himalayan linguistics

A free refereed web journal and archive devoted to the study of the languages of the Himalayas

**Himalayan Linguistics**

# *A Corpus-based study of classifiers and measure words in Khortha*

**Netra P. Paudyal**

University of Kiel

## ABSTRACT

Areal patterns of numeral classifiers have been studied extensively across Asian languages. Emeneau (1956) was probably the first work that focused on the distribution of classifiers while defining India as "linguistic area". While absent in western Indo-Aryan languages like Hindi, classifiers are prevalent in Eastern Indo-Aryan languages like Bengali, Assamese, Maithili and so on. However, Khortha, an Eastern Indo-Aryan language, remains unstudied despite possessing several classifiers, some borrowed from neighboring Munda (Austro-Asiatic) languages due to prolonged contact. This study explores classifiers in Khortha, highlighting their functions and distributions, including their intriguing postnominal use. It describes how these classifiers attach not only to numerals but also to various other linguistic elements including, demonstratives, adjectives, genitives and (past) participles.

## KEYWORDS

classifiers, sortal classifiers, measure words, Eastern Indo-Aryan

This is a contribution from *Himalayan Linguistics, Vol. 23(1): 12–42.*
ISSN 1544-7502

*Himalayan Linguistics, Vol. 23(1).* © Himalayan Linguistics 2024
ISSN 1544-7502

# *A corpus-based study of classifiers and measure words in Khortha*

**Netra P. Paudyal**
University of Kiel

## 1 Introduction[1]

Numeral classifiers are typologically uncommon in western Indo-Aryan languages like Hindi. However, Eastern Indo-Aryan languages, including Khortha, typically feature a moderate number of classifiers. Khortha exhibits seven classifiers, namely *=ʈa, =ʈi, =ʈho, =go/goɽ*, *=muɽ*, *=hʌr* and *=gãɽa,* along with numerous "measure words", which serve essential functions in this language.[2] These classifiers function as bound morphemes or clitics, never appearing in isolation. They can be attached not only to numerals, but also to nouns, pronouns, quantifiers and adjectives, with varying modifying effects. Among these numeral classifiers, the first four classifiers do not have any semantic differences. But a closer examination reveals that *=ʈho* is predominantly used with large numbers, typically exceeding three. Notably, *=ʈi* is generally used with nouns that typically refer to female beings, small objects/animals and tiny equipment. It is worth noting that certain classifiers can attach to both numerals and nominals, while others are confined to solely to numerals. When the numeral classifiers are used postnominally, they mark specificity or definiteness of the object, as observed in examples (1a) and (1b).[3] However, such specificity is not necessarily present when classifiers are attached to numerals, as shown in example (1c).

(1) a. *kʌlʌm=ʈa* *an-o!*
pen=CLF$_1$ bring-IMP
'Bring me the pen (lying over there)!'

---

[1] I am grateful to the Khortha community and my language consultants: Ritu Ghɜ̃si (resident of Kotogaṛā, Bokāro), Shekhar Barnawāl (resident of Hiroḍih, Koḍerma), Mr. Sandeep Kumār Mahto (resident of Bagdā, Bokāro) for their assistance in data recording, transcription and translation during my field visits in Jharkhand, India. I also extend my sincere thanks to Mr. Giridhāri Goswāmi (Akāsh Khũṭi) and Dinesh Dinmani (Dinesh Kumar) for their meticulous comments and native judgments (*māẽkorwā bhākhā*) during our meetings and also via Messenger and WhatsApp. Special thanks to John Peterson for his valuable feedback and support.

[2] In this work, I refer to sortal classifiers simply as "classifiers" and distinguish them from "measure words", which are also known as "mensural classifiers" in various linguistic literatures.

[3] Please note that the data used in this study were collected from the Bokaro variety (also known as Sikhari Khortha). During my fieldwork in Jharkhand, I observed variations in the distribution and usage of Khortha classifiers across different Khortha-speaking regions. Notably, the postnominal use of the Khortha classifiers is absent in the Khortha spoken in Koderma district of Jharkhand.

b. *i* *ghar=ʈa* *sundʌr* *he.*
3SG.PROX house=CLF$_1$ beautiful COP.PRS.3SG
'This house is beautiful.'

c. *ek=ʈa* *kʌlʌm* *an-o!*
one=CLF$_1$ pen bring-IMP
'Bring me a pen (any pen)!'

In addition to a variety of numeral classifiers, Khortha also possesses a diverse range of measure terms. While they serve a similar function to numeral classifiers, measure words are primarily used with mass nouns to indicate a specific quantity of something, such as *ek ʈipik pani* 'one drop of water'. However, it is worth noting that there is limited freedom in combining measure words with nouns, as certain nouns are inherently associated to specific measure words that enable them to be counted. Interestingly, unlike a number of neighboring Indo-Aryan languages, Khortha lacks a human classifier, aligning with the notion that "languages with heavy contact lack numerous sortal classifiers".[4] Given Khortha's prolonged contact with various tribal languages, notably Austro-Asiatic languages like Santali, the absence of a human classifier in Khortha likely stems from this linguistic contact (Paudyal, Forthcoming).

The remaining structure of this paper is as follows: Section 2 describes the position of the Khortha language within the Magadhan group of Indo-Aryan languages. Section 3 discusses numeral classifiers in Khortha, examining their usage in both noun phrases and modification contexts. Section 4 explores the restrictions and preferences related to the absence of classifiers. Section 5 provides an interim summary of the findings related to numeral classifiers. Section 6 introduces additional classifiers known as the "measure words" or mensural classifiers and explores how they describe physical properties, shape, and quantity. Finally, Section 7 summarizes the key findings of the study and includes a table summarizing the various classifiers and their characteristics.

## 2 Background on the Khortha language

Khortha [K$^{h}$or$\underset{\cdot}{t}^{h}$a] belongs to the Indo-Aryan branch of Indo-European. However, the linguistic classification of this language within Indo-Aryan has been a matter of debate in a few previous works. Grierson (1903) views Khortha as a dialect of Eastern Magahi, while the Ethnologue lists it as a Maithili dialect (Eberhard et al. 2020). Verma (2007: 500) suggests it as a subdialect of a regional variety of standard Magahi. These varied classifications are puzzling at a first glance, however, it seems likely that these different classifications are due to the effects of language contact (Paudyal, Forthcoming).

There is no clear evidence on the origin of the term "Khortha". Unlike the names of other surrounding languages, the word "Khortha" is neither the name of any particular place nor the name of the ethnic group that speaks it. Despite lacking a clear etymology, the "Khortha" is favored by contemporary speakers, although it has several other alternative names mentioned in different native booklets. Among these are: Khortha, Khaṭṭahi, Khaṇtali, Khoṛta, Khoṭṭa Deswali, Golwari, Rāmgaṛhiyā, Sikhari Khortha, Sadri Khortha, Khā̃ṭi Khortha, Chikā Chiki Khortha, Khaspailiya

[4] One-Soon Her p.c., 2022.

Khortha, and Bangla Khortha (Paudyal, Forthcoming). However, most of these names, especially those occurring with specifying modifiers, are not necessarily alternative names, because they denote the particular region where it is spoken. Among the Indo-Aryan languages spoken in Jharkhand, Khortha is the largest language in terms of the number of speakers who speak it as their mother tongue. According to the latest census report (GOI 2011), Khortha has approximately eight million speakers in terms of L1 speakers. It is also used as a *lingua franca* in the central part of Jharkhand. Alongside Sadri/Nagpuri, Kụrmali and Panchpargania, Khortha forms a part of the Sadani languages in Jharkhand. These are closely related languages, and all of them are used as *lingua francas*.

During my fieldwork observations and accounts from native writers reveal that Khortha spoken near Maithili-speaking areas is referred to as *chikā chiki* Khortha. This is probably the reason that it has been analysed as a dialect of Maithili by some writers, based on the variety of Khortha spoken in that particular area. Given Khortha's extensive geographic spread amidst diverse linguistic communities, it is obvious that it can be influenced by different languages in different regions (Paudyal, Forthcoming). In addition, not only the Khortha speakers, but also the speakers of other Indo-Aryan languages spoken in Jharkhand, viz. Sadri/Nagpuri, Kụrmali and Panchpargania, view these four languages as separate languages belonging to the Sadan/Sadani group, as shown in Figure 1.

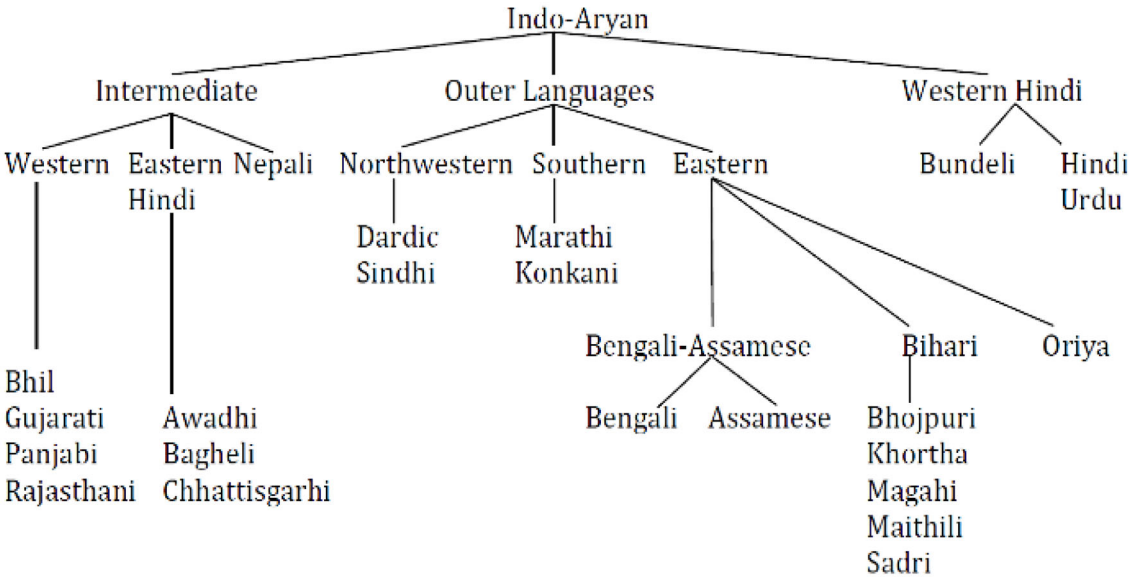


Figure 1. The place of Magadhan within Indo-Aryan,
adapted and simplified from the classification in Eberhard et al. (2019), updated by Paudyal & Peterson (2021)

Paudyal and Peterson (2021) conducted a comprehensive study using the software COG developed by the Summer Institute of Linguistics (SIL) to determine the precise position of Indo-Aryan languages spoken in Jharkhand within the broader Eastern Indo-Aryan group of languages. This computational analysis was performed using lexical items, primarily based on the Swadesh 100 wordlist with slight adaptation to include some locally used lexemes from several languages and

dialects of the Eastern Indo-Aryan languages and beyond, including Nepali, Hindi, Panjabi, Bengali and Darai (Figure 2).

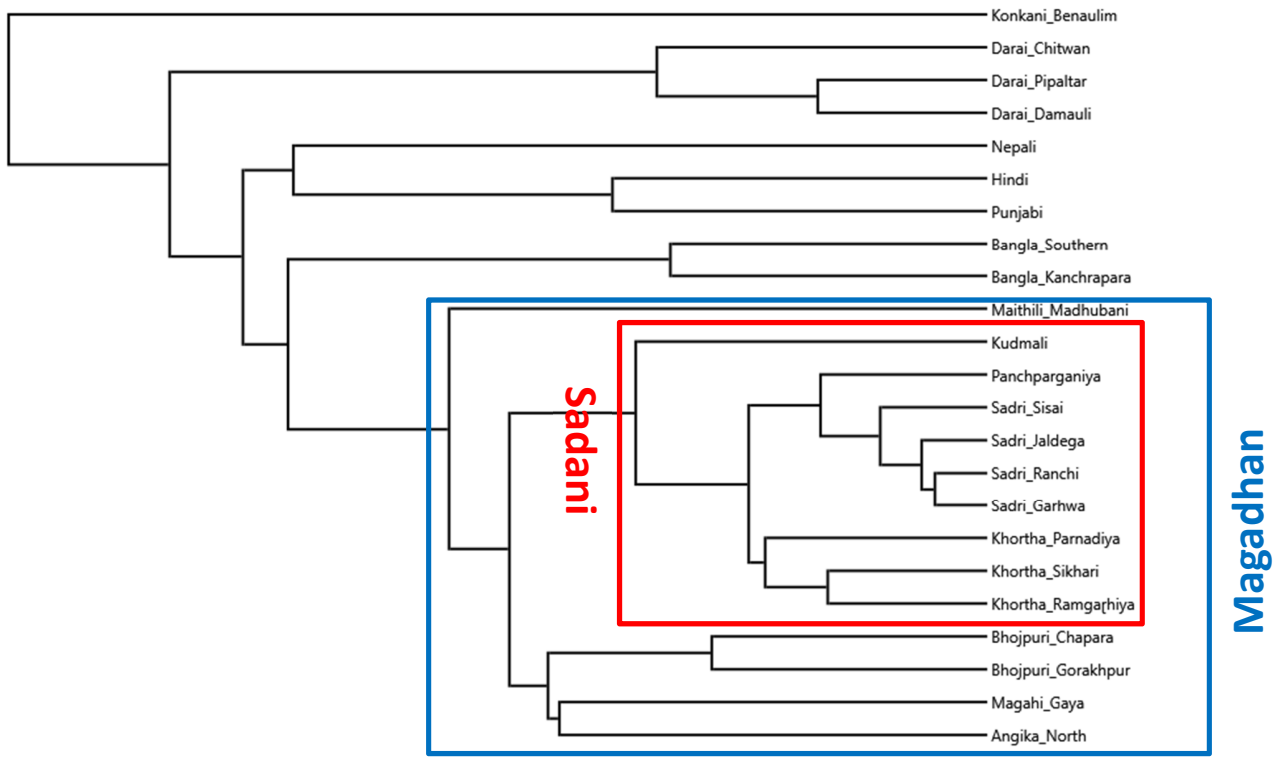


Figure 2. The genealogical relationship of various "Magadhan" languages and dialects with respect to selected Indo-Aryan varieties – Lexical similarity, UPGMA

Paudyal and Peterson's study clearly shows that Khortha, together with the other three Sadani languages (Sadri/Nagpuri, Kụrmali and Panchpargania), form a closely related group of their own within the Magadhan group of Indo-Aryan languages. Their work strongly suggests that Khortha is neither a dialect of Magahi nor of Maithili, as Verma (1985) and Ethnologue (Eberhard et al. 2020) suggest, but rather a language with its own dialects. Figure 2 above shows the genealogical relationship of various "Magadhan" languages and dialects with respect to selected Indo-Aryan varieties, cited from Paudyal & Peterson (2020).

The data presented in this article are derived based from my own fieldwork and research conducted between 2018 and 2022. The study involved four distinct field trips to Jharkhand, India, and also incorporated virtual data collection during the lockdown periods of 2021-2022. A variety of techniques were used, including elicitation sessions and recordings of natural dialogues and conversations.

## 3 Numeral Classifiers

As stated before, there are altogether seven numeral classifiers, *=ʈa, =ʈi, =ʈho, =go/goɽ*, *=muɽ*, *=hʌr* and *=gãɽa* in Khortha. These classifiers are also referred to as sortal numeral classifiers (Gil 2013:1), which are required within the context of enumeration (Aikhenvald 2000:30). Through my data elicitation and corpus analysis, it became evident that these classifiers are obligatory when enumerating in Khortha.[5] As expected, these classifiers commonly appear after cardinal numerals and the quantity word *kʌi* 'how many', but most of them can also attach to nouns, demonstratives, and adjectives. Sections 3.1 to 3.6 of this paper provide a detailed examination of each classifier, together with the relevant primary data from the Khortha corpus. Additionally, Khortha includes a substantial number of mensural classifiers, sometimes known as "measure words", that appear to describe the shape, size and quantity of the respective noun. It's interesting to note that measure words can be used with both count and mass nouns, whereas classifiers are restricted to count nouns.

### *3.1 =ʈa and =ʈi*

The classifiers *=ʈa* and *=ʈi* are the most common classifiers in Khortha. Both classifiers can accompany numerals and follow nouns, demonstratives, adjectives, genitive forms and participles. However, *=ʈa* appears mostly with non-feminine nouns, serving as a generic classifier, while *=ʈi* is primarily used with feminine nouns, as well as nouns denoting small animals, insects, objects and tiny equipment. It can also convey a sense of diminutiveness to some degree. Notably, *=ʈi* is the only classifier that exhibits masculine/feminine gender agreement with the head noun.[6]

In examples (2a) to (2d), the classifier *=ʈa/=ʈi* appears with numerals, preceding both human (*chʌuɽi* 'girl') and nonhuman (*ghar* 'house') nouns in attributive function. In (3), it accompanies the quantity word (*kʌi* 'how many') in numeral use, and in (4a) and (4b), it demonstrates noun phrase use.

(2) a. *du=ʈa* *ghar*
two=CLF$_1$ house
'two houses'

b. *ek=ʈi* *chʌuɽi/ bilair*
one=CLF$_2$.FEM girl/ cat
'a girl/a cat'

c. *okʌr* *ghar=ẽ* *cair=ʈa* *har* *hʌi.*
3SG.GEN home=LOC four=CLF$_1$ plough COP.PRS.3SG
'He has four ploughs at his home.'

d. *hamʌr* *ghar* *cair=ʈi* *chageir* *hʌi.*
1SG.GEN house four=CLF$_2$ goat.FEM COP.PRS.3SG
'There are four goats (fem) at my home.'

[5] By way of contrast, Peterson & Baraik (2023), state that such classifiers are optional but are regularly attested after numerals in Sadri/Nagpuri, one of Khortha's closely related neighboring languages.

[6] As in Khortha, the classifier *=ʈa* maintains gender agreement with the following nouns in Nepali as well. For example, *eu=ʈa keʈo* 'a boy' vs. *eu=ʈi keʈi* 'a girl'. However, unlike in Khortha, it cannot appear after the nouns in Nepali; for example **rukh=ʈa* 'the tree' or **keʈi=ʈ*i 'the girl' are ungrammatical in Nepali, but perfectly grammatical in Khortha.

(3) *tor gha=ẽ har kʌi=ʈa hʌi?*
2SG home=LOC plough how.many=CLF$_1$ COP.PRS.3SG
'How many ploughs are at your home?'

(4) a. *okʌr ghar=ẽ cair=ʈa hʌi.*
3SG.GEN home=LOC four=CLF$_1$ COP.PRS.3SG
'He has four (ploughs) at his home.'

b. *babu ki=ʈa khʌi-l-ʌi?*
boy what=CLF$_1$ eat-PST-3SG
'What (thing) did the boy eat?'

The above examples also illustrate that when a numeral precedes the noun, the classifiers *=ʈa* and *=ʈi* must appear obligatorily together with the numeral. These sentences would be ungrammatical if we drop the classifier from the numeral (5a) or if we place the classifier after the noun instead of the numeral (5b).

(5) a. **hamʌr ghar cair chageir hʌi.*
1SG.GEN house four goat.FEM COP.PRS.3SG
'There are four goats (fem) at my home.'

b. **hamʌr ghar cair chageir=ʈi hʌi.*
1SG.GEN house four goat=CLF$_2$ COP.PRS.3SG
'There are four goats (fem) at my home.'

In addition to cardinal numbers, the classifiers *=ʈa* and *=ʈi* can also accompany with ordinals and collective numbers, such as *sʌbh(e)* 'all'. However, unlike cardinal numbers, where the classifier appears obligatorily, with ordinal numbers, the classifier is optional (6b). On the other hand, when a noun is present together with a collective number, the classifier is paired with the noun rather than with the collective number (7b).

(6) a. *kʌun=ʈa le-b-ẽ?*
which=CLF$_1$ take-2SG
'Which one will you take?'

b. *pʌhil=ʈa khʌrab ho, dusʌr=(ʈa) de na!*
first=CLF$_1$ bad COP.3SG second=(CLF$_1$) give PTCL
'the first one is bad, give me the second one!'

(7) a. *sʌbhe=ʈa le jo.*
all=CLF$_1$ take v2:go
'take all (of them).'

b. *sʌbhe(*=ʈa) gidʌr=ʈa ge-l-a.*
all=CLF$_1$ child= CLF$_1$ go-PST-3PL
'All children went away.'

Moreover, the classifiers *=ʈa/=ʈi* can also directly attach to nouns to mark the definiteness or specificity of the noun phrase. *=ʈi* is commonly used after feminine nouns but also after the names of small animals, objects or as a diminutive marker. Besides, it can also appear after proper nouns. Consider the following examples from the Khortha corpus (8a-f).

(8)

a. *leru=ʈi bahir hʌi*
calf=CLF$_2$ out COP.PRS.3SG
'The baby calf is outside.'

b. *ege gidʌr=ʈi*
VOC child=CLF$_2$.FEM
'Hello girl!'

c. *sup=ʈa*
winnow=CLF$_1$
'the winnow'

d. *sup=li=ʈi*
winnow=DIM=CLF$_2$
'the small winnow'

e. *khãci=ʈa*
bucket=CLF$_1$
'the bucket'

f. *khʌĩc=li=ʈi*
bucket=DIM=CLF$_2$
'the small bucket'

In Khortha, nouns do not specify grammatical gender. When it is necessary to denote whether the noun refers to a male or female, there is a tendency to add the classifiers *=ʈa* and *=ʈi* for non-feminine and feminine, respectively. This can be observed in examples (9a-f), where the classifier *=ʈi* is added after the noun to specify a female reference (9b,d,f) and *=ʈa* to refer to a male (9a,c,e).

(9)

a. *gidʌr=ʈa*
boy=CLF$_1$
'the boy'

b. *gidʌr=ʈi*
girl=CLF$_2$.FEM
'the girl'

c. *chʌuwa=ʈa*
boy=CLF$_1$
'the boy'

d. *chʌuwa=ʈi*
girl=CLF$_2$.FEM
'the girl'

e. *bilair=ʈa*
cat=CLF$_1$
'the cat'

f. *bilair=ʈi*
cat=CLF$_2$.FEM
'the cat'

Alternatively, Khortha indicates the gender of a child by placing the borrowed nouns, especially from Hindi, before the Khortha nouns. For instance *beʈa* 'son' and *beʈi* 'daughter', borrowed from Hindi, are used as, *beʈa chʌuwa* 'boy' and *beʈi chʌuwa* 'girl' (10a). Alternatively, gender can be double marked by both a noun and the classifier *=ʈa/=ʈi*, as shown in (10b).

(10) a. *pʌhile=kʌr beʈi chʌuwa tʌni dɖ̃gʌr sʌkʌt rʌhʌ hʌ-lʌ-th.*
before=GEN daughter child little fearless strong cop AUX-PST-3PL
'The girls used to be very strong before.'
[village_life_talk]

b. *...menek beʈi chʌuwa=ʈi khub luirgʌr, cʌnphʌn.*
but daughter child=clf$_2$.fem very intelligent active
'...but the girl (was) very intelligent and active.'
[Luirgar beti chauwa, Jha 2012: 150]

In addition to common nouns, the classifier *=ʈa* also appears with abstract nouns, as observed in the Khortha corpus. An example is given in (11).

(11) *ar i jankari=ʈa rʌh-ʌe.*
and DEM.PROX information=CLF$_1$ COP.PST-3SG
'...and there was this information.'
[KʌhʌNi_0040]

When there are case markers attached to nouns and demonstratives, the case markers follow the classifiers. Adding classifiers before the case markers highlights the definiteness or specificity of the object being referred to. In (12a), the locative case *=ẽ* follows the classifier, whereas in (12b) and (12c), it is followed by the oblique case.

(12) a. *koir gaich=ʈa=ẽ.*
plum tree=CLF$_1$=LOC
'...on the plum tree'
[KʌhʌNi_0040]

b. *kukur=ʈa=ke kaʈ-i an-i=ke khuin=ʈa de-hak.*
dog=CLF$_1$=OBL cut-LNK bring-LNK=SEQ_ blood=CLF$_1$ give-IMP.3SG
'Bring the dog after cutting it and give it to her (the queen)'
[KʌhʌNi_0040]

c. *gay=ʈi=ke kʌhũ ʈawan naẽ milʌ-l.*
cow=CLF$_2$=OBL Q_ information NEG get-PST
'There was no information/clue about the cow.'

Apart from numerals and nouns, *=ʈa* and *=ʈi* can also appear with demonstratives, genitive-marked personal pronouns, adjectives and participles when they are the head of an NP. However, when demonstratives, pronouns, adjectives and participles function as modifiers of a head noun, the classifier attaches to the head noun.

With Demonstratives

The classifiers are optional when used with demonstratives functioning as noun phrases (NPs) in the sentence. However, as previously stated, the classifiers *=ʈa* and *=ʈi* are absent when the

demonstrative appears in the **attributive function**. In such cases, the demonstrative is followed by a noun, causing the classifier to attach to the head noun rather than the demonstrative.

The optionality of the classifier can be observed in examples (13) and (14), where the optional use of the classifier with a demonstrative is illustrated. The absence of classifier is in the attributive function is demonstrated in (15) and (16).

(13) a. *bʌhut bes, i=ʈa bʌhut bes.*
very good DEM=CLF$_1$ very good
'Very good! This (is) very good.'

[Sp_ShM]

b. *i=ʈa=ẽ konho ho-i jai-t=o.*
DEM.PROX=CLF$_1$=INST something be-LNK V2-FUT=ADDR
'This might cause something or do something (bad) to you.'

[brahman_mahto015]

(14) a. *i=(ʈa) hamʌr lagi.*
DEM=CLF$_1$ 1SG.GEN COP.1SG
'This one is mine.'

b. *i=(ʈa) kekʌr lag-ʌi?*
DEM.PROX=CLF$_1$ who.GEN COP-3SG

*u=(ʈa) okʌr lag-ʌi.*
DEM.DIST=CLF$_1$ 3SG.GEN COP-3SG
'Whose is this (book)?' That one belongs to him.'

(15) a. **i=ʈa kitab kekʌr lag-ʌi?*
DEM.PROX=CLF$_1$ book who.GEN COP-3SG
Intended: 'Whose is this book?'

b. **i=ʈa kitap=ʈa hamʌr lag-ʌi.*
DEM.PROX=CLF$_1$ book=CLF$_1$ 1SG.GEN COP-3SG
'This book is mine.'

c. *i kitab=ʈa kekʌr lag-ʌi.*
DEM.PROX book=CLF$_1$ who.GEN COP-3SG
'Whose is this book?'

(16) a. *i gidʌr=ʈa sundʌr he.*
3SG.PROX child=CLF$_1$ beautiful COP.PRS.3SG
'This boy is beautiful.'

b. *i poka=ʈi=k khortha bhakh=ẽ*
DEM.PROX insect=CLF$_2$=OBL khortha language=LOC

*ki kʌhʌ-l ja hʌe?*
what say-PST PASS COP.PRS.3SG
'What is this insect called in the Khortha language?'

c. *ke ʌisʌn thethʌr hʌi i goʈh=ʈa=ẽ?*
who like.this stubborn COP.PRS.3SG DEM.PROX group=CLF$_1$=LOC
'Who is stubborn like this in this group?'

[brahman_mahto015]

The examples in (13a-b) are derived from natural conversations recorded by the author, where the use of a classifier is observed. On the other hand, examples in (14a-b) are elicited examples that show the optional nature of classifiers with demonstrative as the head of an NP.

Adjectives

The usage of classifiers with adjectives follows a similar pattern to that of demonstratives. The presence of classifiers with adjectives is optional in noun phrase (NP) use, while in attributive function it is ungrammatical. In example (17b), the adjective *boɽ* 'big' is optionally followed by the classifier *=ʈa* when used as the head of an NP. However, the inclusion of the classifier in example (17c), when the adjective is used in a modifying context, results in ungrammaticality.

(17) a. *kʌun=ʈa toẽ kin-b-ẽ?*
which=CLF$_1$ 2SG buy-FUT-2SG
'Which one will you buy?

b. *boɽ=ʈa kinʌ-b /(ek=ʈa) boɽ kinʌ-b.*
big=CLF$_1$ buy-FUT.1SG /(one=CLF$_1$) big buy-FUT.1SG
'I will buy the big one/ I will buy a big one.'

c. **boɽ=ʈa ghar kinʌ-b.*
big=CLF$_1$ house buy-FUT.1SG
'I will buy the big house.'

Genitive-marked forms

Similar to demonstratives and adjectives, the presence of a classifier after the genitive form is optional in the noun phrase (NP) use. However, if there is a classifier in the previous utterance, the speaker tends to copy the classifier, as shown in (18). Conversely, in attributive function, the presence of classifiers after the genitive form is ungrammatical, as depicted in (19c).

(18) a. *kekʌr(=ʈa) ge-l=o?*
Who.GEN go-PST=ADDR
'Whose (child) went away?'

b. *tor(=ʈa)* *ge-l=o.*
2SG.GEN go-PST=ADDR
'Your (child) went away.'

(19) a. *kʌun=ʈa* *ge-l-ʌi.*
which=CLF$_1$ go-PST-3SG
'Which one went away?'

b. *tor=e=ʈa* *ge-l-ʌi.*
2SG.GEN=FOC=CLF$_1$ go-PST-3SG
'Your (child) went away.'

c. **tor=e=ʈa* *gidʌr* *ge-l-ʌi.*
2SG.GEN=FOC=CLF$_1$ child go-PST-3SG
'Your child went away.'

Participles

Similar to the previous instances, the use of classifiers with participles follows a similar pattern. In NP use, the classifiers after participles are optional, but they are considered ungrammatical in attributive function (20c). Consider the examples in (20) and (21).

(20) a. *kʌun* *bihin=ʈa* *khet=ẽ* *lʌgai-b-ẽ?*
which seed=CLF$_1$ field=LOC plant-FUT-2SG
'Which seed will you plant in your field?'

b. *tor* *dõg-ʌl=(ʈa)* *lʌgai-b.*
2SG.GEN collect-PTCP=CLF$_1$ plant-FUT.1SG
'I will plant the one which you collected.'

c. **tor* *dõg-ʌl=ʈa* *bihin* *lʌgai-b.*
2SG.GEN collect-PTCP=CLF$_1$ seed plant-FUT.1SG
'I will plant the seed which you collected.'

(21) a. *bʌhut* *pak-ʌl* *tetʌir*
very ripe-PTCP tamarind
'the tamarind which is very ripened'

b. *okʌr* *rãdh-ʌl* *kʌlwa*
3SG.GEN cook-PTCP lunch
'the lunch which s/he cooked'

c. **okʌr* *rãdh-ʌl=ʈa* *kʌlwa*
3SG.GEN cook-PTCP=CLF$_1$ lunch
'the lunch which s/he cooked'

The ungrammatical examples illustrated above highlight the restrictions on the use of classifiers in specific syntactic contexts. Instances such as (15a-b), (17c), (19c), (20c) and (21c), demonstrate that the demonstrative, adjective, genitive-marked pronoun and participle, respectively, cannot host the classifier =*ʈa* when used attributively and followed by an overt noun. Additionally, (15b) shows that the classifier cannot be repeated in both the demonstrative and the noun.

It's noteworthy that the classifiers =*ʈa* and =*ʈi* are absent in the Parnadiya variety of Khortha, while in the Ramgarh variety, their usage is sporadic. Instead, =*a* and =*wa* are used after nouns to denote specificity, as discussed in detail by Paudyal (in press) concerning the marking of definiteness and indefiniteness in noun phrases. Additionally, it should also be noted that the classifiers =*ʈa* and =*ʈi* are also attested in other neighboring languages, for example Kuṛmali (own work), Bengali (Hanne-Ruth, 2012) and Oriya (Neukom & Patnaik, 2003). However, only =*ʈa* is found in the closely related languages Sadri/Nagpuri (Peterson & Baraik, 2023) and Panchpargania (own work). Both =*ʈa* and =*ʈi* are attested in Nepali, where they function exclusively as numeral classifiers. Unlike in Khortha, they are not used postnominally. [7]

## *3.2 =go/=goɽ*

The classifier =*go/goɽ* [CLF$_3$] is used with both animate and inanimate nouns, as well as with both large and small numerals. Unlike =*ʈa*/=*ʈi*, =*go* is highly common and gender-neutral. For instance, *e=go chʌuwa* means 'a boy', and *e=go chʌuɽi* means 'a girl'. =*go* appears only after the numerals, and there are no instances of it appearing after the noun. If there is a case marker, it follows the classifier.

Examples (22) to (24) illustrate the use of the classifier =*go* with various numerals and other quantifiers:

(22) a. *okʌr* *chagʌir* *kʌi=go chʌuwa* *de-l-ʌi?*
'3SG.GEN goat how.many=CLF$_3$ child give-PST-3SG
'How many baby-goats did his goat deliver?'

b. *ham* *du=go=ke* *kin-bʌ-i.*
1SG two=CLF$_3$=OBL buy-FUT-1
'I will buy two of them.'

(23) a. *e=go bʌrʌd* *hʌ-l-ʌi* *ar* *e=go sãɽ* *hʌ-l-ʌi.*
one=CLF$_3$ ox COP-PST-3SG and one=CLF$_3$ bull COP-PST-3SG
'There was an ox and a bull.'
[Text0020-ox-and-bull-story.Sh-Pramanik]

b. *ham* *okra* *bis=go ʈaka* *de-l-i-ʌi.*
1SG 3SG.GEN twenty=CLF$_3$ money give-PST-1-3SG
'I gave him twenty rupees.'
[Text05-Khoda.admi.SH-Pramanik]

[7] See Allassonnière-Tang and Kilarski (2020) for a study of classifiers in Nepali

c. *gʌɽh=ek* *mʌidhẽ* *pʌcaso=go* *koṭhʌri* *ar*
palace=GEN inside fifty=CLF$_3$ oom and

*dalan* *rʌhʌ=hʌ-l-ʌi.*
passage remain=AUX-PST-3SG
'There were fifty rooms and passages inside the palace.'

[bhadu debi, Jha 2014: 38]

(24) *e=go=k* *nam* *rʌhʌe* *sit* *e=go=k* *basʌnt.*
one=CLF$_3$=GEN name COP.PST.3SG Sit one=CLF$_3$=GEN Basant
'(A king had two sons,) one son's name was Sit and another's name was Basant.'

[KʌhʌniI_0040]

The classifier *=go* can be combined with the interrogative pronoun *kʌi=go* 'how many', as in (22a). However, unlike the classifiers *=ɽa* and *=ɽi*, *=go* has certain restrictions in its usage. It cannot appear after nouns, demonstratives, adjectives, genitive-marked pronouns and participles. The ungrammatical examples in (25) illustrate some of these restrictions.

(25) a. **ham* *cair* *gidʌr=go=ke* *dekh-l-i.*
1SG four child=CLF$_3$=OBL see-PST-1
'I saw the four boys.'

b. **i=go* *gidʌr* *kekʌr* *lagʌi?*
DEM.PROX=CLF$_3$ child who.GEN COP.PRS.3SG
'Whose child is this?'

When two numerals occur together to express approximation, the classifier can occur either with both numerals or only with the second numeral. This is illustrated by the following examples, where (26) takes the classifier in both numerals, whereas in (27a-c), the classifier is only found on the second numeral.

(26) *e=go* *du=go* *gorkhiya* *duwair* *tʌr=ẽ* *nak*
one=CLF$_3$ two=CLF$_3$ herdsman door below=LOC nose

*ɖak-e=he.*
call-LNK=AUX.PRS.3SG
'One or two herdsmen are snoring near the door.'

[Mãjho, Sahu 2018: 1]

(27) a. *ek* *du=go* *khet* *rop-ai-l* *hʌ-l-ʌi.*
one two=CLF$_3$ field plant-PASS-PST aux-PST-3SG
'One or two fields (= some fields) had (already) been planted.'

[Dharam-Karam, Dinmani 2018: 4]

b. *jʌdi* *ek* *du=go* *chʌua puta* *bhʌ-i* *ge-l-a…*
If one two=CLF$_3$ child son be-LNK v2-PST-PL
'If he becomes (a father of) one or two children…'
[Khortha-lok-khatha-jha]

c. *e=khʌn* *roij* *sʌi* *pʌchas=go* *paʈha*
DEM.PROX=TEMP daily hundred fifty=CLF$_3$ he.goat

*kaʈʌ=hʌi*
cut=AUX.PRS.3SG
'These days he slaughters (lit. cuts) approximately 50 to one hundred goats every day.'
[Baiḍh, Giridhāri Goswāmi]

Adjectives

Unlike the classifiers *=ʈa* and *=ʈi*, *=go* is predominantly restricted to numerals. There are no occurrences of *=go* after demonstratives, genitives, or participles. However, there are instances where *=go* follows certain adjectives, particularly shape and color adjectives that are often marked with *ka* for definiteness. In such cases, it is attested both as the head of an NP and in attributive function.

(28) a. *kʌun=ʈa (*=go)* *le-b-ẽ?*
which.one=CLF$_3$ take-FUT-2SG
'Which one will you take?'

b. *bʌɽka=go* *le-b.*
big=CLF$_3$ take-FUT.1SG
'I take the big one.'

(29) a. *gãw=kʌr* *bʌgʌl=e* *bʌɽka=go* *bon* *he.*
village=GEN side=FOC big=CLF$_3$ forest COP.PRS.3SG
'There is a big forest in a side of the village.'

b. *okʌr* *ghar=ek* *dhair=ẽ* *choʈka=go* *gaich* *he.*
3SG home=GEN near=LOC small=CLF$_3$ tree COP.PRS.3SG
'A small tree is near his home.'

c. *ghar=ẽ* *cʌrʌka=go* *paʈha* *he.*
home=LOC white=CLF$_3$ he.goat COP.PRS.3SG
'There is a white goat at his home.'

In our corpus, *=go* is not attested with other adjectives, such as **khãʈ=go*, etc. A native Khortha scholar, A. K. Jha (2010), claims that *=go* and *=goɽ* are equally used in such cases. However, my consultants favor *=go* over *=goɽ*. During my survey, I noticed that speakers mostly prefer *=go* rather than *=ʈa* in Khortha. Moreover, unlike *=ʈa/=ʈi*, *=go* is attested in wider Khortha-speaking areas.

### *3.3 =ʈʰo*

The classifier *=ʈʰo* [CLF$_5$] is commonly used with large-numbers, typically more than three, but it can be associated with any numbers, including small ones like *one* or *two*. Additionally, it is not limited to numerals and can be used with demonstratives, genitive marked-personal pronouns, adjectives and participles. It is found with both human and nonhuman nouns in my data.

(30) a. *toẽ* *kʌi=ʈʰo* *temrʌs* *khai-l-ẽ?*
2SG how.many=CLF$_5$ guava eat-PST-2SG
'How many guavas did you eat?

b. *ham* *(ʌbga)* *ek=ʈʰo* *khai-l-i.*
1SG only one=CLF$_5$ eat-PST-1SG
I had (only) one'

c. *okʌr=ʈʰin* *dʌs=ʈʰo* *temrʌs* *hʌi.*
3SG=LOC ten=CLF$_5$ guava COP.PRS.3SG
'He has ten guavas.'

(31) *dʌs=ʈʰo* *lok*
ten=CLF person
'ten people'

Unlike the classifier *=go* but similar to *=ʈa* and *=ʈi, =ʈʰo* can also optionally appear with demonstratives, genitives, adjectives and participles when they are the head of an NP. However, the demonstratives and the genitive-marked pronouns cannot host a classifier in attributive function. Consider the following examples.

Demonstratives

In (32a) and (32b), the demonstrative itself functions as an NP and hosts the classifier, whereas in (32c), it appears as a modifier of the noun *am* 'mango' and therefore cannot host classifiers.

(32) a. *i=tho* *kekʌr* *lagʌi?*
this=CLF$_5$ who.GEN COP.PRS.3SG
'Whose is this?'

b. *u=tho* *hamʌr* *bhai=k* *lag-ʌi.*
that=CLF$_5$ 1SG.GEN brother=GEN COP.PRS-3SG
'That (mango) is my brother's.'

c. **u=tho* *am* *hamʌr* *bhai=k* *lag-ʌi.*
that=CLF$_5$ mango 1SG.GEN brother=GEN COP.PRS-3SG
'That mango is my brother's.'

Adjectives

In Khortha, *=ʈho* is attested only with selected adjectives in NP use. For example, (33a) shows the classifier with the interrogative pronoun *kʌun*, and (33b) with an adjective as the head of the NP. However, the ungrammatical example in (33c) demonstrates that it cannot be used attributively.

(33) a. *kʌun=ʈho temrʌs de-b-ẽ?*
which=CLF$_5$ guava give-FUT-2SG
'Which guava will you give (me)?'

b. *boɾ=ʈho de-b.*
big=CLF$_5$ give-FUT.1SG
'I will give (you) the big one.'

c. **boɾ=ʈho temrʌs khai-b.*
big=CLF$_5$ guava eat-FUT.1SG
'I will eat the big guava.'

Genitives

Example (34a) illustrates the use of the classifier *=ʈho* with the genitive-marked pronouns in NP use, while example (34b) shows that this is not possible in attributive function.

(34) a. *tor=ʈho?*
2SG.GEN=CLF$_5$
'(Is it) yours?'

b. **tor=ʈho am le-b.*
2SG.GEN=CLF$_5$ mango take-FUT.1SG
'I will take your mango.'

Participles

Example (35a) shows that it is possible to use the classifier in the participle form in NP use but not in attributive function (35b). However, it is possible to use this classifier after the head noun of the NP, as shown in (35c).

(35) a. *ɖõg-ʌl=(ʈho)*
collect-PTCP=CLF$_5$
'the collected (ones)'

b. **ɖõg-ʌl=ʈho mahua*
collect-PTCP mahua
'the collected mahuwa'

c. *ɖõg-ʌl* *mahua=ʈho*
collect-PTCP mahua=CLF$_5$
'the collected mahuwa'

This classifiers is the least productive among all classifiers in the Sikhari variety of Khortha. It is only attested twice in the entire corpus, and my primary consultant Giridhari Goswami rejects its usage, preferring the classifier *=go* instead. However, in the Parnadiya variety of Khortha, spoken in the border region between the states of Jharkhand and Bihar, the classifier *=ʈho* is more prevalent. This region is culturally and linguistically more diverse, with speakers of multiple varieties of Indo-Aryan, including Bhojpuri and Magahi. The occurrence of the classifier *=ʈho* may be explained by Bhojpuri's influence on the Parnadiya variant.

### *3.4 =hʌr*

The classifier *=hʌr* is commonly used when counting domestic animals that come in pairs, such as of oxen. It typically appears with numerals but may also are appear with genitive-marked pronouns and adjectives. The following examples demonstrate the usage of *=hʌr.* In (36a), *=hʌr* attaches to the interrogative pronoun *kʌi-* 'how many', and in (36b) to a numeral in NP use, while (37a-b) illustrates its appearance with numerals in attributive function.

(36) a. *kʌi=hʌr goru kin-l-ʌi?*
how.many=CLF$_6$ ox buy-PST-3SG
'How many pairs of oxen did he buy?

b. *cair=hʌr kin-l-ʌi.*
four=CLF$_6$ buy-PST-3SG
'He bought four pairs.'

(37) a. *du=hʌr kaɽa*
two=CLF$_6$ buffalos
'two pairs of buffalos'

b. *ek=hʌr goru kin-l-ʌi*
one=CLF$_6$ ox buy-PST-3SG
'He bought a pair of oxen.'

Additionally, as indicated above, *=hʌr* is attested with demonstratives, adjectives and participles as the head of the NP. However, it cannot follow the noun itself. It is worth mentioning that the classifier *=hʌr* is also used in attributive function, unlike most other classifiers. The following two examples demonstrate that it can be used in both NP and modification uses in Khortha.

Demonstratives

(38) a. *i=hʌr* *okʌr* *lagʌi?*
DEM.PROX=CLF$_6$ 3SG COP.PRS.3SG
'This pair is his?'

b. *i=hʌr gʌru=ʈa* *kekeʌr* *lagʌi?*
DEM.PROX=CLF$_6$ ox= CLF$_6$ who.GEN COP.PRS.3SG
'Whose pair of oxen is this?'

Adjectives

(39) a. *boɽ=hʌr* *(goru)*
big=CLF$_6$ ox
'the big pair of oxen'

b. **boɽ goru=hʌr*
big ox=CLF$_6$
'the big pair of oxen'

Genitives

(40) a. *tor=hʌr* *(goru)*
2SG.GEN=CLF$_6$ ox
'Your pair of oxen'

b. *tor=hʌr* *goru*
2SG.GEN=CLF$_6$ ox
'Your pair of oxen'

c. **tor* *goru=hʌr*
2SG.GEN ox=CLF$_6$
'Your pair of oxen

Participles

(41) a. *tor* *jot-ʌl=hʌr*
2SG.GEN plough-PTCP=CLF$_6$
'the pair which you plowed'

b. *tor* *jot-ʌl=hʌr* *goru=ʈa*
2SG.GEN plough-PTCP=CLF$_6$ ox=CLF$_1$
'the pair of oxen which you plowed'

c. **tor* *jot-ʌl* *goru=hʌr*
2SG.GEN plough ox=CLF$_6$
'the pair of oxen which you plowed'

Note that the classifier *=hʌr* in Khortha was originally used to count pairs of oxen, which have

been traditionally used for various domestic activities such as pulling carts or ploughing fields. Over time, the classifier *=hʌr* has extended its usage to other pairs of livestock, including *bhenda* 'sheep', *kaɽa* 'buffalo', *perwa* 'pigeon' and *geɽe* 'duck', e.g., *ek=hʌr kaɽa* 'a pair of buffalos'. It is attested not only in Khortha but also in other regional Indo-Aryan languages for the same meaning. It's etymology may be traced back to the noun *har*, which describes a plough that requires two animals to pull it.

### *3.5 =gãɽa*

The classifier *=gãɽa* [CLF$_4$] in Khortha is used to count items in groups of four, including both inanimate and animate objects. This classifier has been borrowed from Santali[8] (*ganḍa*) into Khortha. A cognate form is also observed in Mundari, cf. *ganḍ*, serving the same counting function.[9] Examples (42)-(43) from the Khortha corpus exemplify its usage.

(42) a. *murgi=ʈi* *kʌi=gãɽa* *dim* *de-l-ʌi?*
Hen=CLF$_2$ how.many=CLF$_4$ egg give-PST-3SG
'How many eggs did the hen lay?'

b. *ek=gãɽa* *de-l-ʌi.*
one=CLF$_4$ give-PST-3SG
'She laid four (= 1x4).'

(43) a. *murgi=ʈi* *du=gãɽa* *ḍim* *de-l=hʌe.*
chicken=CLF$_2$.FEM two=CLF$_4$ egg give-PST=AUX.PRS.3SG
'The hen has laid eight eggs (= 2x4).'

b. *okʌr* *ek=gãɽa* *chʌuwa* *hʌe*
3SG.GEN one=CLF$_4$ child COP.PRS.3SG
'He has four children (= 1x4).'

The classifier *=gãɽa* is not only used in Khortha but also with the same meaning in its donor language, Santali. Example (44) presents two instances, one sourced from a historical Santali-English dictionary and another elicited during my own fieldwork in Ranchi.

(44) a. Santali (Munda)
*pon=ganḍa aphor* *menak'=a.*
four=CLF$_4$ rice.seedlings COP=FIN
'There are sixteen bunches of rice seedlings (= 4x4).'
[Campbell 1899: 177]

[8] Santali, also known as Santal/Santhal, is the most widely spoken language of the Munda subfamily of the Austro-Asiatic language family, spoken mainly in several eastern Indian states, including Jharkhand.

[9] I thank the audience at the International Workshop on the Documentation of Endangered Languages and Cultures: with special reference to Jharkhand, organized by the Dr. Ram Dayal Munda Tribal Welfare Research Institute in collaboration with Dr. Shyama Prasad Mukharjee University on 11th and 12th April, 2019, for rightly pointing out the original source of the classifier *=gãɽa*.

b. Santali (Munda)
*bar=ganḍa sim hopon* *menak'=ko=ta=le=a.*
two=CLF$_4$ hen chick cop=3PL=POSS=1PL=FIN
'(At home) we have eight chicks (= 2x4).'[10]

[Elicited-2019, Ranchi]

The classifier *=gãɽa* has a highly restricted distribution in Khortha. It is not attested after nouns, adjectives, genitives, or participle forms. However, it does appear with demonstratives to some extent. There is only one instance where it is used in a specific situation, particularly when the speakers indicate something by pointing it. In example (45), the speaker shows the four eggs on the ground to the addressee and asks. However, (46) is ungrammatical because the classifier *=gãɽa* cannot be used after nouns.

(45) *i=gãɽa=ʈa* *kariya* *murgi* *de-l-ʌi*
this=CLF$_4$=CLF$_1$ black hen give-PST-3SG
'The black hen laid these four.' (showing four eggs on the ground)

(46) **murgi=ʈi* *du ḍim=gãɽa* *de-l=hʌe*
chicken=CLF$_2$.FEM two egg=CLF$_4$ give-PST=AUX.PRS.3SG
'The hen has laid eight eggs.'

In (45), the classifier *=ʈa* is optional, but it is worth noting that most speakers prefer to include it to highlight the specificity of the object. However, as we lose the sense of "group of four" without the classifier *=gãɽa,* it is obligatory in these contexts.

According to Tiwari (1960: 123), the classifier *=ganḍa* is observed in certain dialects of the Bhojpuri language spoken in India. My survey also revealed its usage in the Bhojpuri spoken in the eastern Terai belt of Nepal.[11] However, it is not attested in the Bhojpuri variety spoken in the western Terai districts of Nepal. This suggests that *=ganḍa* may be present in the Bhojpuri variety spoken in proximity to the Santal-speaking area, indicating a potential contact influence.

### *3.6 =mũɽ*

The classifier *=mũɽ* [CLF$_7$] is used when counting cattle and other domestic animals. This classifier possesses a unique characteristic that distinguishes its usage from others. It typically indicates a singular entity, but when it is used in specific constructions, especially when individuals express their wishes or desires to a higher power, it indicates a pair of items.

The use of the classifier *=mũɽ* is relatively uncommon among the younger generation of Khortha speakers. However, it is more prevalent and commonly used by adult speakers; for example, *cair=mũɽ goru* 'four=head cattle'. In our corpus, it is exclusively attested with quantifiers. Consider the following two examples.

[10] Elicited in Ranchi on 07.12.2019 with a native speaker of Santali named Vivek Bhushan Murmu, resident of Lodhkiyari, Bokāro.
[11] Gopal Thakur (a native speaker of Bhojpuri, resident of Bara district, Nepal), p.c., 2022.

(47) a. *kʌi=mũɽ* *kaɽa?*
how.many=CLF$_7$ buffalo
'How many buffalos?'

b. *cair=mũɽ* *hʌ-thin.*
four= CLF$_7$ COP-PRS.3PL
'He has four.'

(48) *okʌr* *cair=mũɽ* *janwʌr* *hʌ-thin.*
3SG four=CLF$_7$ animal COP-PRS.3PL
'He has four cattle.'

As stated earlier, *=mũɽ* cannot be used with demonstratives, adjectives, or genitives. However, there is an example where it is found with a participle followed by the classifier *=ʈa* (49). This sentence would be ungrammatical if we drop the *=ʈa* classifier.

(49) a. *tor* *bech-ʌl=mũɽ=ʈa*
2SG.GEN sell-PTCP=CLF$_7$=CLF$_1$
'(I wanted) what you sold'

b. *tor* *bech-ʌl=mũɽ(*=ʈa)* *goru=ʈa*
2SG.GEN sell-PTCP=CLF$_7$ ox=CLF$_1$
'(I wanted) the ox that you sold'

Additionally, I observed that the classifier *=mũɽ* is used differently in some special constructions, such as wish constructions, where a speaker wishes for something to the god and offers the sacrifice of an animal in fulfillment of a vow. In such a case, *=mũɽ* expresses the meaning of the two. In example (50a) *adh* is used in the sense of a half (pair), while in example (50b) *ek=mũɽ pãʈha* signals 'two goats', and similarly in example (50c) *du=mũɽ pãʈha* means 'four goats'.

(50) a. *adh=mũɽ* *pãʈha* *de-b=o*
half=CLF$_7$ he.goat give-FUT=ADDR
'(If I pass the current exam) I give a goat.' (lit. I sacrifice a half pair of goat.)
[elicited-Sandeep-Giridhari-022]

b. *he* *khũʈababa* *jʌdi* *hamʌr* *kam=ʈa* *pura*
hey Khunta.lord if 1SG.GEN work=CLF$_1$ complete

*ho-e* *ji-t-ʌi*
be-LNK v2-FUT-3SG

*tʌ* *ek=mũɽ* *pãʈha* *de-b=o*
TOP one=CLF$_7$ goat give-FUT=ADDR
'Hey Khunta Lord! If my work is completed I will offer a pair of he-goats.'
[elicited-Sandeep-Giridhari-022]

c. *du=mũɽ pãʈha de-b=o*
two=CLF$_7$ he.goat give-FUT=ADDR
'(If I pass the current exam) I give a pair of goats.'
[elicited-Sandeep-Giridhari-022]

As shown in the above examples, *=mũɽ* is commonly followed by the noun *pãʈha* 'he-goat', which is the most famous animal for sacrifice to a deity in the Khortha community. But it is also compatible with other livestock, such as *bhenɖa* 'sheep', *kaɽa* 'buffalo', *perwa* 'pigeon', and *geɽe* 'duck'.

It is important to note that out of the seven classifiers discussed above, *=hʌr* and *=gaɽã* have been analysed in this work as classifiers for practical reasons. However, during my personal communication with One-Soon Her, he suggested that these two classifiers rather belong to the measure words category. According to his suggestion, if the final result is not an exact match with the numeral, then it is considered a measure word. For example, as shown in (45a) when someone says, *du=gãɽa ɖim,* the meaning is not simply 'two eggs', but rather 'eight eggs', because this classifier is employed to count items in groups of four (2x4=8). Khortha still preserves its traditional counting system, a vigesimal system, where items are counted in pairs (for example, 2, 4, 20, 40 and so on). Both of these classifiers are exclusively employed with countable nouns.

During my analysis of the Khortha corpus, I noticed that one of the common features of all the sortal numeral classifiers is that they exclusively appear with singular nouns. In the entire corpus, there were no instances of the classifier and the plural marker clitic *=gula/guli* appearing together. Although there was a single occurrence of the *=ʈa* classifier preceding the plural marking suffix *-ʌin* in my elicited data, native speakers rejected it during the verification process (51).

(51) **gidʌr=ʈ-ʌin*
child=CLF$_1$-PL
'children'

## 4 Optionality of the numeral classifiers

Khortha is an obligatory classifier language; however, there are specific contexts where direct noun modification by numeral is possible. Our corpus data shows that classifiers are optional when the numeral precedes kinship terms, as shown in the corpus example (52a) and the elicited example (52b). Furthermore, (52c) demonstrates that numerals do not require classifiers when they are followed by postpositions.

(52) a. *e=go gãw=e sat bhae ar ek bʌhin rʌhʌ-th.*
one=CLF$_3$ village=LOC seven brother and one sister COP.PST-3PL

*sat=o bhae mil-i=ke e=go pokhʌir*
seven=ADD brothers being.together-LNK=SEQ one=CLF$_3$ pond

*koɽ-e lag-l-a.*
dig-LNK v2-PST-3PL

'There were seven brothers and a sister in a village. All seven brothers started to dig a pond together.'

[Pokharik dan, Mahto 2018: 75]

b. *hamin* *du(=ʈa)* *bhai* *h-i-a.*
1PL two=CLF$_1$ brother COP-1-PL
'We are two brothers.'

c. *siyan* *how-it* *jhumari=o* *ek* *baʈhe* *ja-i=ke*
young be-IPFV a_girl=ADD one side go-LNK=SEQ

*kʌps-e* *lag-l-i.*
cry-LNK start-PST-3SG.FEM
'The growing Jhumari also started to cry in a corner.'

Another noteworthy feature of numeral classifiers is that it is often grammatically possible to omit the noun and use only the numeral plus classifier. These types of omissions are particularly common in responses to questions. For instance, in response to the question provided in example (42a) above, a typical answer would be *cair=gãɽa*, four=CLF$_4$ 'sixteen (eggs)'.

## 5 Word order

The most common arrangement involving a numeral and classifier is [NUM CLF N], where the numeral precedes the classifier and the noun. However, there are instances where the numeral and classifier can be postposed, appearing as [N NUM CLF]. It's crucial to note that this postposed structure is only possible when the numeral and classifier directly precede the noun. Other formations, such as demonstrative=CLF or genitive=CLF, cannot undergo this postposing. Despite attempts to elicit such utterances, my informants rejected them (53b-d). An example of the postposed word order is seen in (53a).

(53) a. *hamʌr* *ghar* *chageir* *cair=ʈi* *hʌi.*
1SG.GEN home goat four=CLF.FEM COP.PRS.3SG
'There are four goats (fem) at my home.'

b. **kitap* *i=ʈa* *hamʌr* *lagʌi.*
book DEM.PROX=CLF 1SG.GEN COP.3SG
'This book is mine.'

c. **gidʌr* *tor=ʈa* *ge-l-ʌi.*
child 2SG.GEN=CLF go-PST-3SG
'Your child went away.'

d. *ḍim i=gãɽa=ʈa kʌriya murgi de-l-ʌi.
egg DEM.PROX=CLF=CLF black hen give-PST-3SG
'The black hen laid these four eggs.'

In all the corpus data I analyzed, the noun consistently appears adjacent to the numeral. However, there is an example extracted from written literature where the noun is not positioned adjacent to the numeral. Consider the following example where the head noun *bon* 'jungle' appears at the end of the sentence.

(54) e=go rʌhʌ hʌ-l-ʌi khub gajar bon.
one=CLF$_3$ remain COP-PST-3SG very dense jungle
'There was a very dense jungle.'

[sɑ̃pek sʌhai, Jha 2014: 41]

## 6 Interim Summary of classifiers

In this study, we have discussed the numeral classifiers in Khortha, providing relevant examples from the corpus, exploring their usage with numerals, nouns, demonstratives, genitives, adjectives, and participles. Our analysis reveals that except with kinship terms, the classifiers are obligatory with numerals both in noun phrases (NP) and in attributive function. However, they are optional with DEM(monstrative), GEN(itive), ADJ(ective) and Participle in NP use and ungrammatical in modification use. To summarize these occurrences systematically, we present them with individual formulas:

***=ʈa***

| | | | | | |
|---|---|---|---|---|---|
| NP use: | NUM > | $^{OPT}$DEM > | $^{OPT}$GEN > | $^{OPT}$ADJ > | $^{OPT}$V-based participle |
| Attr: | NUM > | $^{*}$DEM > | $^{*}$GEN > | $^{*}$ADJ > | *V-based participle |

***=ʈi***

| | | | | | |
|---|---|---|---|---|---|
| NP use: | NUM > | $^{OPT}$ DEM > | $^{OPT}$ GEN > | $^{OPT}$ ADJ > | $^{OPT}$ V -based participle |
| Attr: | NUM > | $^{*}$ DEM > | $^{*}$ GEN > | $^{*}$ ADJ > | * V -based participle |

***=go/=goʈ***

| | | | | | |
|---|---|---|---|---|---|
| NP use: | NUM > | $^{OPT}$ DEM > | $^{OPT}$ GEN > | $^{OPT}$ ADJ > | $^{OPT}$ V -based participle |
| Attri: | NUM > | $^{*}$ DEM > | $^{*}$ GEN > | $^{OPT}$ ADJ > | * V -based participle |

***=ʈho***

| | | | | | |
|---|---|---|---|---|---|
| NP use: | NUM > | $^{OPT}$ DEM > | $^{OPT}$ GEN > | $^{OPT}$ ADJ > | $^{OPT}$ V -based participle |
| Attr: | NUM > | $^{*}$ DEM > | $^{*}$ GEN > | $^{*}$ ADJ > | * V -based participle |

***=hʌr***

| | | | | | |
|---|---|---|---|---|---|
| NP use: | NUM > | $^{OPT}$ DEM > | $^{OPT}$ GEN > | $^{OPT}$ ADJ > | $^{OPT}$ V -based participle |
| Attr: | NUM > | DEM > | GEN > | ADJ > | V -based participle |

***=gãɽa***

| | | | | | |
|---|---|---|---|---|---|
| NP use: | NUM > | $^{OPT}$ DEM > | $^{OPT}$ GEN > | $^{OPT}$ ADJ > | $^{OPT}$ V -based participle |
| Attr: | NUM > | $^{*}$ DEM > | $^{*}$ GEN > | $^{*}$ ADJ > | * V-based participle |

***=mũɽ***

| | | | | | |
|---|---|---|---|---|---|
| NP use: | NUM > | $^{*}$ DEM > | $^{*}$ GEN > | $^{*}$ ADJ > | $^{OPT}$ V -based participle |
| Attr: | NUM > | $^{*}$ DEM > | $^{*}$ GEN > | $^{*}$ ADJ > | * V -based participle |

In the upcoming sections of this paper, I introduce measure words and their functions in the Khortha language.

## 7 Measure words

In addition to numeral classifiers, there are several dozen measure words (MW) that describe physical properties like shape, size and quantity of both mass nouns and count nouns in Khortha. These measure words are assumed to belong to the class of classifiers and described as "mensural classifiers" in the literature. However, unlike the numeral classifiers, measure words are used in combination with a numeral to indicate an amount or quantity of something often represented by a mass noun (e.g., "two grains of rice"). Numeral classifiers are clitics and function as bound morphemes, while measure words are independent words that can also occur in the absence of numeral. However, like the numeral classifiers, measure words also occupy the same slot, i.e., NUM MW N or N NUM MW. Numeral classifiers and measure words are mutually exclusive in Khortha.

Measure words are of various types, ranging from general ones compatible with several different types of nouns to highly specialized ones, such as *sɑ̃khri* 'a piece of cooked-rice', that are highly specialized. Their usage as a numeral classifier or a measure word may yield different meanings. For example, *e=go kʌira* [one=CLF banana] means simply 'a banana', while *ek kandhi kʌira* [one MW banana] means a whole banana stick containing several bananas. Unlike classifiers, measure words are syntactically very close to nouns. This is probably the reason that measure words, but not classifiers, can be modified by adjectives.

Based on their semantics, measure words, are categorized into the following classes: (i) measure words for partitioning (55a-j); (ii) measure words for grouping or bunches (57a-e); (iii) container words (59a-e); (iv) measure words for indicating amount or quantity (60a-e); and (v) temporal measure words (62a-b).

(i) measure words for partitioning

Partitioning measure words indicate the division or partitioning of a larger quantity into smaller parts. They are extremely common in everyday speech, facilitating the division of objects into smaller units, such as drops of water, grains of rice or seeds, pieces of land and so on.

(55) a. *dana* 'a seed of rice, fruits, or grains' — *ek dana dhan* 'a seed of paddy'
b. *pata* 'a flat thin object' — *ek pata kʌrkʌɽ* 'a piece of tin'
c. *phoɽa* 'a clove of garlic' — *ek phoɽa rʌsun* 'a clove of garlic'
d. *ɽipik* 'a drop of water' — *ek ɽipik pani* 'a drop of water'
e. *sɑ̃khri* 'for a piece of cooked-rice' — *ek sɑ̃khri bhat* 'a piece of rice'

| | | |
|---|---|---|
| f. | *cʌkʌta* 'a slice' | *ek cʌkʌta am* 'a slice of mango' |
| g. | *pahi* 'for a part (straight)' | *ek pahi dhan* 'a line of paddy plants' |
| i. | *ʈukra* 'a piece' | *ek ʈukra bari* 'a piece of land' |
| j. | *kudhi* 'a part' | *ek kudhi mas* 'a portion of meat' |

Some of these measure words, e.g., *pahi*, *ʈukra,* and *kudhi* are illustrated with examples from the Khortha corpus in (56).

(56) a. *ek pahi dhan rop-ai-l hʌe*
one straight.line paddy plant-PASS-PST AUX.PRS.3SG
'(Out of the entire paddy field), one part in a straight line has been planted.'

b. *ghar=ẽ sʌʈʌl=hẽ ek ʈukra bari*
home=LOC connected=FOC one piece field
'(He had also) a piece of land connected with the home'
[Text02-Mae ke lor]

c. *e=go chagʌir=k cair kudhi kʌrʌ-l ar*
one=CLF goat=GEN four part do-PST and

*jʌkʌr=ẽ budhana du kudhi mas uʈhwʌ-l*
RELGEN=LOC Budhana two part meat take-PST
'(Mangra) cut and divided the goat into four pieces and Budhana took two portions out of that.'
[elicited-Giridhari-Goshwami-2022]

(ii) Group or a bunch

Group or bunch measure words are frequently employed to quantify collection or clusters of countable nouns, indicating that items are gathered or organized into groups. Here are some examples of the most typical group or bunch measure words in Khortha:

| | | | |
|---|---|---|---|
| (57) | a. | *kandhi* 'a whole banana stick' | *ek kandhi kʌira* 'a stick of bananas' |
| | b. | *ghoda* 'a bunch of mangos' | *ek ghoda am* 'a bunch of mangos' |
| | c. | *curua* 'a palm full of X' | *cair curua ghi* 'four palm full of ghee' |
| | d. | *gõcha* 'a bundle of dried paddy plants and grass' | *ek gõcha ghãs* 'a bundle of grass' |
| | e. | *jawa* 'a bulb of garlic' | *ek jawa rʌsun* 'a bulb of garlic' |

An example of a group or bunch measure word is exemplified in (58).

(58) *ek ghoda tetʌir par-l-i=o*
one bunch tamarind pluck-PST-1=ADDR
'I plucked a bunch of tamarind (from the tree).'

(iii) Container

Container measure words are used to count or quantify the items that can be contained or

measured within a particular container. Consider the following examples:

| | | | |
|---|---|---|---|
| (59) | a. | *pʌila* 'a measure pot' | *ek pʌila car* 'roughly a kilogram of rice)' |
| | b. | *khomaca* 'a fist full' | *ek khomaca bhat* 'a fist full of rice' |
| | c. | *õjʌra* 'two palms' | *ek õjʌra car* 'one two-palms full of rice' |
| | d. | *pʌĩʈi* 'half a glass' | *ek pʌĩʈi daru* 'a half-glass of local sprit' |
| | e. | *ʌkwair* 'arm-full' | *ek ʌkwair ghas* 'an arm-full of grass' |
| | f. | *khoĩcha* 'lap' | *ek khoĩcha car* (used with grains) 'carrying something by folding the front part of the t-shirt' |

(iv) amount or quantity of something

Amount or quantity measure words are used to quantify a bulk or large quantity of items. These measure words are especially employed when counting a significant amount of a particular item or substance. Here are some examples:

| | | | |
|---|---|---|---|
| (60) | a. | *hoɽ* '1,000 pieces' | *ek hoɽ ĩʈa* '1,000 bricks' |
| | b. | *buda* 'root' | *du buda bãs* 'two farms of bamboos' |
| | c. | *goʈh* 'group' | *ek goʈh goru* 'a group of cattle' |
| | d. | *kʌʈahi* '40 kg' | *ek kʌʈahi dhan* 'a 40 kilograms of rice' |
| | e. | *pai* 'amount' | *ek pai pʌiso* 'a tiny amount of money' |

An example of amount or quantity measure word is exemplified in (61).

(61) *hamʌr hat=hẽ ek pai pʌiso nãe he-l*
1SG.GEN hand=LOC ek some money neg COP-PST
'I had not even a single penny on my hand.'

[Text02-Mae ke lor]

(v) temporal measure words

Terms that denote specific time periods or durations are known as temporal measure words. Here are a few typical instances:

| | | | |
|---|---|---|---|
| (62) | a. | *din* 'day' | *ek din* 'a day' |
| | b. | *ghʌrik* 'a moment' | *ek ghʌrik cand* ' for a short period' |

It is important to note that the choice of measure words is influenced by the characteristics and attributes of the noun being quantified. For example, the measure term for a group or a bunch of items like *ghoda* (58) is also applied to similar, small and round fruits grown on trees, such as litchi, tamarind. Similarly, *curua* (57c) always precedes nouns denoting liquids, such as oil, water, ghee and so on. It's used as a kind of measurement tool formed by bending the four fingers of the palm. Here are some examples from our corpus and elicited data.

In Khortha, it is considered ungrammatical to use double classifiers. For instance, **e=go phoʈa rʌsun* 'a clove of garlic', is ungrammatical because the numeral classifier *=go* and the measure word *phoʈa* cannot coexist in a phrase. However, it is notable that an NP with an internal measure word can be marked as a whole by *=ʈa,* as shown in (63).

(63) *abga ek gõcha poũra=ʈa an-l-ẽ*
only one bundle dried.paddy=CLF$_1$ bring-PST-2SG
'You brought only a bundle of dried paddy plants (I had told you to bring five bundles or more).'

Furthermore, alongside the measure words discussed earlier, Khortha also employs various other nouns as measure words, commonly referred to as 'measure nouns'. These measure nouns are directly combined with a numeral to indicate quantity or measurement. For example, *ek balti pani* meaning 'a bucket of water' illustrates the use of a measure noun.

## 8 Summary

This study provides a comprehensive analysis of previously undocumented classifiers in Khortha, encompassing both sortal and mensural types. The data presented in this research demonstrates that among the numeral classifiers, *=ʈa*, *=ʈi*, *=go*, and *=gãɽa* are compatible for use with both human and non-human nouns, while *=ʈho* is predominantly associated with fruits, and *=hʌr* and *=mũɽ* are primarily used for counting animals. It's noteworthy that the noun cannot intervene between the numeral and the classifier. The most common construction involving a numeral and classifier is [NUM CLF N], but occasionally, the classifier is postposed, i.e., [N NUM CLF]. Numeral classifiers and measure words both occupy the same slot in the construction and are therefore mutually exclusive. Some numeral classifiers can attach to both numerals and nouns, demonstratives, adjectives and quantifiers, while some of the classifiers are confined to numerals. When the numeral classifiers are used postnominally, they mark the specificity or definiteness of the object; however, such specificity is not found when they attach to numerals. Table 1 summarizes the occurrences and orders of numerals, classifiers and nouns.

| Classifiers | Distribution | Gender agreement | Semantic basis |
|---|---|---|---|
| ***=ʈa*** | Numeral=clf$_1$+(modifier)+Noun<br>Quantifier= clf$_1$<br>Noun= clf$_1$<br>Demonstrative=( clf$_1$)<br>Genitive=( clf$_1$)<br>Adjective=( clf$_1$)<br>Participle=( clf$_1$) | Yes | non-feminine nouns (includes both human and non-human nouns) |
| ***=ʈi*** | Numeral=clf$_2$+(Modifier)+Noun.FEM<br>Quantifier=clf$_2$<br>Noun.FEM=clf$_2$<br>Demonstrative=clf$_2$<br>Genitive=clf$_2$<br>Adjective=clf$_2$ | Yes | feminine nouns, small animals, objects, and equipment |
| ***=ʈho*** | Numeral=clf$_5$+Noun<br>Quantifier= clf$_5$<br>Demonstrative= clf$_5$ | No | fruits, human and nonhuman nouns |

| | | | |
|---|---|---|---|
| | Genitive= $clf_5$<br>Adjective= $clf_5$ | | |
| ***=go/goɽ*** | Numeral=$clf_3$+(Modifier)+Noun<br>Quantifier= $clf_3$<br>Adjective=$clf_3$ | No | both animate and inanimate nouns |
| ***=hʌr*** | Numeral=$clf_6$+Noun<br>Quantifier= $clf_6$<br>Adjective=( $clf_6$)<br>Genitive=( $clf_6$)<br>Participle=( $clf_6$) | No | domestic animals in pairs |
| ***=gãɽa*** | Numeral= $clf_4$+Noun<br>Quantifier=$clf_4$<br>Demonstrative= $clf_4$ | No | counts both animate and inanimate nouns in groups of four |
| ***=mũɽ*** | Numeral=$clf_7$+Noun<br>Quantifier= $clf_7$<br>Participle= $clf_7$ | No | common with cattle, occasionally with other livestock such as sheep, pigeon, duck |

Table 1. Khortha classifiers and their distribution

NP use: num > $^{OPT}$dem > $^{OPT}$gen > $^{OPT}$adj > $^{OPT}$v-based participle
Attributive function: num > $^{*}$dem> $^{*}$dem> $^{(*)}$adj> $^{(*)}$v-based participle

(*) depends on classifiers, some permit, others don't.

Measure words in Khortha do not constitute a single homogeneous class. Some, like *sʌ̃khri,* are highly specialized, while others are more general and compatible with several different types of nouns. These versatile measure words often behave like nouns themselves.

## ABBREVIATIONS[12]

| | | | |
|---|---|---|---|
| 1, 2, 3 | 1st, 2nd, 3rd person | IMP | imperative |
| ADDR | addressee | INST | instrumental |
| AUX | auxiliary | LOC | locative |
| $CLF_1$ | classifier I (=*ʈa*) | LNK | linker |
| $CLF_2$ | classifier II (=*ʈi*) | MW | measure word |
| $CLF_3$ | classifierIII(=*go*/*goɽ*) | OBL | oblique |
| $CLF_4$ | classifier IV (=*gaɽã*) | PL | plural |
| $CLF_5$ | classifier V (=*ʈho*) | POSS | possessive |
| $CLF_6$ | classifier VI (=*hʌr*) | PROX | proximate |
| $CLF_7$ | classifier VII (=*mũɽ*) | PRS | present |

[12] Only those not in Leipzig Glossing Rules.

| | | | |
|---|---|---|---|
| COP | copula | PST | past |
| DEM.DIST | demonstrative distal | PTCP | participle (*-ʌl*) |
| FEM | feminine | Q_ | question |
| FUT | future | SEQ_ | sequential converb |
| GEN | genitive | SG | singular |

## Acknowledgements

I extend my sincere gratitude to Masayoshi "Matt" Shibatani and Prashant Pardeshi for their comments and suggestions. Special thanks to Sandip K. Mahto, Dinesh Kumar Dinmani and Giridhari Goswami for their valuable assistance in eliciting a substantial amount of fresh data on various 'measure words' and numeral classifiers. I also express my appreciation to Peter E. Hook for suggesting references to historical works on the related topic, and to One-Soon Her for a very productive discussion via Google Meet. Additionally, I would like to thank two anonymous reviewers for their useful comments and suggestions.

The research reported here was part of the project "Towards a Linguistic prehistory of Eastern-Central South Asia (and beyond)", led by Prof. John Peterson at the University of Kiel and funded by the DFG, which I gratefully acknowledge here.

## References

Aikhenvald, A. Y. 2000. *Classifiers*: *A typology of noun categorization devices*. Oxford: Oxford University Press.

Allassonnière-Tang, Marc; and Kilarski, Marcin. 2020. "Functions of Gender and Numeral Classifiers in Nepali". *Poznań Studies in Contemporary Linguistics* 56.1: 113-168. https://doi.org/10.1515/psicl-2020-0004

COG. 2019. Copyright: SIL International. Version 1.3.4.10016, at http://software.sil.org/cog Accessed Jul 9, 2019.

Dayal, Veneeta. 2012. "Bangla classifiers: Mediating between kinds and objects". *Italian Journal of Linguistics* 24.2: 195-226 at https://www.italian-journal-linguistics.com/app/uploads/2021/05/2_dayal.pdf Accessed Jun 11, 2024.

Eberhard, David M.; Simons, Gary F.; and Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the world.* Dallas: SIL International, at http://www.ethnologue.com Accessed Jan 15, 2022.

Emeneau, M. B. 1956. "India as a linguistic area". *Language* 32.1: 3-16. https://doi.org/10.2307/410649

Gil, David. 2013. "Numeral Classifiers". In: Dryer, Mathew S.; and Haspelmath, Martin (eds.), WALS Online (v2020.3). http://wals.info/chapter/55. Accessed May 9, 2024

GOI (Government of India). 2011. *Census of India 2011.* New Delhi.

Grierson, Sir George Abraham. 1903. *Linguistic survey of India Vol. V: Indo-Aryan Family, Eastern Group Pt. II: Specimens of the Bihari and Oriya Languages*. Calcutta: Superintendent of Government Printing, India.

Jha, A. K. 2012. *Khortha bha ̄sa ̄ ka baijña ̄nik Adhyayan* (A scientific study of the Khortha language).
Ranchi: Jharkhand Jharokha.

Jha, A. K. 2012. "Luirgar beti chauwa" (A strong daughter). In: Jha, A. K. ; Mahto, Chitranjan; Pramanik, Shivnath; Mahto, Shyam Sundar; Ohdar, B. N. ; Dinmani, Dinesh Kumar ; and Goswa¯mi, Giridha¯ri (eds.), *Khortha Loksahit* (Khortha literature). Ranchi: Jharkhand Janjatiya Kalyan Shodh Sansthan.

Neukom, Lukas. 2001. *Santali*. München: Lincom Europa [Languages of the World/Materials 2].

Paudyal, Netra P. ; and Peterson, John. 2019. Language contact between "Sadani" and the Tribal Languages of Jharkhand. Paper presented at the 35th South Asian Language Analysis Roundtable (SALA-35). INALCO, Paris.

Paudyal, Netra P. ; and Peterson, John. 2021. "How one language became four: The impact of different contact-scenarios between 'Sadani' and the tribal languages of Jharkhand". In: Ivani, Jessica K.; and Peterson, John (eds.), *Language contact in South Asia* Special issue *Journal of South Asian Languages and Linguistics* 7.2: 327–358. https://doi.org/10.1515/jsall-2021-2028

Paudyal, Netra P. Forthcoming. *A grammar of Khortha*. Leiden, Boston: Brill.

Peterson, John; and Chik Baraik, Sunil. 2023. *A grammar of Chotanagpuri Sadri: A lingua franca of eastern central India*. Mysore: Central Institute of Indian Languages (CIIL).

Tiwari, Udai Narain. 1960. *The origin and development of Bhojpuri.* Calcutta: The Asiatic Society.

Verma, Sheela. 2007. "Magahi". In: Cardona, George; and Jain, Dhanesh (eds.), *The Indo-Aryan languages,* 498-514. London: Routledge.

Verma, Sheela. 1985. *The structure of the Magahi verb*. Chandni Mahal; New Delhi: Manohar.

Netra P. Paudyal
n.paudyal@isfas.uni-kiel.de
netrapaudyal@yahoo.com