\emptyset$ /*-a* (nominative), *-ko* (accusative), *-na* (dative), *-ni* (genitive), *-ni* (ablative), *-o* (locative), and *-c^ha* (instrumental) (Burling 2011: 43, 52).

1a. *gitu-ra* *ritu-ge* *kub-ale* **Banai**

gitu-ERG ritu-ACC beat-PST

‘Gitu beat Ritu.’

1b. *biraz-ra o-ge* *sithi* *eg-ra* *likh-ile*

biraz-ra o-ge *sit^hi* *eg-ra* *lik^h-ile*

‘Biraj wrote a letter to him.’

- 2a. *durba bitu-ke mar-l-o* **Bengali**
 durba bitu-ACC beat-PST-3
 ‘Durba beat Bitu.’
- 2b. *durba bitu-ke ek-ta sit^hi lik^h-l-o*
 durba bitu-ACCone-CLF letter write-PST-3
 ‘Durba wrote a letter to Bitu.’

The ablative markers *t^hiki* and *t^hokon* are cognates with Bengali ablative marker *-t^heke*. *-t^hiki* and *-t^hokon* are used in free variation. Let us see the following Banai and Bengali examples below:

- 3a. *moi ghor-t^hiki ah-ile* **Banai**
 1SG house-ABL come-PST
 ‘I came from home.’
- 3b. *ami ghor-t^heke as-l-am* **Bengali**
 1SG house-ABL come-PST-1
 ‘I came from home.’

The locative marker *-ni* in Banai is formally identical to the genitive suffix *-ni* found in both Garo and Koch, as shown in (4). However, it functions as a locative in Banai, suggesting that this form was either retained from a TB source with a locative function or was reanalysed after contact. Its formal match with the Garo and Koch genitive *-ni* does not, by itself, demonstrate genetic inheritance; a shared TB origin remains plausible, but the functional divergence merits further investigation:

- 4a. *oi g^hor-ni* **Banai**
 3SG house-LOC
 ‘He is at home.’
- 4b. *ang-ni nok* **Garo**
 1SG-GEN house
 ‘My house.’
- 4c. *nirmal-ni bukt^hil bolan* **Koch**
 nirmal-GEN dress white
 ‘Nirmal’s dress is white.’ (Duttamajumdar2022:137).

The instrumental marker *-diu* is cognate with the Bengali instrumental marker *-die* as in

- 5a. *oi am-ra dau-diu kat-ile* **Banai**
 3SG mango-CLF knife-INST cut-PST
 ‘He cut the mango with a knife.’
- 5b. *ami c^haku-die am-ta kat-l-am* **Bengali**
 1SG knife-INST mango-CLF cut-PST-1
 ‘I cut the mango with a knife.’

Table 5 shows the case markers in Banai, Hajong, Bengali, Assamese, Koch and Garo. The possible cognates for Banai case markers are marked in bold as in Table 5.

	Banai	Indo-Aryan			Tibeto-Burman	
		Hajong	Bengali	Assamese	Koch	Garo
Ergative	-ra			-e		
Nominative	∅	∅	∅, -ra	∅		∅/-a
Accusative	-ge	-ge, -gon	-ke, [e]re -go-re	-[a]k		-ko
Dative	-ge	ge, gon	-ke, [e]re		-na	-na
Genitive	-la	-la	-[e]r -go	[o]r	-ni	-ni
Locative	-ni, -bhaj	-ni, -[o]t	-e, -te	-[o]t		-o
Allative	-bhaj, -ni	bhaj		-oloi	-waj	-chi, -ona
Ablative	-thiki/-thokon	thiki, thokon, thaki	-theke	-pora		-oni
Instrumental	diu	diu, de	die	-re, -ere, - dara		-chi

Table 5: Case markers in Banai, Hajong, Bengali, Assamese, Koch and Garo

The overall pattern in Table 5 is revealing. The majority of Banai case markers — accusative *-ge*, ablative *-^hiki/-^hokon*, instrumental *-diu*, genitive *-la*, and allative *-b^haj* — are systematically cognate with Hajong, and to a lesser degree with Bengali and Assamese. This is strong evidence that Banai's case morphology is predominantly of Indo-Aryan origin. The one clear exception is the locative *-ni*, which aligns formally with the Garo and Koch genitive. This is best interpreted as a TB substratum feature that has been retained in the locative domain. The case system as a whole, therefore, supports an IA genetic classification while simultaneously attesting to TB contact influence in an isolated morphological domain.

5. Typological Comparison of Banai with Indo-Aryan and Tibeto-Burman Languages

This section offers a typological comparison of Banai with Indo-Aryan and Tibeto-Burman languages. The framework for Indo-Aryan follows Masica (1991; 2001), and for Tibeto-Burman follows Burling (2007: 43; 2008: 173). Examples from the Indo-Aryan group are drawn from Assamese, and those from Tibeto-Burman are drawn from Garo. It must be emphasised at the outset that typological convergence resulting from prolonged contact can produce structural similarity between languages of different genealogical lineages. The patterns documented below must therefore be interpreted alongside, not instead of the lexical and morphological evidence presented in Sections 2 and 4.

5.1 Basic Word Order

SOV order is the dominant pattern in Tibeto-Burman languages, with exceptions among Bai and Karen which follow SVO (Burling 2007: 43). NIA languages are likewise verb-final. Banai also exhibits SOV order, as illustrated in (6):

- 6a. *namchi git ringa* **Garoiⁱⁱ**
namchi song sing-NEUT
 'Namchi sing(s) a song.' (Ingty 2008: 74)
- 6b. *xi gaan ga-i* **Assamese**
 3SG song sing-3
 'He sings songs.'
- 6c. *oi gaan ga-i* **Banai**
 3SG song sing-PRES
 'S/he sings songs.'

SOV order is shared by both the IA and TB compranda and by Banai. This feature therefore cannot distinguish between IA and TB affiliation in Banai's case. Since SOV is an areal feature of South and Northeast Asia more broadly (Masica 1991), its presence is expected regardless of genetic origin.

5.2 Noun Phrase

The following subsections examine modifier-head ordering within the Banai noun phrase, comparing it with Assamese (IA) and Garo (TB).

5.2.1 Noun-Demonstrative

In both TB languages (such as Garo) and IA languages (such as Assamese), demonstratives precede the nouns they modify. Banai patterns identically:

- 7a. *[ia paŋ-ko] daraŋ-ba* **Garo**
 [this tree-ACC] anyone-IND
denʔ-nabe ia-de aŋ-ni-sa
 cut-NEG.IMP this-but 1S-GEN-only
 'Don't anyone cut[this tree]. It is mine.' (Burling 2004:217)
- 7b. *I hadam*
 DEM place
 'This place.'
- 7c. *ei lora-tu* **Assamese**
 This boy-CLF
 'This boy.'

7d. *u* *gari*

Banai

DEM car

‘This car.’

Demonstrative-noun order is shared across both comparanda and thus provides no basis for distinguishing IA from TB affiliation in Banai.

5.2.2 Genitive-Possessor Construction

In both Garo (TB) and Assamese (IA), the possessor precedes the possessed. Banai follows the same pattern:

8a. *ang-ni* *a-chak*

Garo

My-GEN dog

‘My dog.’

8b. *mur* *kukur*

Assamese

1SG.GEN dog

‘My dog.’

8c. *mola* *kukul*

Banai

1SG.GEN

dog

‘My dog.’

5.2.3 Plural Marking

Plural markers are suffixed to the noun in Garo, Assamese, and Banai alike:

9a. *mande-dan-n* *i* *ha?-ba-soŋ-ba*
person-PL-GEN fields- also-village-also

Garo

baŋ?-a-miŋ

much-NEUT-PST

‘The Mandis’ property (fields and villages) were plentiful.’ (Burling 2004: 180)

9b. *manuh-bur* *gol*
man-PL went.PST

‘Men went.’

9c. *mun-gilu* *dʒa-le*
man-PL go-PST

‘Men went.’

Plural markers are *-dan* (Garo), *-bur* (Assamese), and *-gilu* (Banai). Suffix placement is shared across both groups.

5.2.4 Adjective– Noun Order

Garo permits both AN and NA ordering (Burling 2004: 133). Assamese and Banai require adjectives to precede nouns — NA order is not permitted:

- | | | | | | |
|------|--|---|------|--------------------------------------|-----------------------------|
| 10a. | <i>dalʔ-gipa</i>
NMZ | <i>achak</i>
dog | 10b. | <i>achakdalʔ-gipa</i>
dog big-NMZ | Garo big- |
| | | | | | |
| | | ‘Bigdog.’ | | | ‘Bigdog.’ (Burling2004:133) |
| | | | | | |
| 11. | <i>d^hunija</i>
beautiful | <i>suwali</i>
gi
rl ‘Beautiful girl.’ | | | Assamese |
| | | | | | |
| 12. | <i>b^hala</i>
good | <i>mun</i>
man | | | Banai |
| | | ‘Good man.’ | | | |

The fixed AN order in Banai aligns with Assamese and diverges from Garo, which is the one typological feature in the NP domain that distinguishes Banai's profile from the TB comparandum.

5.2.5 Classifiers

In Garo, the classifier and numeral follow the noun (N-Num-CLF). In Assamese and Banai, the classifier and numeral precede the noun (Num-CLF-N):

- | | | | |
|------|--------------------------------|---------------------------------|--------------------------|
| 13a. | <i>teʔrik roŋ-sa</i>
Banana | N-Num
CLF-one | Garo |
| | | ‘one banana.’ (Burling2004:248) | |
| | | | |
| 13b. | <i>e-ta</i>
one-CLF | <i>am</i>
Mango | Num-N
Assamese |
| | | ‘One mango.’ | |
| | | | |
| 13c. | <i>eg-ra</i>
one-CLF | <i>am</i>
Mango | Num-N
Banai |
| | | ‘One mango.’ | |

Banai's Num-CLF-N order aligns with Assamese (IA) and diverges from Garo (TB). This is a meaningful typological distinction that supports the IA affiliation of Banai's NP structure.

5.2.6 Subject-Verb Agreement

TB languages such as Garo lack subject-verb agreement — the verb form remains invariant across person and number (Ingty 2008: 62–63). IA languages such as Assamese mark agreement obligatorily. Banai sides with TB in lacking agreement:

- | | | | | |
|------|----------------------|------------------------|----------------------------|---|
| 14a. | <i>anga</i>
I-NOM | <i>bol</i>
tree-ACC | <i>gadoa</i>
climb-NEUT | Garo |
| | | | | |
| | | | | ‘I climb(the) tree.’ Ingty(2008: 62-63) |

14b. *na-simang bol gadoa*
 you-PL tree-ACC climb-NEUT
 ‘You (plural) climb (the) tree.’ (ibid)

14c. *Arjun hoi gadoa*
 arjun-NOM tree-ACC climb-NEUT
 ‘Arjun climbs (the) tree.’ (ibid)

It is seen in examples 14(a-c) that the verb *gadoa* ‘climb’ is the same in all these sentences. The verb form does not change irrespective of its number and person. It is only with regard to tense that the form of the verb changes.

Let us look into the Assamese sentences given in (15):

15a. *moi b^hat k^ha-l-u*
 1SG rice eat-PST-1
 ‘I ate rice.’

15b. *tohot-e b^hat k^ha-l-i*
 2PL-ERG rice eat-PST-2
 ‘You ate rice.’

15c. *gita-e b^hat k^ha-l-e*
 gita-ERG rice eat-PST-3
 ‘Gita ate rice.’

16a. *moi b^hat k^ha-le* **Banai**
 1SG rice eat-PST
 ‘I ate rice.’

16b. *tora b^hat k^ha-le*
 2PL rice eat-PST
 ‘You ate rice.’

16c. *manik-ra b^hat k^ha-le*
 manik-ERG rice eat-PST
 ‘Manik ate rice.’

Examples (16a-c) confirm that the verb *kha-le* remains invariant in Banai regardless of person or number, replicating the Garo (TB) pattern in (14a-c) and contrasting with Assamese (IA) in (15a-c). This is one of the most diagnostically significant typological features of Banai because agreement is obligatory in Assamese and Bengali but absent in TB languages of the region. The absence of agreement in Banai therefore points unambiguously toward a TB substratum influence, even if the language is genetically classified as IA on the basis of its lexicon and core case morphology.

From the above discussions, the various inflectional markers of Banai which have retained its origin are shown in Table 6.

Banai	Tibeto-Burman		Inflectional markers
	Koch	Garo	
-ra		-daŋ	Plural marker with 2 nd and 3rd person pronominals
-bun/-bun/-ibun/- ibun	-mung	-min	Perfective

Table 6: Inflectional Markers with Probable TB Origin

Table 6 presents two inflectional markers in Banai that appear to be TB-derived rather than IA-borrowed. The perfective *-bun/-bun* corresponds formally to the Garo perfective *-min* (Burling 2002: 155–158) and the Koch form *-mung/-mun*, providing evidence that the verbal aspect system retains TB morphological material even as the lexicon has shifted predominantly toward IA. These markers are not merely borrowed items but form part of Banai's core grammatical architecture, reinforcing the case for a significant TB substratum.

6. Conclusion

This paper has undertaken a preliminary analysis of the linguistic classification of Banai through comparative-historical lexical examination, morphological analysis of case and inflectional markers, and typological profiling against Indo-Aryan (Assamese, Bengali, Hajong) and Tibeto-Burman (Garo, Koch) comparanda. The central finding is that Banai is best classified as a primarily Indo-Aryan language with a substantial and systematic Tibeto-Burman substratum.

The evidence for IA affiliation is robust. Basic vocabulary items — including personal pronouns (*moi* 'I', *ami* 'we', *toi* 'you sg.') — closely parallel Hajong and Assamese forms, and personal pronouns are among the least susceptible categories to borrowing (Campbell 2004). The case morphology of Banai is predominantly cognate with IA systems: the accusative *-ge* matches Bengali *-ke* and Hajong *-ge*; the ablative *-t^hiki/-t^hokon* matches Bengali *-t^heke* and Hajong *-t^hiki/-t^hokon*; the instrumental *-diuu* matches Bengali *-die* and Hajong *-diuu*. The classifier-numeral ordering (Num-CLF-N) aligns with Assamese and distinguishes Banai from Garo. Taken together, these features across the lexicon and morphology constitute a coherent IA profile.

At the same time, several features of Banai cannot be explained by IA inheritance alone and point clearly to a TB substratum. The phoneme /u/, absent from Assamese and Bengali but present in Garo and Bodo, occurs systematically in Banai. The demonstrative system (*iduu/uduu* for 'this'/'that') parallels Koch and Garo proximal/distal distinctions. The perfective morpheme *-bun/-bun* is formally comparable to the Garo *-min* and Koch *-mung*. Most significantly, the complete absence of subject-verb agreement in Banai — a structural feature that is obligatory in Assamese and Bengali — replicates the TB pattern. These cannot plausibly all be contact-induced convergences; they suggest that the Banai people originally spoke a TB language that shifted over generations toward IA under prolonged sociolinguistic pressure, leaving identifiable TB traces in phonology, morphology, and grammar.

This interpretation has broader implications for the linguistics of Northeast India. Banai represents a case of incomplete language shift — a community that ethnically identifies with TB groups (Koch) but whose language has been restructured predominantly along IA lines. This type of contact situation, in which genetic affiliation diverges from ethnic identity and typological profile, is well attested in contact linguistics and points to the importance of integrating sociolinguistic history into genealogical classification. Northeast India, as a high-density contact zone between the IA, TB, Austro-Asiatic, and Dravidian families (Masica 2001), is particularly significant for understanding such dynamics. Banai's profile supports the hypothesis that TB-speaking communities in this region have been progressively incorporated into IA speech networks without fully abandoning their grammatical heritage.

Several limitations of this study must be acknowledged. The comparative analysis of PTB roots in Section 3 is exploratory; establishing genuine PTB-Banai correspondences would require systematic reconstruction of regular sound correspondences across a broad lexical sample, which is beyond the scope of this paper. Similarly, the borrowing status of individual items requires further investigation through corpus-based elicitation and verification with multiple speakers across the Assam and Meghalaya communities. Future research should also examine clause-level syntax (including causatives, complementation, and relative clause formation) and prosody, domains which may yield additional evidence for or against the TB substratum hypothesis.

Notwithstanding these limitations, the weight of the evidence — lexical, morphological, typological, and phonological — converges on a classification of Banai as an Indo-Aryan language bearing substantial Tibeto-Burman structural inheritance, and positions Banai as a critically important case study for contact linguistics and language classification in South and Southeast Asia particularly Northeast India.

Despite these features, a comparative analysis of Banai with both Indo-Aryan and Tibeto-Burman languages provides sufficient evidence to classify Banai primarily within the Indo-Aryan language group, as summarized in Table 7.

Feature	Banai	Indo-Aryan			Tibeto-Burman	
		Hajong	Bengali	Assamese	Koch	Garó
Word Order→	SOV	SOV	SOV	SOV	SOV	SOV
Morphological Type →	Agglutinative	Agglutinative	Agglutinative	Agglutinative	Agglutinative	Agglutinative
Tense→	PRES, PST, FUT	PRES, PST, FUT	PRES, PST, FUT	PRES, PST, FUT	PRES,PST, FUT	PRES, PST, FUT
Subject- Verb Agreement →	X	X	√	√	X	X
Pro-drop→	X	X	√	√	X	X
Case system→	NOM-ACC with split Ergativity	NOM-ACC	NOM-ACC	NOM-ACC with split Ergativity	NOM-ACC	NOM-ACC
Num-CLF-N Order	Yes	—	Yes	—	—	No (N-CLF-NUM)

Table 7: Typological Comparison of Banai with IA and TB Languages

References

- Burling, R. 1961. Digitized in 2008. *A Garo Grammar* (Deccan College Monograph Series, No. 25). Deccan College.
- Burling, R. 2004. *The Language of the Modhupur Mandi (Garo): Grammar*. Bibliophile South Asia.
- Burling, R. 2007. The Tibeto-Burman Languages of North Eastern India. In Thurgood, G & Lapolla, Randy, J., (eds.), *The Sino-Tibetan Languages*. Routledge.
- Campbell, L. 2004. *Historical linguistics: An Introduction* (2nd ed.). Edinburgh University Press.
- Datta Majumdar, S. 2022. *A Socio-Linguistic Profile of Koch: An Endangered Language of North-East India*. DVS Publishers.
- Ingti, A. R. W. 2008. *Garo morphology: A Descriptive Analysis* (Doctoral dissertation). North-Eastern Hill University.

- Masica, C. P. 1991. *The Indo-Aryan Languages*. Cambridge University Press.
- Masica, C. P. (2001). The Definition and Significance of Linguistic Areas: Methods, Pitfalls, and Possibilities with Special Reference to the Validity of South Asia as a Linguistic Area). In Bhaskara Rao, P. and Subbarao, K.V. (Eds.), *The Yearbook of South Asian Languages and Linguistics*. Sage Publications.
- Wood, D. 2011. An Initial Reconstruction of the Proto-Boro-Garo Noun Phrase. In G. Hyslop, S. Morey, & M. W. Post (Eds.), *North East Indian Linguistics* (Vol. 3, pp. xx–xx). Cambridge University Press.

ⁱ[https://socialsci.libretexts.org/Bookshelves/Geography_\(Human\)/Book%3A_Introduction_to_Human_Geography_\(Dorrell_and_Henderson\)/05%3A](https://socialsci.libretexts.org/Bookshelves/Geography_(Human)/Book%3A_Introduction_to_Human_Geography_(Dorrell_and_Henderson)/05%3A)

[The_Geography_of_Language/5.03%3A_Classification_and_distribution_of_Languages](#)

ⁱⁱ Garo Morphology: A Descriptive Analysis by Angela R. Wa'tre Ingty. 2008.