# Classifiers in Dimasa and (in-)definite marking

Jonathan P. Evans
Institute of Linguistics, Academia Sinica

Dimasa (ISO 639-3 dis) is a Bodo-Garo language within the Tibeto-Burman family of languages. Like other Tibeto-Burman languages of Northeast India, it has a system of numeral classifiers which occur prefixed to their respective numerals. Across the dozen or so languages of Bodo-Garo, it has been noted that the classifier-numeral word can be placed either before or after the counted noun. The existence of numeral classifiers, as well as the CLF-NUM order may have arisen in this area due to contact with Tai languages during the Ahom kingdom period from the 1200's to 1800's. In the Tai languages, classifiers precede the number "one". By comparing occurrences of classifiers in Dimasa texts, it can be seen that the order N CLF-NUM corresponds to definite nouns, while the order CLF-NUM N marks indefinites. In addition to classifier placement, Dimasa definiteness is also marked by demonstratives, case marking, topic marking, and/or combinations of these strategies.

**Keywords:** Bodo-Garo, constituent order, definite, indefinite

## 1. Introduction and overview of Dimasa

The present study focuses on the numeral classifiers of Dimasa (ISO 639-3 dis), a language of the Bodo-Garo branch of the Tibeto-Burman family. Dimasa is spoken mainly in the Dima Hasao district of Assam state, northeast India. Smaller populations live in neighboring districts, as well as in Nagaland and Manipur. The language has about 110,000 speakers. Older speakers tend to be fluent in both Dimasa and Bengali, while younger speakers are usually fluent in Dimasa, Hindi, and English. Those who receive education in Guwahati, the capital of Assam, are also fluent in Assamese. There are many lexical borrowings from Indo-Aryan languages, especially Bengali; for some borrowed words, it is not clear which Indo-Aryan language is the source. Many newer words have been borrowed from English. The language has a young but vigorous writing culture, with newspapers, blog posts,

https://doi.org/10.1075/alal.22007.eva
*Asian Languages and Linguistics* 3:2 (2022), pp. 181–201. ISSN 2665-9336 | E-ISSN 2665-9344

and literature written in Dimasa, as well as new music and drama transmitted via online videos. Younger speakers prefer writing Dimasa in Roman letters, in a semi-standardized script.

Grammatical analysis of Dimasa in this study is based on Evans and Langthasa (2023); most of the supporting examples cited below are drawn from texts presented in that source. About sixty oral and written texts were collected by Dhrubajit Langthasa primarily between 2016 and 2020. These texts are being compiled into the CoRSAL digital archive (https://corsal.unt.edu/dimasa-language-resource), which is designed to be freely accessible. The collection covers a wide range of genres, including folk tales, personal reminiscences, explanations of processes, accounts of historical events and figures, news stories, drama, proverbs, spontaneous dialogues, etc. All data collection and storage has been conducted according to Academia Sinica IRB standards.

As in most Tibeto-Burman languages, Dimasa sentences are typically verb final. Like other Bodo-Garo languages, Dimasa has an agglutinative morphology, with more suffix positions than prefix positions. Dimasa verbs can take up to two prefixes, and as many as three or four suffixes. Verb concatenation is common and some grammatical markers descend historically from verbs, such as benefactive /-ɾi/, from the homophonous verb "give". Like many Tibeto-Burman languages, there is no syntactic class of adjectives. Adjectival concepts are expressed either as verbs or as nouns, depending on syntactic context. Case marking is cliticized onto the end of noun phrases.

Noun phrases can occur with accusative, genitive, dative, benefactive, ablative, instrumental or comitative clitics. Nominative case is unmarked. Marking of other cases correlates with the definiteness of the noun referent (Section 3); differential case marking has been documented across Bodo-Garo and other Tibeto-Burman languages (DeLancey 2011; Haokip and Brahma 2018). Nouns that are [+ human] can be marked for plural; less animate nouns do not take plural marking. Verbs have temporal marking which combines features of tense and aspect (Table 1).

In terms of transitivity, causatives, benefactives and applicatives increase valence, while reciprocals, reflexives, and passives reduce valence. At the clause level, grammatical relations are marked by a combination of constituent order and clitics. Without topicalization, constituent order is AOV/SV. Some utterances like equatives do not require a verb, as in "This is a dog" (1), which only has topic marking on the arguments.

(1) Verbless declarative[1]
*police-raw ti -ka 'i-bu de sisa se'*
police-PL say-PRF PRX-DEM TOP dog TOP
"The police officers said, 'This is a dog.'"

**Table 1.** Tense-aspect marking

| | Label | Tense | Aspect |
|---|---|---|---|
| /-re/ | "Split imperfective" | Matrix verb: Past<br>Aux: present | Matrix verb: Habitual<br>Aux: Stative |
| /-ba/ | Non-Future Imperfective | Non-Future | Imperfective |
| /-du/ | Present Imperfective | Present | Imperfective (includes continuous, habitual, futurate, generic) |
| /-bi/ | Stative | Non-Future | Imperfective |
| /-saj/ | Continuative | any | Continuous |
| /-ko/ | Durative Progressive | Non-Future | Progressive |
| /-ka/ | Perfect | Non-Future | Perfect; Change of State; Inceptive |
| /-ma/ | Prospective | Non-Past | Prospective |

Clauses with a retrievable referent allow elision (2b). PRO drop also occurs in imperatives and identity clauses.

(2) a. Elided NP.
*Rahul nadi klaj-ba?*
PN what do-NF:IPFV
"What is Rahul doing?
b. *tu-ba*
sleep- NF:IPFV
"Sleeping."

Multiple tense/aspect and mood markers can be affixed to a verb, as in (3)

(3) Tense/aspect marking with two mood markers and one tense-aspect marker
*bu dini doŋ-pu-du-mu*
3SG today EXST-POT-PIPFV-SBJV
"He could have stayed today."

Dimasa numerals are used only cardinally, typically prefixed by a classifier. The Dimasa numeral system combines decimal and vigesimal patterns. The numbers from one to nineteen reflect a decimal structure (Table 2). The number "twenty" is monomorphemic and does not appear to be related to "two". "Fifteen" shows

**1.** Dimasa words are transcribed phonologically. Proper nouns and English borrowings are transcribed according to their own traditions. Counted noun phrases have the classifier-numeral word in bold and the noun underlined.

phonological reduction of the morphemes /zi/ "ten" and /bŋwa/ "five". Numbers in the twenties follow a similar pattern as the teens: /CLF-kon-se/ "twenty-one".

**Table 2.** Numerals from one to twenty

| Numeral | Gloss | Numeral | Gloss |
|---|---|---|---|
| / CLF-se/ | "one" | / CLF-zi-se/ | "eleven" |
| / CLF-gin/ | "two" | / CLF-zi-gin/ | "twelve" |
| / CLF-tam/ | "three" | / CLF-zi-tam/ | "thirteen" |
| / CLF-bɾi/ | "four" | / CLF-zi-bɾi/ | "fourteen" |
| / CLF-bŋwa/ | "five" | / CLF-zɾa/ | "fifteen" |
| / CLF-do/ | "six" | / CLF-zi-do/ | "sixteen" |
| / CLF-sni/ | "seven" | / CLF-zi-sni/ | "seventeen" |
| / CLF-zaj/ | "eight" | / CLF-zi-zaj/ | "eighteen" |
| / CLF-sku/ | "nine" | / CLF-zi-sku/ | "nineteen" |
| / CLF-zi/ | "ten" | / CLF-kon/ | "twenty" |

Except for "fifty", the multiples of ten from "forty" to "ninety" are vigesimal, where the bound form for "twenty" /bisa/ is borrowed from Bengali. Thus, "ninety" is "four *bisa* and ten", etc. (Table 3). The classifier occurs before the single digit of a morphologically complex number, as in "sixty-four bullets" (4). As shown in Table 3, round numbers built on /bisa/ do not take classifiers.

**Table 3.** Round numerals

| Numeral | Morpheme gloss | Lexical gloss |
|---|---|---|
| /CLF-kon/ | CLF-"twenty" | "twenty" |
| /CLF-tim-zi/ | CLF-"three"-"ten" | "thirty" |
| /bisa-gin/ | "twenty"-"two" | "forty" |
| /CLF-dan/ | CLF-"fifty" | "fifty" |
| /bisa-tam/ | "twenty"-"three" | "sixty" |
| /bisa-tam CLF-zi/ | "twenty"-"three" CLF-"ten" | "seventy" |
| /bisa-bɾi/ | "twenty"-"four" | "eighty" |
| /bisa-bɾi CLF-zi/ | "twenty"-"four" CLF-"ten" | "ninety" |
| /ɾza-si/ | "hundred"-"one" | "hundred" |
| /ɾziŋ-si/ | "thousand"-"one" | "(one) thousand" |

(4) Classifier before single digit of complex number ("sixty-four")
*Operation klaj-jaraw dol* ***ma-si*** *AK47* ***ma-bri*** *magazine*
operation do-REL group CLF-one AK47 CLF-four magazine
*odehe* ***bisa-tam ma-bri*** *gulitaj* ***ma-si*** *AR maj-ba.*
and 20–3 CLF-four bullet CLF-one AR get-NF:IPFV
"The group doing the operation found one AK 47, four magazines and 64 bullets; (they) got one automatic rifle."

Ordinal expressions borrow numbers from English, as in (5a–c).

(5) Examples of ordinal numerals "first", "second", "third" from texts.
a. *odehe HSLC odehe HS=ha first Div la-hi pass za-jaraw*
and HSLC and HS=LOC 1st division take-LOC pass be-REL
"In addition, the ones who secured 1st Division in the HSLC and HS,..."
b. *police-raw second bosta ke zik pa=ba=ha bosta bisiŋ nipraŋ*
police-PL 2nd sack ACC kick attach=NMLZ= LOC sack inside ABL
*mjaw mjaw briŋ-ka*
ONP ONP emit.sound-PRF
"When the policemen aimed and kicked the 2nd sack, the sound "Meow Meow" came from inside the bag,..."
c. *u-laj niŋ third station=ha Asampa bu hadisa ke seŋ pin-ka*
DIST-like TOP 3RD station=LOC PN DEM Bengali ACC ask repeat-PRF
"Similarly, at the third station Asampa asked the Bengali person again."

## 2. Counted nouns and classifiers

Classifiers in Dimasa occur only as bound morphemes that are prefixed to numerals. The same pattern is found throughout the Bodo-Garo branch of the Tibeto-Burman family, as well as numerous other Tibeto-Burman languages in Northeast India and surrounding areas (Evans 2022). Outside of this area, few languages have been observed to place classifiers before numerals. Notably, in Warekena, the numbers one and two take classifiers that are prefixed or fused with the numeral; higher numbers, which are borrowed from Portuguese, do not take classifiers (Aikhenvald 2000: 109). The Chibchan language Chimila of Columbia also has classifiers that are prefixed to numerals (Malone 2004; Aikhenvald forthcoming, Chapter 3). Dimasa does not have noun classifiers; classifiers always occur with a number.

The sortal classifiers that have been identified thus far are in Table 4. The inventory of classifiers is relatively small, compared to classifier-rich languages of Asia such as Chinese and Thai. Most classifiers clearly originate from repetition (Aikhenvald 1994, 2000). That is, they formed in the same way that some languages repeat (part of) a noun to form a classifier; e.g., Mandarin for "three

rooms" 三間房間 *sān jiān fángjiān* "three" "CLF:room" "room" (Jackson T.-S. Sun, p.c.). The largest subset of classifiers with a recognizable origin are of this autoclassifier sort. The next group of classifiers in Table 4 come from a word ("flat", "body") that is semantically associated with the noun referent, but is not a repetition. The classifier /saw-/ for person has a suppletive form /g-/ used only with numbers one and two. Finally, there is a small residue whose origin remains obscure.

Dimasa classifiers categorize noun referents with regard to humanness, shape, inherent nature, and arrangement. In addition, there are some highly specific or unique classifiers, such as *baw-* "thought", *pin-* "times", *soŋ-* "bamboo", and *poŋ-* "bamboo container". "Bamboo" and "bamboo container" are culturally salient objects, hence typical specific or unique classifiers (Aikhenvald 2021). The classifier *taj-* "fruit" has undergone extensive semantic extension to be used for other round objects like eyes, and also words or parts of a text. The default classifier *ma-* has the highest textual frequency, and is applied to types of nouns not indicated in Table 4, such as "group" (42), weapons (42), "truck" (9), "village" (41), animals (37), and many other cases.

**Table 4.** Dimasa sortal classifiers and their origins

| | | |
|---|---|---|
| *Repeater origin* | | |
| /taj-/ | "fruit-like" (fruit, eyes, words) | from /btaj/ "fruit" |
| /poŋ-/ | "bamboo container" | from /poŋ/ "container" |
| /groŋ-/ | "seed like"(rice, grains) | from / bgroŋ/ "seed" |
| /baw-/ | "thought" | from /baw/ "think" |
| /pin-/ | "times, repetitions" | from /pin/ "return" |
| /paŋ-/ | "plant" | from /sampaŋ/ "grassy plant" |
| /dep-/ | "branch" | from /bdep/ "branch" |
| /do-/ | "chunk" (meat) | from /bdo/ 'solid chunk' |
| /di-/ | "egg like" | from /di/ 'egg' |
| /diŋ-/ | "hair like" | from /bdiŋ/ 'creeper' |
| /to-/ | "rupees" | from /b-tho/ 'skin lump' (coins used to be very small) |
| /baɾ-/ | "flower" | from /bubaɾ/ 'flower', 'blossom' |
| /graŋ-/ | "flat" (book, clothes) | from /bgraŋ/ 'wing' |
| /poŋ-/ | "blows, strikes" | ? from onomatopoeia (homophonous with "container") |
| /goŋ-/ | "stick like" (pen, gun) | ? from /deɾgoŋ/ "river bank" |

**Table 4.** *(continued)*

| | | |
|---|---|---|
| *Semantic association origin* | | |
| /per-/ | "flat" (biscuit) | from /gpeɾ/ 'flat' |
| /saw-/ | "human" | from /saw/ 'body' |
| *Residue of unknown origin* | | |
| /soŋ-/ | "bamboo" | |
| /ma-/ | default | |
| /alaŋ-/ | "instance" | |

Some classifiers have cognates in other Bodo-Garo languages (Table 5). The Dimasa forms in Table 5 are semantically transparent, so the table may reflect parallel developments. Up to this point in time, classifiers have not been identified that have a shared origin that goes back further in time than Proto-Bodo-Garo, such as Proto-Sal or Proto-Tibeto-Burman.

**Table 5.** Sample of Bodo Garo cognate classifiers (Langthasa and Evans, 2018)

| **Dimasa** | **Bodo** | **Kokborok** | **Garo** | **Rabha** | |
|---|---|---|---|---|---|
| paŋ-si | pʰaŋ-se | pʰaŋ-sa | pʰaŋ-sa | pʰaŋ-sa | CLF:plant-one |
| saw-si | sa-se | kʰorok-sa | sak-sa | sak-sa | CLF:person-one |

Mensural classifiers indicate amounts and seem to form an open class which is often semantically transparent; e.g., /kbaw-/ "handful", /ktaj-/ "packet", /kep-/ "piece of", etc. They follow the same morphosyntactic patterns as sortal classifiers: /mkam kbaw-tam/ food CLF:handful-three "three handfuls of food".

There is no direct evidence as to why so many Tibeto-Burman languages of this area have the phenomenon of placing classifiers before numerals. However, it is noteworthy that Assam was ruled by the Tai speaking Ahom people from 1228 until the 1800's. Tai languages are known for their rich classifier systems and could be a source of classifiers in this area. The Tibeto-Burman languages that lack classifiers are all to the west of the Northeast India language area (Evans 2022). In the Aiton language (one of the Tai Ahom languages), classifiers occurred before the numeral "one" (6), although after the higher numerals, a pattern found throughout Tai languages. Perhaps this order, along with the concept of classifiers, was borrowed into the Northeast India language area during this era.

(6) Aiton Tai classifier before numeral (Morey 2005: 228)
*khɔt¹ luk³ wai³* ***phuu³ lɯŋ²*** *caaj²* *niu²/diu²*
leave son keep CLF one male only.one
"He left one son"

In terms of degree of contact, we note that there was formerly a Bodo-Garo language called Moran (Gurdon 1904), which was spoken in eastern or "upper" Assam, in a kingdom which bordered the Ahom kingdom for several centuries, during which the Moran speakers heavily assimilated Ahom language before their kingdom dissolved. Jacquesson and van Brueghel (2017) present evidence that Moran is an extinct dialect of Dimasa. Intense language contact between Ahom and Moran could have been a vehicle for the grammatical concept of classifiers to enter Bodo-Garo, and for the Tai constituent ordering CLF-one to become the standard morpheme order. Another argument for a Tai origin of classifiers in Tibeto-Burman languages of Assam is that the Assamese language, historically spoken further to the west than Moran and Ahom, is claimed to have developed classifiers under Tai influence as well (Emeneau 1956; Barz and Diller 1985). Further to the west, in Newar, classifiers are used with nouns in non-counted noun phrases, which is a feature of Tai languages, but not of Bodo-Garo (Hyslop 2008)

Texts from young adult speakers seem to show a similar range of classifier usage as do those from older speakers. Although Dimasa is under heavy linguistic pressure from Indo-Aryan languages and English, the rise of online oral, multimedia, and written literature in recent years may aid in preserving grammatical structures that might otherwise be lost in younger speakers.

Dimasa does not overtly distinguish mass/count quantifiers, such as "little/few" and "much/many" (7). Classifiers do not occur with these quantification strategies. Likewise, classifiers do not occur with demonstratives, possessives, etc.

(7) Marking of "little/few" and "much/many".
a. small amount of mass noun "work" marked with /kisa/
*kisa kusi daŋ-baj*
a.little work do-without.intent
"Do at least a little work."
b. small amount of count noun "house" marked with /kisa/
*iɾa=ha kisa no doŋ*
here=LOC a.little house EXST
"There are a few houses here."
c. Large amount of mass noun "spice" marked with /baŋ/
*mosla baŋ=jaba samlaj=zaŋ tajsa ɾi-pa-du*
spice be.much-REL dish=COM lemon give-APPL-PIPFV
"Lemon is served with dishes which have more spice."
(lit., "With dishes where the spice is much, lemon is also given")

d. Large amount of count noun "animal" marked with /baŋ/.
*ode hagra ha bo mi baŋ-bi*
and jungle LOC TOP animal be.much-STAT
"And in the jungle, there were many animals."

Many measure words occur in the same location as classifiers, as shown in (8). For measure words that are identifiable objects, such as "truck (full)", the measure word itself takes a classifier (9).

(8) Classifier (a) and measure word (b) show pre-numeral post-nominal order.
a. *tajlik* ***taj-si*** *laŋ-pa*
banana CLF:fruit-one take-COM
"Take a banana along with you."
b. *ada tajlik* ***kjoŋ-tam*** *labu-ba*
brother banana bunch-three bring-NF:IPFV
"Brother brought three bunches of bananas."

(9) measure word that takes its own classifier
*hazeŋ truck* ***ma-si***
sand truck CLF:default-one
"a truckload of sand"

In Dimasa, nouns can be counted using the combinations exemplified below. The default counted noun phrases have the structure N CLF-NUM or CLF-NUM N. Noun phrases with the order N CLF-NUM tend to be definite (10).

(10) N CLF-NUM (definite)
*yao* ***gong-gin*** *plaw-hi dung-ɾi=ha-ka*
hand CLF-two stretch-hi hot-CAUS=LOC-PRF
"(The sun) extended (his) two hands and made it hot."

For indefinite counted noun phrases CLF-NUM N is generally used (11). The interaction between classifier placement and (in-)definiteness marking is explored further in Section 3.

(11) CLF-NUM N (indefinite)
*DSU=ni* ***sao-si*** *dang-hoja ti-ba*
PN=GEN CLF-NUM work-leader say-NF:IPFV
"A leader of DSU said…"

Many counted noun phrases omit either the classifier or the noun. Round number approximations are expressed by reduplication of the number and elision of the classifier. The reduplicated NUM NUM occurs without a classifier and before a noun when the noun is being estimated (12). Similarly, NUM NUM occurs with comitative marking and no classifier when it has an adverbial sense, as in (13). In both examples, NUM NUM occurs in indefinite position, which seems to be typi-

cal for approximations. There are no examples in the corpus of smaller numbers being used in this way (e.g., "by pairs" or "by tens"), although mensural classifiers can be repeated for an adverbial sense (14).

(12) Use of NUM NUM N to express approximates
*bu* **ɾza** **ɾza** *msep* *kraŋ-ba*
3SG hundred hundred buffalo raise-NF:IPFV
"He reared hundreds and hundreds of buffalo."

(13) Use of NUM NUM COM for adverbial approximation
*mel=ha* ***ɾziŋ*** ***ɾziŋ*** *zaŋ* *subuŋ* *paj-ba*
meet=LOC thousand thousand COM person come-NF:IPFV
"People came to the meeting by the thousands."

(14) repetition of mensural classifier as adverb
*bu* *mkam kbaw* *kbaw* *zi-ka*
3sg rice handful handful eat-PRF
"He ate food by the handful."

Classifiers are also often omitted when counting humans (15). A similar pattern is found in Hungarian, in which none of the six classifiers is used when counting humans (Aikhenvald, forthcoming Chapter 3).

(15) Omission of classifier when counting humans ("six").
*mnaŋ* *gda=ha* *saw-si* *banzik-ni* *basa-tu* ***do*** *doŋ-ba*
before period=LOC CLF:human-one widow-GEN son-PL six EXST-NF:IPFV
*=ne*
=QUOT
"It is said that a long time ago a widow had six sons."

Counted events are expressed with classifier /pin-/ "times" and a numeral, but without a noun. The morpheme /pin-/ could be considered a verbal classifier "times" (16), and appears to come from the verb "return" (Table 4). Counted expressions often omit the noun when it is understood from context, as in "lemon" (17); this sentence also demonstrates the verbal classifier "three times".

(16) Verbal classifier "times"
*police-raw* *pin-tam* *pin-bri* *bu* *bosta=ke* *zik* *pa=ba*
police-PL CLF:times-three CLF:times-four DEM sack=ACC kick attach=NMLZ
*de* *bisiŋ* *nipraŋ* *garaŋ* *bajgo-ka*
TOP inside ABL voice come.out-PRF
"When the policemen kicked the sack a third or fourth time, then a voice came from inside,"

(17) Omission of counted noun ("lemon").
***kep-si*** *zaŋ ham-ja tika bo proŋ sajnzer sajnbli*
CLF:slice-one INS good-NEG if TOP morning noon evening
***pin-tam*** *liŋ-ka niŋ ham-du*
CLF:repeat-three drink-PRF TOP good- PIPFV
"If having one slice (of lemon) isn't good (enough), then drink it three times, morning, noon, and evening."

Combinations of CLF-"one" can become lexicalized. Subsequently, /baw-si/ "CLF: thought-one" has become lexicalized to mean "be the same", as in (18), where it functions as a predicate. In (19), /baw-si/ functions as an adverb "in the same way".

(18) lexicalized /baw-si/ "CLF:thought-one" to mean "the same"
*Rahul=zaŋ Arup zu-ba* ***bawsi***
PN=COM PN tall-NMLZ be.same
"Rahul and Arup have the same height."

(19) lexicalized /baw-si/ "CLF:thought-one" to mean "in the same way"
*saw=ha bemaɾ giri ɾi=ba-zaŋ* ***bawsi*** *bemaɾ*
body=LOC sickness INEXST CAUS=NMLZ-COM same.way sickness
*gba-gin=jaba ke bo tajsa mtaw-du*
spread-SPCF=REL acc top lemon stop-PIPFV
"Along with wiping out the diseases of the body, it also causes the diseases that can be contracted from another person to be stopped."

The generic classifier /ma-/ combines with "one" to form /ma-si/ which occurs in (20) as a predicate "be exactly the same".

(20) lexicalized /ma-si/ CLF:generic-one to mean "be exactly the same"
*Rahul=zaŋ Arup=ni daŋ-taj* ***ma-si***
PN=COM PN=GEN work-NMLZ be.exactly.same
"Rahul and Arup"s working behavior is exactly the same".

Non-lexicalized CLF-NUM can function as a predicate; e.g., when giving prices of things. Given the relatively low frequency of an expression like "fifty rupees", the predicate in (21) is not considered to be lexicalized.

(21) CLF-NUM as predicate
*homao* ***pher-shi****=ha* ***tho-dan***
rice.cake CLF:flat-one=LOC CLF:rupees-fifty
"One rice cake (costs) 50 rupees."

The examples above exemplify counted noun phrases in Dimasa, showing that classifiers occur only as bound morphemes prefixed to numerals. Depending on semantics and context, classifiers or counted nouns are sometimes elided. The fol-

lowing section shows the interaction between (in-)definiteness and the relative ordering of noun and CLF-NUM.

## 3. Classifier placement and (in-)definiteness

In many Indo-European languages, definiteness and indefiniteness are marked with special grammatical items; e.g., in English with *the* and *a*. However, in many Sino-Tibetan languages, (in-)definiteness is indicated by constituent order, rather than by overt lexical marking. For example, in Sinitic languages, noun phrases that occur before the verb tend to be definite, and those that follow the verb tend to be indefinite (Xu 1995; Li and Bisang 2012).

Among Bodo-Garo languages, definiteness has not been well described. Before looking at the interaction between classifier placement and (in-)definiteness, I would like to establish general principles of (in-)definite marking in Dimasa. First, proper nouns are definite, since they refer to an entity that is uniquely identifiable. For example, in (22) the protagonist (Veer Sambudhan Phonglo) is definite, as is "the village of Longkhor." The definiteness of the village is made additionally clear by the locative case marker /ha/. Example (23) gives a proper place name without locative marking, showing that definite place names also occur without locative marking.

(22) Proper nouns as definite.
***Veer Sambudhan Phonglo*** *26th February bsajn marbari 1850 majtaj=ha*
PN 26th February day Tuesday 1850 year=LOC
***Majbaŋ*** *rgoŋ=ha* ***Loŋkor nolaj***=*ha* *hazaj-ba*
PN near=LOC PN village=LOC born-NF:IPFV
"Veer Sambudhan Phonglo was born in the year 1850 on 26th February on Tuesday in the village of Longkhor near Maibang."

(23) Definite place name without locative marking (/hasaw razi/).
*gdema* *hagzer=ke* ***hasaw*** ***razi*** *ni* *Gandhi ti-pu-du*
respected.person clan.name= ACC hill.region society GEN PN say-POT-PIPFV
"Respected Hagjer can be called the Gandhi of Hasao district."

For similar reasons to proper nouns (e.g., identifiability), possessed nouns are also definite; e.g., "foreigners' leader" (24).

(24) Definite marking of /gdeba/ "leader" by possession
*bu* *ha* ***gpusa-raw*** ***ni*** ***gde=ba*** *za hi* *doŋ-ba* *Major Boyd*
DEM LOC foreigner-PL GEN big=NMLZ be SUCC EXST-NF:IPFV PN
"At that time, the head of the British was Major Boyd."

Case marking generally marks definiteness on non-subject nouns (differential case marking). In (25), although several place words are used, only "plains" is definite, marked with locative, while "paddy fields" is indefinite and does not take case marking.

(25) Definite ("plains") and indefinite ("paddy fields") locations.
***haplaj=ha hadi*** *daŋ-du maj maj-du*
plains=LOC paddy.field do-PIPFV rice get-PIPFV
"In the plains, paddy fields are worked and rice is obtained."

Accusative marking on objects is another example of case marking coinciding with definiteness. In (26) "jungle" is marked for definiteness both by possession ("there" GEN) and by accusative marking /ke/.

(26) Accusative marking of definite object.
*gbin hatan=ni subuŋ ura=ni* ***hagra=ke*** *nu-hi*
other place.of.dwelling=GEN person there=GEN jungle=ACC see-SUCC
*mzaŋ-zaw-re*
beautiful-PSV-SIPFV
"People from other places find the jungle there to be beautiful,"

On the other hand, direct objects that are indefinite do not take accusative marking, as in "groups" (27).

(27) Lack of accusative marking on indefinite direct object ("groups").
*odehe nolaj nolaj taŋ palaŋ-hi* ***dol*** *ka palaŋ-ba*
and village village go along-SUCC group form along-NF:IPFV
"And went along village by village forming groups along the way."

Topic marking also indicates definiteness. In (28), "executive member", marked by the topic marker /bo/ is definite. On the other hand, in (29), "Rajya Sabha member", which is indefinite, is not marked by case or topic marker.

(28) Topic marking of definite "Executive Member".
*gdema hagzer hasaw razi=ni district.council=ha*
respected.person clan.name hill.region society=GEN district.council=LOC
***executive-member bo*** *za-ka*
executive.member TOP be-PRF
"Respected Hagjer also became the Executive Member of the District Council"

(29) Lack of topic or case marking on indefinite "Rajya Sabha member".
*bu ni jahon ha 1958 nisiŋ 1962 so-si Delhi ni* ***Rajya.Sabha***
DEM GEN after LOC 1958 ABL 1962 reach-COND PN GEN Rajya.Sabha
***member*** *za-ba*
member be-NF:IPFV
"After that he also became a member of Delhi's Rajya Sabha from 1958 to 1962."

In addition to indicating definiteness by case and topic markers, Dimasa also has demonstratives which indicate definiteness. In (30), /bu grasa/ indicates "this/that old man". From the context of the story, "pig" and "land" are semantically definite, as both have been mentioned in the story before this point. Although they are semantically definite, they are not overtly marked for definiteness, a recurring issue crosslinguistically (Lyons 1999: 275).

(30) Use of demonstrative /bu/ to mark definiteness
***bu*** *grasa hono ha luŋ-ku luŋ-klaj min min klajba*
DEM old.man pig land root-upwards root-downwards ONP do-NOMZ
*nu-ka*
see-PRF
"The old man saw the pig harrowing up and down the earth."

The demonstrative marker can be combined with a plural morpheme to become a marker of definite plural. Examples (31a, b) demonstrate the use of demonstratives in an identificational use (Dixon 2003: 84). In each case, the noun referent ("drains", "needed items") is neither physically present, nor uniquely identifiable. The use of the demonstrative here is more identificational or recognitional than deictic. An example in English would be *I like <u>that kind</u> of dog you can cuddle with*. Note that DEM-PL is placed after its noun referent.

(31) DEM-PL /bu-tu/ as a marker of plural definite.
a. *drain* ***bu-tu****=ni di-lama giri*
drain PRX-DEM=GEN water-way INEXST
"The(se) drains' outlets are missing."
b. *samlajdik bawku-ma=ni skaŋ=ha naŋ-jaba* ***bu-tu****=ke bzom*
dish place-NMLZ=GEN before=LOC need-NMLZ DEM-PL=ACC gather
*klaj*
do
"Before placing the dish (on the stove) collect all the(se) needed items."

In many cases, demonstratives occur with case and topic marking (32), making for redundant marking of definiteness.

(32) Definiteness marked with both demonstrative and case marking ("time").
***bu*** *somoj=**ha la*** *school tanjaraw bo kom*
DEM time=LOC TOP school go-REL TOP less
"At that time, the school attendees were also fewer."

For nouns that are counted, Dimasa consistently uses the order CLF-NUM N for indefinite nouns (33) and the order N CLF-NUM for definite nouns (34).

(33) CLF-NUM N order of indefinite noun phrase.
*DSU=ni **sao-si*** *dang-hoja ti-ba*
PN=GEN CLF-NUM work-leader say-NF:IPFV
"A leader of DSU said…"

(34) N CLF-NUM order of definite noun phrase.
*yao **gong-gin** plaw-hi dung-ɾi-ha-ka*
hand CLF-two stretch-hi hot-CAUS-TRNS.LOC-PRF
"(The sun) extended (his) two hands and made it hot."

Dimasa postnominal classifier placement encodes definiteness in a variety of contexts, as exemplified in the following examples. When a participant in a narrative is introduced, the entity may be given as definite (35). This use of postnominal CLF-NUM marks identificational or recognitional meaning (Dixon 2003: 84) in the same way that the plural demonstrative marks it in (31). Definite marking in Dimasa texts seems to be the most common strategy for introducing characters in stories.

(35) Definite marking to introduce character in a narrative.
*bu somoj=ha subung **sao-si** hem-hi paj-ba*
DEM time=LOC person CLF-num walk-INFV come-NF:IPFV
"At that time one (DEF) person came walking."

On the other hand, introduction of an NP that does not recur in the story is likely to be indefinite, whether human (36) or not (37)

(36) Indefinite marking introducing a non-character.
*Delhi=niprang **sao-si*** *golf mlao=yaba paj-ba*
PN=ABL CLF-num golf play=REL come-NF:IPFV
"A golf player from Delhi came."

(37) Indefinite marking introducing a non-character
*bu goɾaj zaŋ maɾ dada aɾ **ma-sni** goɾaj*
DEM horse COM like NF and CLF-seven horse
*paj-pa-ba-ko=ne*
come-APPL-NF:IPFV-DPROG=QUOT
"And it is said that seven more horses came along with that horse."

Topicalization also cooccurs with syntactic definiteness (38). In (38a), the entity that is syntactically marked as definite is simultaneously claimed to be nonexistent. Nevertheless, "the strong one" is followed by the topic marker /bo/, which further emphasizes its definiteness. In (38b), there is no flower in the story. However, the topic marker /siniŋ/ "only" licenses definite word order in the preceding NP.

(38) Topic marking and definite marking of nonexistent entity
  a. *aŋ-laj* *rao=jaba* ***sao-si*** *bo* *giri*
  1SG-like strong= NMLZ CLF-NUM top INEXST
  "There is no one as strong as I."
  b. *kim* ***bar-si*** *sining* *sajn-ba-mu*
  flower CLF-one TOP:only ask-NF:IPFV-SBJV
  "(I) asked for only a flower." Lit., "Only asked for the flower."

As in many other languages, expected entities can take definite marking. For example, in (39), since airplanes have individuals who fly them, the previously unmentioned "pilot" /biɾ-jaba/ can take definite marking.

(39) Expected entity takes definite.
*bu* *singao=ha* *biɾ-jaba* ***sao-si***, *Longma-sa-ɾao=ni* *pdain=ha* *glai*
DEM airplane=LOC fly-NMLZ CLF-one PN-person-PL=GEN paddy=LOC fall
*paj-ba*
come-NF:IPFV
"The person flying in the airplane came falling down at Longma people's field."

Likewise, referents that are retrievable from context are marked definite, even if in translation indefinite semantics make more sense. In (40a), the farmer's son was injured while riding one of the seven horses mentioned above in (37). Although the horse mentioned could be any one of the seven horses, it is given as definite. (40b) shows how the sentence would be changed to encode indefinite "a horse".

(40) Retrievable referent takes definite.
  a. Original definite version from text
  *hagra=ni* *goraj* ***ma-si*** *=ha* *ga-hi* *graj-ba*
  forest=GEN horse clf-num =LOC ride-INFV fall-NF:IPFV
  "(He) fell down riding one of the wild horses." ("Riding on the wild horse, he fell.")
  b. Slightly altered version to yield indefinite.
  *hagra=ni* *goraj* *=ha* *ga-hi* *graj-ba*
  forest=GEN horse =LOC ride-INFV fall-NF:IPFV
  "(He) fell down riding a wild horse."

In some cases, it is not clear why noun phrases are given with formal definiteness. In (41a), "village" is marked definite, although the semantics seem to be indefinite.

It is possible that definiteness here is used to make the story more vivid. Likewise, (41b) shows the first use in the story of "sack" as definite, although "sack" doesn't meet any of the typical criteria for definite marking. Since (41(b)) comes from a folk tale, the speaker may have assumed accessibility of "sacks" on the part of the listener, hence licensing definiteness (the sacks play an important role in the rest of the story).

(41) Formal definiteness with semantic indefiniteness.
a. *nolaj* ***ma-si***=*ha* *grasa* *grajik* *doŋ-ba.*
village CLF-one=LOC old.man old.woman EXST-NF:IPFV
"In one village there lived an old man and an old woman."
b. *gini* *bo* *bosta* ***ma-si*** ***ma-si*** *la-ka*
CLF.two TOP sack CLF:one CLF:one take-PRF
"Both took one sack each."

To summarize definite marking in Dimasa, we find that postnominal placement of CLF-NUM corresponds with definiteness. Although Dimasa lacks a definite article, definite marking is still an important component of texts. In some cases, definite marking may serve to make the entity more vivid, or to bring an entity into the linguistic foreground. Definiteness in counted noun phrases is indicated by a postnominal classifier-numeral compound. In non-counted noun phrases (and some counted noun phrases) definiteness can be marked by demonstratives, topic marking, case marking, or a combination of these.

Counted noun phrases can be overtly marked as indefinite. For example, in (42), the items listed occur in a classic indefinite scenario, in which neither the news reporter nor the listeners are likely to be aware of the exact physical entities being referenced in the list of contraband. The order CLF-NUM N encodes indefinite status.

(42) Classsic indefinite counted nouns
*Operation klaj-jarao dol* ***ma-si*** *AK.47* ***ma-bri*** *magazine odehe* ***bisa-tam***
operation do-REL group CLF-NUM AK.47 CLF-four magazine and 20–3
***ma-bri*** *gulitaj* ***ma-si*** *AR mai-ba.*
CLF-four bullet CLF-one ar get-NF:IPFV
"The group doing the operation found one AK 47, four magazines and 64 bullets, (and) one automatic rifle."

Another stereotypical use of indefinite occurs in introducing a new participant in a story. Although such noun phrases can be definite (35), they can also occur as indefinite (43).

(43) Indefinite introduction of story character
*mining gda=ha* ***sao-si*** *hadi daŋ-jasa doŋ-ba =ne*
before age=LOC CLF-one paddy work-AGT-person EXST-NF:IPFV QUOT
"A long time ago, there was a farmer."

Proper nouns are a classic category of definite nouns. However, indefiniteness can be imposed on a proper noun, much like the use of "a certain __" in English (44; names have been removed from the corpus and replaced with "PN" in cases where embarrassment might occur).

(44) Formal indefiniteness of a proper noun.
*ulaj-ba=ha* ***sao-si*** *PN=ke police maj-ka*
DEM-NMLZ=LOC CLF-one PN=ACC police get-PRF
"At that time, police found a certain PN."

On the other hand, CLF-one is used with well-known individuals' names in a definite construction (45).

(45) Formal definiteness of known entity.
*JB Hagjer* ***sao-si*** *Dimasa=ni hoja dong-ba.*
PN CLF-one Dimasa=GEN leader EXST-NF:IPFV
"The one JB Hagjer was a Dimasa leader."

To summarize indefinite marking, pre-posing the CLF-NUM word before the counted noun is a marker of indefiniteness in Dimasa. Most instances of syntactic indefinite correspond to semantic indefiniteness. In some cases, such as introducing an entity into a story, speakers exercise choice as to whether to mark the noun as definite or indefinite.

Grammatical descriptions of other Bodo-Garo languages also note that CLF-NUM can occur either before or after the counted noun, although the connection with (in-)definiteness does not seem to have been specified in earlier works. It is commonly reported that N CLF-NUM occurs with a higher text frequency than CLF-NUM N. The tendency has been noted in Rabha (Joseph 2007), Bodo (Basumatary 2015), Garo (Burling 1961), Kokborok (Debbarma 2015), and Atong (van Breugel 2014).

## 4. Summary

There are two issues that Dimasa classifiers bring to the fore. First, Dimasa, along with many other Tibeto-Burman languages of the Northeast India language area, places classifiers before numerals in counted noun phrases. This order is extremely rare in the world's languages; outside this area it has been documented

in Warekena (Aikhenvald 2000: 109) and Chimila (Malone 2004; Aikhenvald forthcoming, Chapter 3). As in other Bodo-Garo languages, classifiers are bound morphemes that only occur prefixed to numbers. Most classifiers have transparent origins in (nearly) homophonous nouns. Despite the typologically highly unusual nature of the ordering CLF-NUM, it has spread throughout several Tibeto-Burman language branches in this area. The origin of this word order remains a mystery. It does not appear to be related to AOV/SV constituent order for two reasons. First, almost all Tibeto-Burman languages have AOV/SV constituent order. Outside of this language area, Tibeto-Burman languages with classifiers arrange them after numerals. Second, outside of the Tibeto-Burman family, other languages with AOV/SV order also lack the sequence CLF NUM. For example, Japanese (AOV/SV), Thai (AVO/SV or free constituent order, depending on the analysis) both present the pattern N NUM CLF. It is possible that the concept of classifiers was borrowed from Tai Ahom during the period of Tai dominance in Assam from the medieval period until the nineteenth century. Tai languages pre-pose classifiers to the numeral "one", which may have led to the order CLF-NUM in the languages of this area that borrowed the classifier concept.

The second theoretical issue relates to the marking of (in)definiteness by pre-posing (indefinite) or post-posing (definite) CLF-NUM relative to the counted noun. Whether counted or not, definite nouns can be marked by demonstratives, case marking, topic marking or a combination of these. Indefinite nouns that are not counted have no special marking.

It is hoped that analysis of classifier placement in other languages of Northeast India will reveal additional semantic properties of classifier placement, and perhaps add clarification to the origin of classifier-numeral order.

## Funding

All remaining errors are my own. Support for the research was generously provided by the "Science and Technology Council grant #108-2410-H-001-048-MY3".

## Acknowledgements

The author wishes to acknowledge the kind and thoughtful contributions of Alexandra Aikhenvald, Shobhana Chelliah, Benjamin Hull, Dhrubajit Langthasa, Jackson Sun, Dorian Roehrs, William Salmon, and audiences at the Institute of Linguistics, Academia Sinica and the University of North Texas.

## Abbreviations

The following non-standard abbreviations have been used in this paper. Abbreviations that appear in the Leipzig glossing rules are not listed.

| | | | |
|---|---|---|---|
| DPROG | durative progressive | POT | potential |
| EXST | existential | PRX | proximal |
| INEXST | inexistential | PSV | passive |
| INFV | infinitive | QUOT | quotative |
| N | noun | SIPFV | split imperfective |
| NF:IPFV | nonfuture imperfective | STAT | stative |
| NUM | numeral | SUCC | successive |
| ONP | onomatopoeia | TOP | topic |
| PIPFV | past imperfective | TRNS.LOC | translocative |
| PN | proper noun | | |

## References


Aikhenvald, Alexandra Y. (1994). Classifiers in Tariana. *Anthropological Linguistics*, 407–465.

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

Aikhenvald, Alexandra Y. (2021). One of a kind: on the utility of specific classifiers. *Cognitive Semantics 7*(2), 232–257. https://doi.org/10.1163/23526416-07020001

Aikhenvald, Alexandra. Y. Forthcoming. *A guide to gender and classifiers*. Oxford University Press

Barz, R.K., and Diller, A.V. (1985). Classifiers and standardisation: some South and South-East Asian comparisons. *Papers in Southeast Asian Linguistics No. 9: Language policy, language planning and sociolinguistics in South-East Asia*. Pacific Linguistics.

Basumatary, Guddu P. (2015). Numeral classifiers in Bodo. *Nepalese Linguistics*, 30,19–24.

Burling, Robbins. (1961). *A Garo Grammar*. Deccan College Postgraduate and Research Institute.

Van Breugel, Seino. (2014). *A grammar of Atong*. Brill. https://doi.org/10.1163/9789004258938

Debbarma, Biman. (2015). Word order in Kokborok. *Language in India*, 15(5). 51–58.

DeLancey, Scott. (2011). "Optional" "ergativity" in Tibeto-Burman languages. *Linguistics of the Tibeto-Burman Area*, *34*(2), 9–20.

Dixon, Robert M.W. (2003). Demonstratives: A cross-linguistic typology. *Studies in Language 27*(1), 61–112. https://doi.org/10.1075/sl.27.1.04dix

Emeneau, Murray B. (1956). India as a lingustic area. *Language* 2 (1), 3–16. https://doi.org/10.2307/410649

Evans, Jonathan P. (2022). Classifiers before numerals: Origin of an atypical morpheme order. *Linguistics of the Tibeto-Burman Area*, 45(1), 1–21. https://doi.org/10.1075/ltba.21013.eva

Evans, Jonathan P., and Langthasa, Dhrubajit. (2023). *Dimasa language: structure and texts*. Language and Linguistics.

Gurdon, P.R.T. (1904). The Morans. *Journal of the Asiatic Society of Bengal* 73, 36–48.

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

Hyslop, Gwendolyn. (2008). Newar Classifiers: A summary of the literature. *Newah Vijaanan* (Journal of Newar Studies) 6, 28–41.
Jacquesson, F. and van Breugel, S. (2017). The linguistic reconstruction of the past: The case of the Boro-Garo languages. *Linguistics of the Tibeto-Burman Area*, *40*(1), 90–122. https://doi.org/10.1075/ltba.40.1.04van
Joseph, Umbavu V. (2007). *Rabha*. Brill.
Li, Xuping, and Bisang, Walter. (2012). Classifiers in Sinitic languages: From individuation to definiteness-marking. *Lingua*, 122(4), 335–355. https://doi.org/10.1016/j.lingua.2011.12.002
Langthasa, Dhrubajit, and Evans, Jonathan P. (2018). Numeral Classifier and Word Order in Dimasa and Bodo-Garo. In *Proceedings of the 51st International Conference on Sino-Tibetan Languages and Linguistics*.
Lyons, Christopher. (1999). *Definiteness*. Cambridge University Press. https://doi.org/10.1017/CBO9780511605789
Malone, Terry. (2004). Classifiers in Chimila (Chibchan). *Nominal classification, Special issue of Sprachtypologie und Universalienforschung 57*(2/3), 117–143. https://doi.org/10.1524/stuf.2004.57.23.144
Morey, S. (2005). *The Tai languages of Assam--a grammar and texts*. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
Xu, Liejiong. (1995). Definiteness effects on Chinese word order. *Cahiers de linguistique-Asie orientale*, *24*(1), 29–48. https://doi.org/10.3406/clao.1995.1465

## Address for correspondence

Jonathan P. Evans
Academia Sinica Humanities and Social Sciences Building, 南棟5樓
Institute of Linguistics, Academia Sinica
Section 2, Academia Rd, No. 128
11529 Taipei City, Nangang District
Taiwan
jonathan@sinica.edu.tw

## Publication history

Date received: 3 May 2022
Date accepted: 28 September 2022