

慶應義塾
外国語教育研究
別冊（論集）第1号

JOURNAL
of
FOREIGN LANGUAGE EDUCATION

Special Issue Vol. 1



慶應義塾大学外国語教育研究センター

KEIO RESEARCH CENTER FOR FOREIGN LANGUAGE EDUCATION

Special Issue: Classifying Nominalization in Cross-linguistic Perspective I

Table of contents

Kazuhiro Kawachi, Prashant Pardeshi, Yu Li, and Maho Naito

Classifying nominalization in cross-linguistic perspective: Introduction to special issue I
1–19

1. Masayoshi Shibatani and Marília Ferreira
Dynamic functional typology of Tukanoan classifying nominalizations
20–100
2. Masaaki Shimizu and Trang Phan
Nominalization in Vietnamese: Roles of classifiers and the genitive marker *của*
101–117
3. Satoko Shirai
Classifying nominalizers in Northern nDrapa
118–128
4. Yu Li and Miyagishi Tetsuya
Nominalization in Guoli Zauzou
129–150
5. Tetsuya Miyagishi and Yu Li
Grammatical nominalization in the Jiangmo dialect of Zauzou
151–181
6. Yoshihisa Taguchi
Classifiers and nominalization in Lan Hmyo
182–192
7. Fang Wu and Hongyong Liu
Classifiers as nominalizers in Gouliang Miao
193–208
8. Tatjana Ilic and Kazuhiro Kawachi
So-called gender markers as classifying nominalizers in Swedish
209–232
9. Prashant Pardeshi and Andrej Bekeš
Distribution of gender marking in Slovene
233–252

10. Shuichiro Nakao
Gender marking and nominalization in Bari
253–274
11. Makoto Furumoto
Noun class system and grammatical nominalisations in the Kimakunduchi dialect of
Swahili
275–292
12. Kazuhiro Kawachi
Gender marking on grammatical nominalizations in Sidaama
293–335
13. Gitanjali Bez
Classifier marking in Assamese
336–346
14. Manasi Kelkar and Prashant Pardeshi
Distribution of classifier and gender marking in Southern Marathi as compared to Pune
Marathi
347–372
15. Tetsuya Miyagishi
Nominalizations with classifiers and gender markers in Sinhala: Usage patterns from
online sources
373–414
16. Prashant Pardeshi and Govind Mopkar
Distribution of classifier and gender marking in Konkani
415–445
17. Niranjan Uppoor
Gender and classifier marking in Dravidian languages
446–502

Classifying nominalization in cross-linguistic perspective: Introduction to special issue I

Kazuhiro Kawachi¹, Prashant Pardeshi², Yu Li³, and Maho Naito⁴

1. INTRODUCTION

This special issue presents a collection of articles reporting on the use of classifying nominalizers (strictly speaking, classifying grammatical nominalization markers)—especially so-called classifiers and gender markers—in deriving grammatical nominalizations in various languages, as framed within the nominalization theory proposed in Shibatani (2009, 2017, 2018a, 2018b, 2019, 2021, 2023; Shibatani and Marilia this issue). Each article examines whether classifying nominalizers are used obligatorily, optionally, or not at all with different types of structures to derive grammatical nominalizations functioning as the head of a noun phrase or as a nominal modifier, and tests one of Shibatani’s hypotheses concerning their distribution across these structure types. According to Shibatani’s (2021, 2023) hypothesis, which specifically concerns classifying nominalizers including classifiers and gender markers, these markers are more likely to occur obligatorily with structures that rank higher on the implicational hierarchy ‘NUM ((cardinal) numerals) > DEM (demonstratives) > GEN (genitive noun phrases) > ADJ (adjectives) > V-based NMLZ (verb-based nominalizations such as relative clauses)’ than those lower on the scale, while their function remains the same,⁵ and are more likely to occur with a given type of structure when it is used as the head of a noun phrase (noun-phrase use, or NP use) than when it is used to modify a nominal (modification use). While this hypothesis has diachronic implications for the development of the use of classifying nominalizers (Shibatani and Shigeno 2013), most of the articles in this issue focus primarily on the synchronic distribution of these marking patterns.

This issue comprises two comparative studies—one on Dravidian languages in India and the other on Tukanoan languages in South America—and fifteen investigations of individual languages: four in Africa, nine in Asia, and two in Europe. For two of the nine Asian languages (Marathi and Zauzou), two dialects were investigated.

This collection makes a meaningful contribution to typological research on nominal classification. It fills gaps in data concerning the marking of different types of structures with

¹ Faculty of Business and Commerce, Keio University, 4-1-1 Hiyoshi, Koohoku-ku, Yokohama, Kanagawa 223-8521, Japan. email: kawachi@keio.jp

² Research Department, National Institute for Japanese Language and Linguistics, 10-2 Midoricho, Tachikawa, Tokyo 190-8561, Japan.

³ School of Chinese Language and Literature, Wuhan University, A218 Zhenhua Building, Wuhan, Hubei, China.

⁴ Department of Linguistics, Graduate School/Faculty of Arts and Letters, Tohoku University, 27-1 Aoba-ku, Kawauchi, Sendai, Miyagi 980-8576, Japan.

⁵ In this hierarchy, numerals are restricted to cardinal numerals, and Shibatani does not include ordinal numerals in it. Nevertheless, it would be interesting to see how ordinal numerals are marked with classifying nominalizers, similar to or differently from cardinal numerals and other types of structures. Some of the articles include data on ordinal numerals as well. As for verb-based nominalizations, each article presents data on relative clauses at least, and some articles include data also on complement clauses and other language-specific verb-based nominalizations such as verb participles.

classifying nominalizers, which serve to nominalize the structures and classify their denotations. It also enables language-internal and cross-linguistic comparisons within a unified framework, namely, Shibatani's Dynamic Functional Typology. The significance of these contributions is evident in the following respects. First, the reports on modern languages, which are primarily based on speaker elicitations, provide information on whether the use of classifying nominalizers is obligatory, optional, or disallowed—information that is often difficult to obtain from the existing literature. Second, all reports include data on the NP use of grammatical nominalizations, which are rarely documented even in comprehensive reference grammars. Third, many of the articles describe classifying nominalizers in relatively understudied languages, thereby expanding the empirical foundation of typological research. Fourth, some reports examine languages that possess more than one type of classifying nominalizer—whether different types of classifiers, different types of gender markers, or both—enabling systematic comparisons of their distribution and function within the same language. Fifth, because all the articles adopt a common analytical framework, they facilitate consistent cross-linguistic comparisons of the use of classifying nominalizers with different types of structures.

Before moving to the next section, a word of caution about terminology is necessary. Shibatani (also Shibatani and Marília this issue) argues that so-called classifiers and gender markers do not constitute discrete categories, traditionally defined in terms of referent categorization for counting (e.g., Greenberg 1974, Allan 1977) and agreement with a head noun (e.g., Corbett 1991), respectively. Instead, classifier marking and gender marking can be treated as a unified phenomenon forming a continuum, with gender markers being more grammaticalized than classifiers (Craig 1986, Grinevald 2000, 2002, Grinevald and Seifart 2004).⁶ First, the use of classifiers is not limited to numerals but often extends to structures such as demonstratives and genitive noun phrases, and even to adjectives and verb-based nominalizations (e.g., Guoli Zauzou in Li and Miyagishi this issue, Jiangmo Zauzou in Miyagishi and Li this issue, Assamese in Bez this issue). Second, numerals are not exclusively associated with classifiers; in many languages with grammatical gender, they are also typically accompanied by gender markers. Third, in some languages, gender is not expressed through agreement at all. Fourth, what appears to be agreement frequently emerges across noun phrase components involving classifiers. Finally, there are languages whose classifying nominalizers cannot clearly be categorized as either classifiers or gender markers. For example, as shown by Shibatani and Marília (this issue), and by François (2017: 322–325) in his discussion of “gender classifiers” in Hiw, an Oceanic language of northern Vanuatu (ISO 639-3: hiw), some markers express humanity, gender, and number simultaneously. Another example is Swedish, which has traditionally been characterized as having gender. It distinguishes between common and neuter, but common gender results from the merger of the previously distinct feminine and masculine genders, and animates tend to be categorized as common. Thus, the Swedish system is more like an animacy-based classifier system. We follow the view that certain classifying nominalizers defy clear classification as either classifiers or gender markers. Nevertheless,

⁶ So-called classifiers and gender markers also each differ in the degree of their grammaticalization. “Classifiers” can function as nominals or can form a closed class. “Gender markers” may be based on biological gender and apply only to human or animate referents or may be extended to inanimate referents.

most of the articles in the present issue continue to use the terms ‘classifiers’ and ‘gender markers’ as traditionally employed, recognizing them as useful descriptive labels. These could alternatively be replaced by more specific terms, such as ‘shape-based or animacy-/humanhood-based classifying nominalizers’ and ‘biological gender-based classifying nominalizers’, or simply ‘animate classifiers’ (e.g., animate/inanimate, human/non-human, female/male) and ‘inanimate classifiers’ (e.g., shape, size, consistency, arrangement, function), respectively. Thus, in this issue, the terms ‘classifiers’ and ‘gender markers’, when used, should be understood as implicitly prefaced by ‘so-called’, even when this qualifier is not explicitly stated.

2. BACKGROUND: OVERVIEW OF SHIBATANI’S NOMINALIZATION THEORY AND HIS HYPOTHESES

Traditionally, the term ‘nominalization’ has been used in two senses. In one sense, it refers to a morphological process of derivational affixation, through which nouns are formed from words in other parts of speech, especially verbs (e.g., *sing* → *singer*). In the other sense, it refers to a syntactic process involving the use of a form that combines with an element that may or may not contain a noun to constitute a noun phrase (e.g., *the enemy’s destruction of the city*).⁷ Both processes involve verb-based or predicate-based nominalization.

By contrast, Shibatani’s nominalization theory distinguishes between two types of nominalization: verb-based and nominal-based. This theory not only takes nominal-based nominalization into account but also places particular emphasis on it. Table 1 summarizes Shibatani’s conception of nominalization with English examples. According to him, whether verb-based or nominal-based, nominal structures derived through the process of nominalization—can serve two functions: NP use, in which a nominalization serves as the head of a noun phrase (e.g., (1)), and modification use (Mod use), in which a nominalization modifies a nominal (e.g., (2)).⁸ (Both verb-based and nominal-based nominalizations in NP use are characterized as ‘noun phrases without nouns’ by Dryer 2004.⁹) Nominalizations in either use have denotations similar to nouns. The denotation of a nominalization (e.g., ‘four entities’) is metonymically associated with that of its base structure (e.g., ‘four’ as the denotation of *yi*³¹), and in modification use, it restricts the denotation of the nominals that it modifies (e.g., *nia*⁵³*ni*³³ ‘cats’).

Guoli Zauzou

- (1) *ŋu*⁵⁵ *yi*³¹ *ko*³³ *na*³³
 1SG four CLF:animal like
 ‘I want four (animals).’ (adapted from Li and Miyagishi this issue)

⁷ The distinction between the two processes is not clear-cut, as the syntactic process can also involve the morphological process (e.g., *destruct* → *destruction* in *the enemy’s destruction of the city*).

⁸ See Dryer (2004) for his arguments against the traditional analyses that assume the existence of a head in a noun phrase.

⁹ ‘Nouns’ in ‘noun phrases without nouns’ refers to so-called ‘head nouns’, which do not exist, for example, in a genitive noun phrase in noun-phrase use (e.g., *John’s*).

Guoli Zauzou

- (2) ɲu⁵⁵ nia⁵³ni³³ yi³¹ ko³³ na³³
 1SG cat four CLF:animal like
 ‘I want four cats.’ (Li and Miyagishi this issue)

Table 1. Illustration of Shibatani’s conception of nominalization with English examples (based on Shibatani 2019: 39)

Type/subtype of nominalization			Noun-phrase use	Modification use
Verb-based	Lexical nominalization	Event nominalization	<i>(the) [flight] (lasted three hours), (the) [employment] (of Bill)</i>	<i>[flight] schedule</i>
		Argument nominalization	<i>(a) [cook], (a) [judge], (an) [employer], (an) [employee]</i>	<i>[employee] benefits</i>
	Grammatical nominalization	Event nominalization	<i>[[<i>(that) John employed Bill</i>]_{NMLZ}]_{NP}, [[<i>(John('s) employing Bill</i>]_{NMLZ}]_{NP}, [[<i>to employ Bill</i>]_{NMLZ}]_{NP}</i>	<i>[the fact [<i>(that) John employed Bill</i>]_{NMLZ}]_{NP}</i>
		Argument nominalization	<i>[[<i>what broke John</i>]_{NMLZ}]_{NP}, [[<i>what John broke</i>]_{NMLZ}]_{NP}</i>	<i>[[<i>(the one) [who employs Bill</i>]_{NMLZ}]_{NP}, [[<i>(the one) [who(m) John employs</i>]_{NMLZ}]_{NP}, [[<i>(the one) [which broke John</i>]_{NMLZ}]_{NP}, [[<i>(the one) [which John broke</i>]_{NMLZ}]_{NP}</i>
Nominal-based	Lexical nominalization		<i>[childhood]</i>	<i>[childhood] memories</i>
	Grammatical nominalization		<i>I bought [[<i>three</i>]_{NMLZ}]_{NP}/ [[<i>those</i>]_{NMLZ}]_{NP}/ [[<i>those three</i>]_{NMLZ}]_{NP}</i>	<i>I bought [[<i>three</i>]_{NMLZ} cars]_{NP} / [[<i>those</i>]_{NMLZ} cars]_{NP} / [[<i>those</i>]_{NMLZ} [<i>three</i>]_{NMLZ} cars]_{NP}</i>
			<i>[John's/His]_{NMLZ}]_{NP} (is ...)</i>	<i>[[<i>John's</i>]_{NMLZ} book]_{NP}</i>
			<i>[cotton]_{NMLZ}]_{NP}</i>	<i>[[<i>cotton</i>]_{NMLZ} shirt]_{NP}</i>

There are two types of verb-based nominalization: lexical and grammatical. Lexical verb-based nominalization converts a non-nominal element—such as a verb or an adjective—into a noun, whereas grammatical verb-based nominalization typically produces a complex nominal structure headed by a nominalized verb form, possibly accompanied by its argument nominals.

Both lexical and grammatical verb-based nominalization can be further divided into event and argument types. Event nominalization derives a nominal denoting an event, or related

notions such as a fact or proposition metonymically associated with that event. Argument nominalization, by contrast, produces a nominal denoting one of the verb’s arguments or other entities metonymically linked to the verb’s meaning.

Nominal-based nominalization can also be lexical or grammatical. Lexical nominal-based nominalizations are nouns formed from other nouns. Grammatical nominal-based nominalizations are nominals derived from base structures such as numerals, demonstratives, and genitive noun phrases, with their denotations coming from these structures.

Shibatani’s theory focuses primarily on grammatical nominalizations, encompassing both nominal-based and verb-based types. He proposes two main hypotheses. The first posits that nominalization markers develop diachronically from nominal-based to verb-based nominalizations, and from NP use to modification use. This argument is supported by the observation that there are more attested marking patterns for nominal-based nominalizations than for verb-based ones, and more for NP use than for modification use. In contrast to traditional analyses, which treat grammatical nominalizations in modification use as basic and those in NP use as derivative (often through omission of a head noun), Shibatani’s analysis begins with NP use.

Table 2. Shibatani’s (2023: 21) typology of nominalization

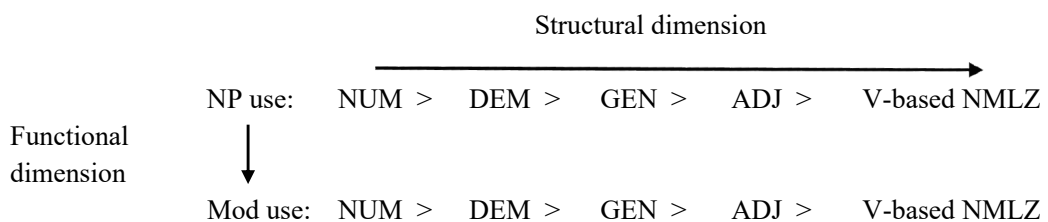
Nominalization	Non-classifying	No marking	
		Marking	Verbal/nominal inflection
			Derivational affix
			“conjunction”, “relativizer”, “subordinator”
			Adnominal particle
			Suprasegmental feature
	Classifying	No marking	
		Marking	Less grammaticalized (aka “(numeral classifier system”)
			More grammaticalized (aka “gender system”, “noun-class system”)

Shibatani (2021, 2023) recently developed an additional hypothesis concerning classifying nominalizers. He distinguishes between two types of nominalization—non-classifying and classifying—as shown in Table 2. Classifying nominalizers include classifiers and grammatical gender markers, encompassing noun-class markers. Like Craig (1986), Grinevald (2000, 2002), and Grinevald and Seifart (2004), Shibatani views these markers as forming a continuum—from less grammaticalized classifiers, which are semantically associated with the properties of noun-phrase referents, to more grammaticalized genders, which are largely conventionalized with individual lexical items (aside from their association

with biological gender), with noun classes occupying an intermediate position. Shibatani further explores this continuum in the context of nominalization, observing that classifiers and gender markers exhibit similar cross-linguistic patterns in marking structures in order to nominalize them. For illustrative examples from both classifier-marking and gender-marking languages that demonstrate this parallelism, see Shibatani (2023) and Shibatani and Ferreira (this issue). Classifying nominalizers nominalize a base structure to derive a nominalization and categorize its denotation according to the language’s classification system (e.g., ‘three feminine-class entities’). (By contrast, non-classifying grammatical nominalization markers only nominalize a base structure to derive a nominalization without categorizing its denotation.)

Using the implicational hierarchy shown in Figure 1 (the Shibatani hierarchy), Shibatani proposes a hypothesis about the synchronic distribution of classifying nominalizers: (i) along the structural dimension, classifying nominalizers are more likely to occur obligatorily with structures that rank higher on the implicational hierarchy than with those lower on the hierarchy, while their function remains the same; and (ii) along the functional dimension, they are more likely to occur obligatorily with a given type of structure when it serves as the head of a noun phrase than when it is used to modify a nominal (modification use).

Figure 1. Hierarchy constraining classifying grammatical nominalizer marking (adapted from Shibatani 2023: 35)



Shibatani also hypothesizes about the diachronic development of classifying nominalizers: their use extends from numerals downward (i.e., rightward in the figure) to other types of structures along the structural dimension, and from NP use to modification use along the functional dimension.¹⁰ This hypothesis is based on the observation that classifying nominalizers are more commonly and obligatorily applied to structures higher on the hierarchy than to those lower down, and that they are more likely to be obligatorily used with a given structure in NP use than in modification use.

Thus, this hypothesis not only addresses the synchronic distribution of classifying-nominalizer marking along the implicational hierarchy, but also carries diachronic implications for how such marking develops across both the structural and functional dimensions of the hierarchy. Notably, the proposed patterns of development are expected to apply not only to so-called numeral classifiers, which are prototypically used with numerals, but also to other types of markers, whether traditionally categorized as classifiers or gender markers.

¹⁰ The diachronic development of nominalization marking using non-classifying nominalizers—from noun-phrase use to modification use—was proposed earlier by Shibatani and Shigeno (2013).

Shibatani's (2021, 2023) hypothesis is grounded in his analysis of data from a wide range of languages. However, only a small number of in-depth studies on grammatical nominalizations in individual languages have so far been conducted within his framework (e.g., Shirai 2022, Pardeshi and Shibatani 2023). To evaluate this hypothesis more fully, it must be tested against data from a broader range of languages. To this end, the articles in the present issue present data from both individual languages and well-defined language families, examining which types of structures are marked with classifying nominalizers and whether such marking is obligatory or optional. All contributions provide synchronic analyses of marking patterns involving classifying nominalizers, with some also addressing their diachronic development. Diachronic analysis is particularly important, as Shibatani's hypothesis concerns not only synchronic marking patterns but also their historical development. For such investigation, comparisons among dialects and closely related languages are especially valuable—a perspective adopted by several of the contributions, Shibatani and Ferriera *inter alia*.

3. LANGUAGES DOCUMENTED AND CONTRIBUTORS IN THE PRESENT ISSUE

Table 3 summarizes the contents of this issue, listing the names of the languages (or language families) or dialects documented in this issue, along with their genetic and geographical information, ISO 639-3 codes, and the names of the contributors. Article 1 investigates a language family in which the traditional typological distinction between classifiers and gender markers is called into question. Articles 2–7 and 13 describe languages traditionally identified as having classifiers. Articles 8–14 report on languages with what are commonly termed gender markers. Articles 15–18 examine languages that possess both classifiers and gender markers. Table 3 lists the names of the languages (or language families) or dialects documented in this issue, along with their genetic and geographical information, ISO 639-3 codes, and the names of the contributors.

Table 3. Languages documented in the present issue and their contributors

Article	Language, language family, dialect	Country/countries	ISO 639-3	Contributor(s)
1	Tukanoan languages	Colombia, Brazil, Ecuador, and Peru	Tuyuka: tue, Tukano: tuo, Kotiria: gvc, Retuarã: tnc, Tanimuka: tnc	Masayoshi Shibatani and Marília Ferreira
2	Vietnamese (Austroasiatic)	Vietnam	vie	Masaaki Shimizu and Trang Phan
3	Northern nDrapa (Qiangic, Tibeto-Burman)	China	zhb	Satoko Shirai
4	Guoli dialect of Zauzou (Ngwi, Tibeto-Burman)	China	zal	Yu Li and Tetsuya Miyagishi

Table 3. Languages documented in the present issue and their contributors (continued)

5	Jiangmo dialect of Zauzou (Ngwi, Tibeto-Burman)	China	zal	Tetsuya Miyagishi and Yu Li
6	Lan Hmyo (Hmongic/Miao)	China	hml	Yoshihisa Taguchi
7	Gouliang Miao (a sub-dialect of Xiangxi Miao) (Hmongic/Miao)	China	muq	Fang Wu and Hongyong Liu
8	Swedish (North Germanic)	Sweden and Finland	swe	Tatjana Ilic and Kazuhiro Kawachi
9	Slovene (South Slavic)	Slovenia	slv	Prashant Pardeshi and Andrej Bekeš
10	Bari (Eastern Nilotic)	South Sudan	bfa	Shuichiro Nakao
11	Kimakunduchi dialect of Swahili (Bantu, Niger-Congo)	Tanzania	swa	Makoto Furumoto
12	Sidaama (Highland East Cushitic)	Ethiopia	sid	Kazuhiro Kawachi
13	Assamese (Indo-Aryan)	India	asm	Gitanjali Bez
14	Southern Marathi and Pune Marathi (Indo-Aryan)	India	mar	Manasi Kelkar and Prashant Pardeshi
15	Sinhala (Indo-Aryan)	Sri Lanka	sin	Tetsuya Miyagishi
16	Konkani (Indo-Aryan)	India	kok	Prashant Pardeshi and Govind Mopkar
17	Dravidian languages	India (also, Sri Lanka and Pakistan)	Arebase (Arebhashe): arb, Kannada: kan, Beary: bnw, Malayalam: mal, Koraga: kfd/vmd, Tamil: tam, Tulu: tcy, Telugu: tel	Niranjan Uppoor

4. FINDINGS REGARDING SHIBATANI'S HYPOTHESIS ABOUT CLASSIFYING NOMINALIZERS

4.1 Summary of the findings of each report

This subsection briefly summarizes the findings of each contribution with respect to Shibatani's hypothesis on the use of classifying nominalizers to derive grammatical nominalizations. As the summaries presented here are necessarily concise and somewhat generalized, readers should consult the individual articles for more detailed information.¹¹

1. According to Shibatani and Ferreira, in Tukanoan languages like Tuyuka, the obligatory use of classifying nominalizers applies to grammatical nominalizations in their NP use, regardless of any semantic or syntactic factor. In contrast, for grammatical nominalizations in their modification use, classifying nominalizers are obligatory only when the referent is animate, or when it is inanimate and the marker appears postnominally. When the referent is inanimate and the marker is prenominal, its use is optional. The marking patterns, however, are different in other Tukanoan languages, where both animate and inanimate classifiers pattern alike, indicating that there is no motivation for distinguishing the former as gender and the latter as classifiers.
2. According to Shimizu and Phan, Vietnamese has two core classifiers. These classifiers obligatorily appear on numerals in both their NP and modification uses. They must also occur with demonstratives, adjectives, and verb-based nominalizations in their NP use, although they cannot be used with other types of structures than numerals in their modification use. However, classifiers are only optionally used with genitive noun phrases in their NP use. Therefore, the distribution of Vietnamese classifiers slightly deviates from Shibatani's prediction along the structural dimension, though it supports his prediction along the functional dimension.
3. Shirai shows that in Northern nDrapa, classifiers must follow basic numerals in both NP and modification uses, but do not occur with other types of structures—such as demonstratives, genitive noun phrases, non-reduplicated adjectives, or verb-based nominalizations—in either use. Nevertheless, there are exceptions: loan numerals representing large numbers may modify nouns without a classifier, and classifiers can be attached to adjectives that are nominalized through reduplication.
4. According to Li and Miyagishi, classifiers in the Guoli dialect of Zauzou are obligatorily used with numerals, demonstratives, and adjectives, and optionally with verb-based nominalizations (relative clauses) in NP use. In modification use, they are obligatory with numerals and demonstratives, and optional with adjectives and verb-based

¹¹ As mentioned in Footnote 5, some of the reports include data on not only cardinal numerals but also ordinal numerals, but these summaries deal only with cardinal numerals. Also, depending on the report, verb-based nominalizations may be limited to relative clauses or may include other types of verb-based nominalizations. In the summaries, the term 'verb-based nominalizations' is used even when it refers only to relative clauses.

nominalizations. However, classifiers alone cannot be used with genitive noun phrases in either NP or modification use; instead, the genitive enclitic must be used with genitive noun phrases when the possesum is implicitly identified. When the possesum is explicitly mentioned, the genitive enclitic is obligatory in inalienable possession constructions but optional in alienable possession constructions. Nevertheless, classifier marking on demonstratives in modification use exhibits inter-speaker variation, with some speakers considering it optional.

5. Miyagishi and Li show that in the Jiangmo dialect of Zauzou, the two types of classifying nominalizers—classifiers and pluralizers—are generally used obligatorily to derive all types of grammatical nominalizations in both NP and modification uses, except with genitive noun phrases in either use. In genitive constructions, the genitive enclitic is required in addition to the classifier or pluralizer when the possesum is implicitly recognized, but is optional when the possesum is overtly expressed.
6. Taguchi shows that in Lan Hmyo, classifiers are obligatorily used with numerals and demonstratives, and optionally used with adjectives and verb-based nominalizations (relative clauses), but they are not used with genitive noun phrases, when these structures occur in NP use. In contrast, in modification use, classifiers are used only with numerals, but not with demonstratives, genitive noun phrases, adjectives, or verb-based nominalizations.
7. Wu and Liu demonstrate that Gouliang Miao, a sub-dialect of Xiangxi Miao, deviates from Shibatani's hypothesis regarding the distribution of classifying nominalizers along the structural dimension, though it supports the hypothesis concerning their distribution along the functional dimension. This language exhibits interaction between classifiers and several non-classifying nominalizers. In the absence of a non-classifying nominalizer (i.e., the genitive marker and two dedicated nominalizers), classifiers are obligatorily used with numerals, demonstratives, genitives, and verb-based nominalizations (relative clauses) in both NP and modification uses, but they are not permitted to nominalize adjectives—representing a deviation from Shibatani's hypothesis along the structural dimension. When a non-classifying nominalizer is present, a classifier may optionally be used in genitive noun phrases, adjectival phrases, and verb-based nominalizations in both NP and modification uses.
8. Ilic and Kawachi show that in Swedish—which has traditionally been described as having gender (with a common/neuter distinction) but is more accurately characterized as an animacy-based classification system—its classifying nominalizers are of two types: different common/neuter forms (or suffix forms) and different common/neuter forms of singular articles. The common/neuter distinction is made by the former for demonstratives, but only restrictedly or not at all for other types of structures. With the latter, the common/neuter distinction is not made for numerals or genitive noun phrases, though it can occur with demonstratives, adjectives, and verb-based nominalizations

(relative clauses). This challenges Shibatani's hypothesis along the structural dimension while remaining compatible with it along the functional dimension.

9. Pardeshi and Bekeš show that in Slovene, gender markers are obligatorily attached to all types of structures on the Shibatani hierarchy—numerals, demonstratives, genitive noun phrases, adjectives, and verb-based nominalizations—in both NP and modification uses. Thus, the gender marking in this language supports Shibatani's hypothesis along both the structural and the functional dimensions.
10. According to Nakao, Bari marks gender on demonstratives, genitive noun phrases, most types of adjectives, and verb-based nominalizations in their NP and modification uses, but does not mark gender on numerals in either use. Therefore, the distribution of gender marking in this language contradicts Shibatani's hypothesis along the structural dimension, though it is compatible with it along the functional dimension.
11. Furumoto reports that in the Kimakunduchi dialect of Swahili, noun-class marking obligatorily occurs on all types of structures on the Shibatani hierarchy—numerals, demonstratives, genitive noun phrases, adjectives, and verb-based nominalizations (relative clauses)—in both NP and modification uses, though noun-class marking on numerals in both uses is restricted to 'one', 'two', 'three', 'four', 'five' and 'eight'. Therefore, the data approximately supports Shibatani's hypothesis along both the structural and the functional dimensions.
12. Kawachi shows that Sidaama has two types of classifying nominalizers that distinguish between feminine and masculine gender. One is an enclitic used to form a singular noun phrase and the other consists of allomorphs of the nominative, the genitive, and the dative-locative case suffixes. The enclitic marks gender obligatorily on genitive noun phrases and verb-based nominalizations (relative clauses) in NP use and optionally on these structure in modification use. It does not mark gender on numerals or adjectives in either use. In contrast, the case suffixes obligatorily mark gender on numerals and adjectives in both NP and modification uses, but not on genitive noun phrases in either use. Both types of markers are analyzed as included in demonstratives. Therefore, while each type of gender marker supports Shibatani's hypothesis along the functional dimension but deviates from it along the structural dimension, the two types of gender markers together appear to conform to the structural predictions of the hypothesis. Nevertheless, the optional gender marking on genitive noun phrases in modification use reveals a gap in the hierarchy.
13. Bez demonstrates that classifier marking in nine regional dialects of Assamese supports Shibatani's hypothesis along both the structural and functional dimensions. Specifically, all investigated dialects exhibit obligatory classifier marking on numerals, demonstratives, genitive noun phrases, adjectives, and verb-based nominalizations in

their NP use. In modification use, however, obligatory classifier marking is consistently restricted to numerals—and to numerals and demonstratives in two dialects.

14. Kelkar and Pardeshi show that Southern Marathi (Thanjavur Marathi) and Pune Marathi (Standard Marathi), both of which have classifier and gender markers, generally require the classifier to be attached to numerals in NP use. Although Southern Marathi must also use the classifier with most numerals in modification use, Pune Marathi does not permit its use in this function. In neither variety can the classifier be used to derive any other type of grammatical nominalization, whether in NP or modification use. On the other hand, in both varieties, gender marking is mostly obligatory on the numeral ‘one’ (though not on other numerals), demonstratives, adjectives, and verb-based nominalizations. However, while gender marking on genitive noun phrases in both NP and modification uses is obligatory in Pune Marathi, it does not occur in Southern Marathi. Therefore, the gender marking in Southern Marathi does not support Shibatani’s hypothesis along the structural dimension, although that in Pune Marathi, as well as classifier marking in both varieties, does. Both the classifier and gender marking, however, support Shibatani’s hypothesis along the functional dimension.
15. According to Miyagishi, Sinhala has three classifiers and a pair of gender markers. Each classifier must appear on numerals in both their NP and modification uses. Two of the classifiers must also occur on demonstratives, adjectives, and verb-based nominalizations, but only in their NP use (not in their modification use), and may optionally be used with genitive noun phrases in NP use. In this respect, the distribution of these two classifiers does not perfectly support Shibatani’s hypothesis along the structural dimension, though it confirms it along the functional dimension. The third classifier cannot be used with demonstratives, genitive noun phrases, adjectives, or verb-based nominalizations in either NP or modification use. The gender markers must appear on numerals in both NP and modification uses, as well as on demonstratives in NP use. Additionally, they may optionally occur on genitive noun phrases with first-pronoun possessors in NP use. Notably, the numeral for ‘one’ independently makes a gender distinction regardless of the type of structure it occurs with.
16. Pardeshi and Mopkar, who describe the patterns of marking with one sortal classifier and three gender markers in Konkani, show that while the use of the classifier is restricted to numerals other than ‘one’ in both NP and modification uses, the gender marker occurs obligatorily and almost obligatorily to derive grammatical nominalizations in NP and modification uses, respectively. However, gender marking is limited to the numerals from ‘one’ to ‘four’—the numeral for ‘one’ must carry the gender marker, but not the classifier, while the numerals for ‘two’, ‘three’, and ‘four’ obligatorily take both the classifier and the gender marker simultaneously or only the classifier. All other numerals must be accompanied by the classifier. This conforms to Shibatani’s hypothesis along both the structural and the functional dimensions.

17. Uppoor reports that gender marking in eight Dravidian languages—Arebase, Beary, Kannada, Koraga, Malayalam, Tamil, Tulu, and Telugu—supports Shibatani’s hypothesis along the functional dimension. In these languages, gender marking in grammatical nominalization is obligatory across all five categories—numerals, demonstratives, genitives, adjectives, and verb-based nominalizations—in NP use, except in Malayalam, where genitives lack gender marking. For modification use, gender markers occur only on numerals (and not on other categories), obligatorily in Arebase, Kannada, Koraga, and Tulu, and optionally in Beary. They appear only optionally on demonstratives in Arebase, Koraga, and Tulu, but not in the other five languages. In Malayalam, Tamil, and Telugu, gender marking cannot occur on any category in modification use. Gender marking in seven of the eight languages also supports Shibatani’s hypothesis along the structural dimension. In addition to being obligatory across all five categories in NP use, when these categories occur in modification use, gender marking—where possible—is largely restricted to numerals. With respect to classifier marking, it is typically found with numerals in NP use, while its occurrence with other categories remains to be investigated. In modification use, classifier marking is restricted to numerals.

4.2 General findings

Importantly, one part of Shibatani’s hypothesis applies consistently to all the languages investigated in this data report collection—namely, the hypothesis along the functional dimension. In each language, classifying nominalizers occur with equal or greater likelihood and obligatoriness with a given type of structure in NP use than in modification use. Classifying nominalizers are never attested with a wider range of structures in modification use than in NP use. Among the languages reported in this issue, none contradicts this pattern, making it, at least so far, an absolute universal. This finding challenges the traditional view that the primary function of grammatical nominalizations is to serve as noun modifiers structurally dependent on a head noun, rather than as full noun phrases in their own right.

The data from many of the languages investigated also confirm Shibatani’s hypothesis along the structural dimension. However, the use of classifying nominalizers in nine languages exhibits patterns not predicted by the hypothesis.¹² Setting aside the unusual absence of classifying nominalizer marking on numerals in Bari (Nakao this issue) and the requirement to use nominalizers on adjectives in addition to classifying nominalizers in Gouliang Miao (Wu and Liu this issue), the most common deviation—found in seven languages—involves genitive noun phrases, where the use of classifying nominalizers is either optional or disallowed, despite their obligatory or optional use with adjectives: Lan Hmyo, Malayalam, Southern Marathi, Sidaama, Sinhala, Swedish, and Vietnamese.

The exact reason for the frequent gap in genitive noun phrases is not entirely clear at this point. Nevertheless, there seem to be possible contributing factors. There appear to be language-specific factors influencing the absence of classifying nominalizer marking on genitive noun phrases. For example, some languages (e.g., Lan Hmyo in Taguchi this issue,

¹² The severity of these violations varies across languages. Minor violations tend to occur in either noun-phrase or modification use, whereas more serious violations often appear in both.

Vietnamese in Shimizu and Trang this issue) use a genitive marker that functions as a non-classifying nominalizer, which may obviate the need for classification. On the other hand, there is a motivated, though not universal, factor affecting classifying nominalizer marking on genitives, which always involve two entities—a possessor and a possessum—unlike other types of structures: the identifiability of the possessum relative to the possessor. The motivation for this factor varies across languages. Typically, the possessum in a possessive construction is definite (Haspelmath 1999) and therefore likely identifiable to the addressee, so its category need not be specified, whereas the possessor may be unidentifiable and is often marked for its category; further specification of the possessum’s category on the possessor noun can increase cognitive load. Nevertheless, when the possessor is clearly identifiable, the possessum may be less so, and marking the possessum’s category on the possessor pronoun can be informative—for example, only on first- and second-person possessor pronouns in Swedish (Ilic and Kawachi this issue).

5. REMAINING ISSUES AND PROSPECTS FOR FUTURE RESEARCH

Based on the data reported in this collection, several issues remain that are specific to individual languages or particular types of languages, and these will need to be addressed in future research. Some of these issues involve accounting for deviations from Shibatani’s predictions or the application of his hypothesis to one type of noun phrase but not to another. Classifier and gender marking may be analyzed as a unified phenomenon, with the two types of markers forming a continuum rather than being entirely distinct or unrelated. Nevertheless, it would be worthwhile to investigate whether certain phenomena—such as specific types of deviations from Shibatani’s hypothesis—are more likely to occur with one type of classifying nominalizer or with another type and if so, why this is the case.

A first issue is that there are languages with the singular-plural number distinction in which marking with a relatively small number of categories distinguished by classifying nominalizers is restricted to the singular and does not apply to the plural (e.g., Sidaama in Kawachi this issue, Swedish in Ilic and Kawachi this issue). This pattern appears to be explainable: a singular referent can be readily identified, whereas plural referents may consist of individuals of different genders, making marking with a classifying nominalizer more complex or ambiguous. This raises the question of whether a similar restriction to the singular is also possible in languages with a large number of classifiers.

Second, in some languages with a gender marking system (e.g., Romance languages, Slavic languages), only the numeral ‘one’ or lower numerals have gendered forms or receive gender marking, while higher numerals do not (e.g., Sidaama in Kawachi this issue, Sinhala in Miyagishi this issue, Kimakunduchi dialect of Swahili in Furumoto this issue, Konkani in Pardeshi and Mopkar this issue, Marathi in Kelkar and Pardeshi this issue). The likely motivation is the same as for the first issue: whereas plural referents may include individuals of different genders, the gender of a singular referent is more easily identifiable. Although there are instances where lower numerals receive special treatment in terms of classifier form selection (e.g., the use of *tsu* in Japanese for numerals ‘one’ to ‘nine’), this pattern seems to be more common with gender markers than with classifiers in languages. The reason for this discrepancy warrants further investigation.

Third, in some languages, the use of a particular classifying nominalizer is largely restricted to definite referents (e.g., the noun phrase enclitic in Sidaama in Kawachi this issue, definite gender markers in other Afroasiatic languages). Among various types of grammatical nominalizations, a noun phrase consisting solely of a demonstrative or containing a demonstrative is inherently definite. Similarly, a noun phrase composed only of a genitive noun phrase or an ordinal numeral in NP use or one containing such a structure in modification use is typically interpreted as definite. If a classifying nominalizer functions as a definiteness marker, its occurrence in a noun phrase may be optional. It may appear for added emphasis when the definiteness is already contextually clear, or it may be omitted to avoid redundancy. As discussed in Section 4.2, some languages exhibit a gap in the implicational hierarchy with respect to genitive noun phrases. This may be explained by the principle that definiteness in such constructions is already assumed, rendering further marking unnecessary (Ilic and Kawachi this issue). Notably, definiteness appears to correlate more strongly with languages that have a small number of classifying nominalizers than with those that have a larger number, although the precise relationship between definiteness and the number of classifying nominalizers remains unclear. This issue remains open for future research.

Fourth, there is a question of whether gender marking that participates in grammatical nominalization and the existence of different gender forms of grammatical nominalizations (e.g., different gender forms of ‘one’ in some languages, for example, Sidaama) should be treated as the same type of phenomenon or different types of phenomena. In some languages, a gender marker occurs as part of a grammatical nominalization. For example, in many Afroasiatic languages, the feminine-masculine distinction is encoded in elements of demonstratives, which trace back to the Proto-Afroasiatic distinction between *t* (feminine) and *k* (masculine). The question is whether this should be viewed as an obligatory gender distinction expressed by demonstratives or as a case where no additional marking made beyond what is already inherent in the demonstratives.

A related issue concerns whether grammatical nominalizations in any language have distinct classifier forms. In some languages, all members of the (often, singular) demonstrative set have different gender forms. This is likely because demonstratives, unlike other types of grammatical nominalizations, belong to a closed class. However, such distinctions do not appear to be made for classifiers.

Fifth, some languages have distinct (typically, two) types of classifying nominalizers. These may consist of two kinds of classifiers (e.g., general vs. specific, or more grammaticalized vs. more nominal-like), two types of gender markers, or a combination of classifiers and gender markers. Based on the data discussed in the present study, languages with two types of classifying nominalizers that exhibit different morpho-syntactic properties—specifically Sidaama and Swedish—each show deviations from Shibatani’s hierarchy. However, taken together, they may better conform to his hypothesis, with each language covering different, discontinuous segments of the hierarchy. This raises the question of whether the range of each type of marker should be analyzed individually, or whether the usage of all markers should be considered collectively. A further issue is whether discontinuous portions of the hierarchy can be filled by any combination of classifying nominalizers—whether this

involves two types of gender markers, two types of classifiers, or a combination of classifiers and gender markers.

Sixth, there are cases in which it is unclear whether a construction should be analyzed as a noun phrase consisting of a grammatical nominalization in modification use plus the noun that it modifies or as two juxtaposed noun phrases made up of a noun phrase—one consisting of a grammatical nominalization in NP use and the other containing only a noun. Identifying the constituent structure of such noun phrases can be particularly challenging when word order is flexible and dependency relations are not overtly marked (e.g., Kawachi this issue). According to Shibatani's theory, the referent of a noun phrase including one consisting only of a grammaticalization in NP use is identified metonymically through the discourse context and Gricean maxim of relevance. It is necessary to explain how a construction that permits both interpretations is processed differently depending on the interpretation assigned.

Seventh, there are languages with classifying nominalizers that are neither classifiers nor gender markers, yet can still be analyzed as grammatical nominalization markers. Examples include the pluralizers in the Jiangmo dialect of Zauzou (Miyagishi and Li this issue) and the singular and plural nominalizer particles in Kupsapiiny, a Southern Nilotic language of Uganda. It would be worthwhile to investigate how these types of classifying nominalizers behave similarly to or differently from classifiers and gender markers, and to what extent they can be situated on the continuum between the two types.

Finally, there are grammatical phenomena that seem closely related to classifying nominalization, but are not addressed in the present issue, and therefore it is not clear how they align with Shibatani's hypothesis. For example, beyond the commonly discussed numeral classifiers, there are other types—such as verbal classifiers, possessive classifiers, locative classifiers, and deictic classifiers (Aikhenvald 2000), as well as root classifiers (Grinevald 2024). Possessive classifiers in Oceanic languages, which may retain the Proto-Oceanic classification of alienably or indirectly possessed items into food, drink, and other items, are restricted to genitive noun phrases (François 2023).

6. CONCLUSION

The Shibatani hierarchy under his framework of Dynamic Functional Typology serves as a useful analytical tool for examining both the synchronic distribution of classifying nominalizers used to derive grammatical nominalizations and the diachronic development of this distribution across languages. Based on recurrent distribution patterns and common developmental paths, it becomes evident that phenomena in individual languages that deviate from these patterns require explanation. At the same time, although it is not yet clear whether the hierarchy applies equally to different types of classification systems, it may prove useful for distinguishing among them.

ABBREVIATIONS

ADJ	adjective
CLF	classifier
DEM	demonstrative
GEN	genitive noun phrase
Mod	modification
NMLZ	nominalization
NP	noun-phrase
NUM	(cardinal) numeral
SG	singular

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to our language consultants for their invaluable contributions to this research. Our deepest thanks go to Masayoshi Shibatani for his helpful comments on various aspects of the study and on several of the individual reports. Any remaining errors or shortcomings are the sole responsibility of the first author. This research has been supported by grants from the Japan Society for the Promotion of Science (JSPS KAKENHI Project IDs: 22H00659 and 23K21931, ‘Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers’, PI: Kazuhiro Kawachi) and the National Institute for Japanese Language and Linguistics (‘Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspectives’, PI: Prashant Pardeshi). Additional support to the first author was provided by the Keio Gijuku Fukuzawa Memorial Fund for the Advancement of Education and Research, the Keio University Academic Development Funds, and the Ushioda Memorial Fund.

REFERENCES

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Allan, Keith. 1977. Classifiers. *Language* 53.2, 285–311.
- Corbett, Greville G. 1991. *Gender*. Cambridge University Press.
- Craig, Colette G. 1986. Jacalteco noun classifiers: A study in grammaticalization. *Lingua* 70.4, 241–284.
- Dryer, Matthew S. 2004. Noun phrases without nouns. *Functions of Language* 11.1, 43–76.
- François, Alexandre. 2017. The economy of word classes in Hiw, Vanuatu: Grammatically flexible, lexically rigid. *Studies in Language* 41.2, 294–357.
<https://doi.org/10.1075/sl.41.2.03fra>
- François, Alexandre. 2023. Grammars of possession in the Torres–Banks linkage: A comparative study. Paper given at the 2023 Vanuatu languages conference, organized by Max Planck Institute–EVA. Port Vila, Vanuatu, on July 13, 2023.
- Greenberg, Joseph. 1974. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Proceedings of the Eleventh International Congress of Linguists, Bologna-Florence, Aug. 28-Sept. 2, 1972*, 17–37.

- Grinevald, Colette. 2000. A morpho-syntactic typology of classifiers. In Gunter Senft. (ed.) *Systems of nominal classification*, 50–92. Cambridge University Press.
- Grinevald, Colette. 2002. Making sense of nominal classification systems: Noun classifiers and the grammaticalization variable. In Ilse Wischer and Gabriele Diewald. (eds.) *New reflections on grammaticalization*, 259–275. Amsterdam: Benjamins.
- Grinevald, Colette. 2024. The ‘determinatives’ of Ancient Egyptian hieroglyphs are ‘root classifiers’. How egyptology and linguistics met to show it. In Niv Allon and Racheli Shalomi-Hen. (eds.) *Signs of life: Ancient Egyptian script, language, and writing studies in honour of Orly Goldwasser*, 63–74. Turnhout, Belgium: Brepols.
<https://doi.org/10.1484/M.MRE-EB.5.137005>
- Grinevald, Colette, and Frank Seifart. 2004. Noun classes in African and Amazonian languages: Towards a comparison. *Linguistic Typology* 8, 243–285.
- Haspelmath, Martin. 1999. Explaining article-possessor complementarity: Economic motivation in noun phrase syntax. *Language* 75, 227–243.
- Li, Yu, and Tetsuya Miyagishi. This issue. Nominalization in Guoli Zauzou.
- Pardeshi, Prashant, and Masayoshi Shibatani. 2023. Marathi relative and complement clauses in nominalization perspective. In Rajesh Kumar and Om Prakash. (eds.) *Language studies in India: Cognition, structure, variation*, 321–352. Singapore: Springer Nature.
https://doi.org/10.1007/978-981-19-5276-0_19
- Shibatani, Masayoshi. 2009. Elements of complex structures, where recursion isn’t: The case of relativization. In Talmy Givón and Masayoshi Shibatani. (eds.) *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution*, 163–198. Amsterdam & Philadelphia: John Benjamins.
- Shibatani, Masayoshi. 2017. Nominalization. In Masayoshi Shibatani, Shigeru Miyagawa, and Hisashi Noda. (eds.) *Handbook of Japanese syntax*, 271–331. Berlin & Boston: De Gruyter.
- Shibatani, Masayoshi. 2018a. Nominalization in crosslinguistic perspective. In Prashant Pardeshi and Taro Kageyama. (eds.) *Handbook of Japanese contrastive linguistics*, 345–410. Berlin & Boston: De Gruyter.
- Shibatani, Masayoshi. 2018b. Nominalization. In Yoko Hasegawa. (ed.) *Handbook of Japanese linguistics*, 431–462. Cambridge: Cambridge University Press.
- Shibatani, Masayoshi. 2019. What is nominalization? Towards the theoretical foundations of nominalization. In Roberto Zariquiey, Masayoshi Shibatani, and David W. Fleck. (eds.) *Nominalization in languages of the Americas*, 15–168. Amsterdam & Philadelphia: John Benjamins.
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.
https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf-char/en

- Shibatani, Masayoshi, and Marília Ferreira. This issue. Dynamic functional typology of Tukanoan classifying nominalizations.
- Shibatani, Masayoshi, and Hiromi Shigeno. 2013. Atami nominizations. *International Journal of Okinawan Studies* 4.1, 107–139.
- Shirai, Satoko. 2022. Classifiers in nDrapa: Definition and categorization. *Gengo Kenkyu* 162, 25–46.
https://ls-japan.org/wp-content/uploads/2024/02/162_shirai.pdf

Dynamic functional typology of Tukanoan classifying nominalizations

Masayoshi Shibatani¹ and Marília Ferreira²

ABSTRACT

The Tukanoan languages of the Amazon Basin provide compelling evidence for the ongoing debate on gender and classifier systems, recently reanalyzed by Shibatani (2023) as *classifying nominalizations*—structures that simultaneously nominalize numerals, demonstratives, etc. and categorize their referents (e.g., ‘two **feminine-class** entities,’ ‘this **round-shaped** entity’). An analysis of patterns of convergence and divergence of animate classifiers (based on animacy features such as animate/inanimate, human/nonhuman, and masculine/feminine) and inanimate classifiers (based on physical attributes and function) across several Tukanoan languages reveals that the former consistently outpace the latter in grammaticalization, thereby shaping specific evolutionary trajectories of classifier marking patterns. This precedence of animate classifiers over inanimate classifiers is attributed to their grounding in animacy features that are most salient to human cognition and social life, resulting in distinct marking patterns between the two. Over time, however, inanimate classifiers tend to converge with animate ones, gradually diminishing the distinctions between them—only for another cycle of change to begin anew. These findings challenge the traditional view of gender and classifiers as distinct grammatical phenomena, instead revealing a more unified system nuanced by varying degrees of grammaticalization. The paper further highlights the analytical power of Shibatani’s framework of Dynamic Functional Typology, particularly his proposed hierarchy predicting both the diachronic evolution and synchronic distribution of gender/classifier marking.

1. INTRODUCTION

This theoretical exploration of Tukanoan gender and classifier systems builds on Janet Barnes’ seminal studies of Tuyuka (Barnes 1984, 1990; Barnes and Malone 2000), and on Shibatani’s (2021, 2023) study of classifying nominalizations, which argues that what are commonly referred to as gender and (nominal) classifiers are, in fact, part of a unified system of grammatical nominalizations that classify nominal denotations—things and thing-like entities, in contrast to ordinary nominalizations that do not serve a classificatory function. The languages of the Tukanoan family, along with those from several other families in the Amazon Basin, offer intriguing evidence for the typology of gender and classifiers and their relationship to nominalization—particularly because they simultaneously feature animate classifiers (based on animacy features, distinguishing between animate and inanimate, human and non-human, as well as feminine and masculine) and inanimate classifiers (based on physical properties and functional attributes). This paper focuses on three central questions: (i) whether animate and inanimate classifiers in Tukanoan represent two distinct systems of nominal classification (e.g.,

¹ Professor Emeritus, Rice University/Kobe University. email: Matt.shibatani@gmail.com

² Professor, Federal University of Pará. email: mariliaferreira1@gmail.com

a gender system versus a classifier system); (ii) what grammatical functions gender and classifiers serve; and (iii) the locus of gender/classifier marking³. Before presenting the analytical framework we adopt—specifically, Shibatani’s recent analysis of nominalization—and prior to our detailed examination of the Tukanoan data, we begin by reviewing past achievements and identifying the limitations of existing research on the key issues as they manifest in the Amazonian context.

1.1. Gender and classifiers

Tukanoan studies have made significant strides toward a unified treatment of gender-like animate classifiers and inanimate classifiers marking numerals and several other structures. Like Shibatani’s proposal unifying them as classifying nominalizations (see Section 2), traditional Tukanoan treatments⁴ group these two categories under the rubric of classifiers, distinguishing between animate classifiers and inanimate classifiers (e.g., Barnes (1990) on Tuyuka, Jones & Jones (1991) on Barasano). More recent research continues this line of inquiry, proposing that animate and inanimate classifiers form part of a single system organized along a grammaticalization continuum. This view is expounded by studies such as Gomez-Imbert (1982, 2007) on Tatuyo, Stenzel (2004/2013) on Kotiria, and Farmer (2015) on Máihiki⁵. Advocates of the unified approach point to morphosemantic continuum between more grammaticalized animate classifiers and inanimate ones that tend to be less grammaticalized⁶. Observe Table 1, representing such an effort.

³ We maintain that nominal classification constitutes a unified system embracing both what have traditionally been referred to as *gender* and *classifiers*. However, to highlight the inclusion of conventional gender categories in the discussion that follows, we will frequently use the notation “gender/classifier (marking)”.

⁴ Largely works by linguists such as Janet Barnes and others affiliated with the Summer Institute of Linguistics.

⁵ These works are all inspired by Colette Grinevald’s seminal works on the typology of gender and classifiers, especially Grinevald (2000).

⁶ In addition, the phonological dimension of grammaticalization—phonological reductions from disyllabic to monosyllabic, etc.—has been discussed, but our focus in this paper is on the morphosemantic properties.

Table 1. Morphosemantic continuum of Tatuyo classifiers (Gomez-Imbert 2007: 409)

LEX → GRAM						
<i>a</i>	<i>b</i> cv(c)v	<i>c</i> cvv	<i>d</i> cvv	<i>e</i> cv	<i>f</i> cv	<i>g</i> (c)v
<i>ope_L</i> 'hole'	<i>~pái_H</i> 'blade'	<i>-baa_L</i> 'basket'	<i>-ra(a)_L</i> 'fiber'	<i>-wĩ</i> 'tubular'		{ <i>-kĩ</i> , <i>-ĩ</i> } 'AN M'
<i>wĩ_{HL}</i> 'house'	<i>~toó_{LH}</i> 'bunch'	<i>-haa_L</i> 'packet'	<i>-ti(i)_L</i> 'pill'	<i>~we</i> 'filiform'		{ <i>-ko</i> , <i>-o</i> } 'AN F'
<i>peta_L</i> 'port'	<i>-rika_L</i> 'cylinder' ¹¹			<i>~jo</i> 'palm.tree'		{ <i>-he</i> , <i>-je</i> , <i>-e</i> } 'INAN'
<i>-bétó_H</i> 'hoop'						{ <i>-ka</i> , <i>-a</i> } 'round'
<i>~bahá_{LH}</i> 'people'						{ <i>~ra</i> / <i>~da</i> / <i>-a</i> } 'AN PL'
<i>-áhé_{HL}-ro</i> 'skin'						{ <i>-ro</i> / <i>-to</i> / <i>~do</i> } 'INAN.SG'
<i>kaha_L-wo</i> 'trellis'				<i>-wo</i> 'heap'		
<i>júkt_H</i> 'wood'						{ <i>-kĩ</i> / <i>-ĩ</i> } 'CYL'
<i>ria_L</i> 'river'					<i>-ja</i>	
<i>~baa_L</i> 'path'				<i>~wa</i>		
<i>hot_{HL}</i> 'pot'				<i>-rĩ</i>		

The table is organized along a continuum, with the right edge—representing the grammatical or most grammaticalized end—featuring classifiers that are opaque in terms of both their meanings and their morphological connection to the lexical items from which they may have originated. At the opposite end, on the lexical side (the left edge), are classifiers that retain clear morphosemantic connections to their lexical sources, most often nouns; e.g., Barasano *bãa yao-ri-bãa* [path long-NZR-CLF_{PATH} 'a long path' (Jones and Jones (1991: 26). These are commonly referred to in the field as “repeater classifiers” and are also widely observed in East Asian languages such as Thai and Burmese. Positioned between these two poles are classifiers that exhibit varying degrees of morphosemantic transparency with respect to their lexical origins—for example, those that appear as truncated or phonologically reduced forms of the lexical items from which they are derived.

It is noteworthy that Gómez-Imbert (1982, 2007) identifies certain inanimate classifiers that have grammaticalized in ways comparable to animate classifiers, exemplifying a case of convergence between the two types. Table 1 illustrates this continuum and the points of convergence in terms of the inanimate classifier *ka/a* (used for round objects) that aligns with animate classifiers in the rightmost column, representing the most advanced stage of grammaticalization.

What has been lacking in this type of approach to the typology of gender/classifiers in terms of the grammaticalization continuum is a systematic investigation of the syntactic distribution and marking patterns of animate and inanimate classifiers. This paper precisely addresses that gap, demonstrating that the significant convergence in their distribution and marking patterns supports a unified analysis of the two.

These efforts stand in contrast to the widely held view that gender and classifier systems are fundamentally distinct—a position advanced by Dixon (1986), Corbett (1991), and others. This perspective has been influential in the field, as also seen in analyses of Amazonian languages such as Aikhenvald’s (2007) work on the Arawakan language Baniwa, Chacon (2022) on the Central Tukanoan language Kubeo, Eraso’s (2015) analysis of Tanimuka (also Central Tukanoan), and Vallejos (2021) on the Western Tukanoan language Sekoya.

1.2. Functions of gender/classifier marking

In the Amazonian context, there is little need to argue that gender/classifiers serve nominalization as a core function, as this is widely acknowledged in the existing literature. This is in line with a recurring observation in wider contexts beyond the Amazon Basin. In one of the earlier studies on Asian classifiers, Conklin (1981: 365, 374) observed that “[a]ll the Tai languages use classifiers as noun substitutes and as nominalizers”, “WAN [Western Austronesian] classifiers serve as noun substitutes and sometimes as nominalizers or indicators of nominal intent”, and that “[c]lassifiers are more like[ly] to be used if it is necessary to indicate nominal status of an expression”.

Despite such wide recognition of nominalization as a central function of gender/classifiers, few studies—whether on Amazonian or other languages—probe deeply into the true nature of nominalizations (or “deverbals”) themselves. They often overlook the cognitive foundations of nominalization, fail to distinguish between different types, and lack a thorough understanding of their distributional patterns and functional roles in grammar. Some of these problems are taken up in Section 2, where we review Shibatani’s work on nominalization.

Another problematic aspect of the thinking surrounding the functions of gender/classifier marking concerns its role as an inflection/agreement marker. This issue is also a focal point in the study of Amazonian languages, as highlighted in passages like the following: “Barasano has an extensive system of noun classifiers, which provide concordance (agreement) within the noun phrase” (Jones and Jones 1991: 49), where morphological agreement is also seen, as in (1a)⁷. Observe:

⁷ This observation is significant because it undermines the alleged distinction between gender and classifiers proposed by Corbett (1991), who defines gender specifically in terms of agreement. Observe: “the determining criterion of gender is agreement” (1991: 4); “**classifiers** fall outside our study [of gender] because they **do not show agreement**.” (1991: 5: emphasis added)

(1) Barasano (Jones and Jones 1991: 59, 62)

- a. **gūtahāi** [hia-**hāi**/*-gi]
stone.CLF_{FLAT.THIN} two-CLF_{FLAT.THIN}/CLF_{HAMMOCK}
‘two stone slabs’
- b. **hūi** [ĩ ya-**gi**/*-hāi]
hammock 3M.SG GEN-CLF_{HAMMOCK}/CLF_{FLAT.THIN}
‘his long hammock’

While the patterns shown in the Barasano examples with inanimate classifiers above perfectly align with definitions of agreement such as “some syntactic covariance between a semantic or formal property of one element and a formal property of another” (Steele 1978: 610), scholars who classify animate classifiers as gender and distinguish them from inanimate classifiers tend to recognize agreement only in the former. This view overlooks the semantic and morphological concord exhibited by inanimate classifiers as seen in the Barasano examples above, thereby obscuring the broader unity of the two. A more pertinent question is whether classifier marking—animate or inanimate—is obligatory (or optional) in modification contexts. Animate classifiers generally display a higher degree of grammaticalization, evidenced by their stronger tendency toward obligatory marking in such contexts—a pattern amply demonstrated in the following study⁸. This distinction offers a more meaningful basis for evaluating classifier behavior, especially given that both animate and inanimate classifiers may exhibit agreement with the head noun when co-occurring within a noun phrase.

On another front, the dichotomy between the nominalization/derivational function and the inflection/agreement role of gender and classifiers has posed an ongoing challenge for Doris Payne, a specialist in the Peba-Yaguan language Yagua of the Amazon Basin, dating back to her dissertation (Payne 1985: Chap. 4). However, this issue must be examined in light of the actual distribution of classifier marking. As Barnes (1990) notes, classifier-marked forms typically occur in contexts where there is no agreement-controlling head noun. She challenges the appropriateness of the definition of classifiers presented in Allan’s (1977) influential paper⁹, observing the actual usage of Tuyuka classifiers as follows:

[I]n Tuyuca “the associated noun” in Allan’s definition [i.e., the head noun] is not necessarily expressed explicitly in the sentence and discourse in which the classifier occurs; nevertheless, the referent is understood by context or gesture. Noun phrases which include both the referent [the agreement-controlling head noun] and a descriptive word containing a classifier [agreeing with the head noun] are rare in natural speech... (Barnes 1990: 273)

⁸Obligatory use in appropriate contexts is a hallmark of grammaticalization.

⁹Allan (1977: 285) defines classifiers in terms of the two criteria: “(a) they occur as morphemes in surface structures under specifiable conditions; (b) they have meaning, in the sense that a classifier denotes some salient perceived or imputed characteristic of the entity to which an associated noun refers (or may refer)”. Barnes objection pertains to criterion (b).

A similar observation is made on the Western Tukanoan language Ecuadorean Siona by Bruil (2014) as follows:

because numerals, demonstratives, adjectives and nominalized verbs in combination with a nominal classifier can be considered to be part of a noun class, they do not need to occur in combination with a nominal head. Since they are nominal, nominalized verbs can head a noun phrase themselves. This is how these forms are predominantly used. They occur mostly as the head of a noun phrase without any head noun. (Bruil 2014: 137)

These observations resonate the conclusions of Shibatani (2021, 2023), who show that, crosslinguistically, gender/classifier marking occurs more prevalently in contexts where there is no head noun to control agreement. By contrast, marking is less common in modification contexts with a head noun, where it is often either optional or impossible. These findings suggest that Payne’s dilemma is a pseudo-problem. Agreement is not an essential function of gender/classifier marking; rather, it emerges as a byproduct of the modification use of **forms that are gender/classifier marked independently of agreement with the head noun**¹⁰ (see Sections 2.3 and 2.4). Once gender and classifier marking are properly recognized as intrinsic to the grammatical process of classifying nominalization—as an autonomous derivational mechanism—the notion of an agreement controller becomes irrelevant. After all, does anyone claim that a head noun controls agreement in lexical nominalizations marked by genders and classifiers such as *assassin-o/assassin-a* ‘male/female killer’ in Portuguese and *gũta-hãi* (stone-CLF_{FLAT.THIN}) ‘stone slabs’ in Barasano? We argue that the same holds true in grammatical nominalizations based on numerals, demonstratives, adjectives etc. Such independently gender- or classifier-marked forms as Portuguese *duas* [two.F] ‘two feminine-class entities’, *lind-o* (beautiful-M) ‘a handsome male-class entity’ or Barasano *hua-rahe* [two-CLF_{HOLLOW.CYLINDRICAL}] ‘two hollow, cylindrical entities’ and *sũa-ri-hãi* [red-NZR-CLF] ‘a red, thin, flat entity’ may be used as modifiers (see Sections 2.3 and 2.4). Whether they satisfy the semantic congruity required of modification constructions is a separate matter (cf. *square triangle*, *black albino rabbit*, Portuguese *duas homens* ‘two.feminine men’, Barasano *hua-rahe sudihãi* [two-CLF_{HOLLOW.CYLINDRICAL} cloth] ‘two hollow, cylindrical pieces of cloth’).

The broader implications of these points are evident for those who define gender and classifiers and their functions in terms of the agreement-controlling head noun, as proposed by Corbett (1991) for gender and Aikhenvald (2019: 9) for both.

1.3. Locus of multiple classifier-marking

The final major issue addressed in this paper is the concept of ‘multiple classifiers’ (Aikhenvald, 2000; Aikhenvald, 2007), a phenomenon widely recognized by scholars of Amazonian languages. This term refers to the use of gender/classifier marking across various structures, including numerals, demonstratives, and adjectives. The observation that so-called numeral classifiers can extend beyond numerals is not new. It can be traced back to Haas (1942),

¹⁰There are, in fact, both gender- and classifier-marked constructions that cannot modify a head noun, barring the possibility of agreement marking by a head-noun as controller—see Shibatani (2021, 2023) for details.

who questioned the suitability of the term “numeral classifiers” for Thai classifiers, thereby highlighting that this typological feature is not unique to Amazonian classifier systems.

The main problem here lies in understanding the nature of the domain of ‘multiple gender/classifier marking’. Aikhenvald (2007: 479) described it as “seemingly disparate environments” in her study of the Arawakan language Baniá. This same perception—that the relevant domain consists of random, unstructured environments—can be seen in the varying ways specialists of Tukanoan languages enumerate the multiple domains of classifier marking. As for Tuyuka, Barnes offers the list below:

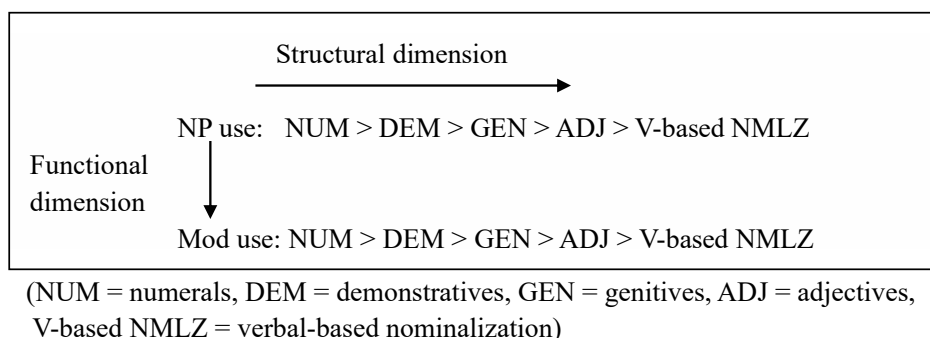
(2) Tuyukua classifier-marking targets (Barnes 1990: 274)

1. Nouns
2. Demonstratives
3. Numerals
4. Genitives
5. Nominalized verbs

More broadly, she notes that “[i]n all the Eastern [Tukanoan] languages plus the Central Kubeo, classifiers are found suffixed to **numerals, nouns, demonstrative adjectives, possessives and verbs** (to form nominalized verbs)” (Barnes (1999: 218-219), emphasis added, also in the below quotes). Other researchers similarly offer unstructured lists of elements that participate in multiple-classifier environments; e.g., “NP modifiers in Kubeo [which can be nominalized by classifiers] include: **adjectives, numerals, quantifiers, possessive pro-forms, personal pronouns and demonstratives**” (Chacon 2022: 20); “elements to which Máihĩki classifiers...may be suffixed...: **adjectival roots; numeral roots; demonstrative roots; bare verb roots; verb roots that have been suffixed with a marker of nominal tense**” (Farmer (2015: 153) on the Western Tukanoan Máihĩki); “[t]he examples below show that the classifiers can be used on **nouns, numerals, demonstratives, adjectives and verbs**” (Bruil (2014: 35) on Ecuadoran Siona, another Western Tukanoan language). Finally, Rose (2024) lists **numerals, adjectives, nominals (N/DEM), verbs** in this order for Mojeño Trinitario, an Arawakan language.

In stark contrast to the seemingly random distributions of gender and classifier marking implied by these enumerations, Shibatani (2021, 2023) proposes a highly structured, hierarchically organized domain of so-called multiple classifiers, as in Diagram 1 below:

Diagram 1. Hierarchy constraining development and distribution of nominalization markers



This domain is mapped along both structural and functional dimensions. The proposed hierarchies aim to account for the diachronic expansion of gender/classifier marking along two dimensions of structures and use/function. They also provide explanatory power for the resulting synchronic distribution patterns observed both within individual languages and across dialects and language families, and beyond.

One of the primary objectives of this paper is to demonstrate the value of Shibatani's framework in analyzing the classifier phenomenon in Tukanoan languages. Before delving into this main topic, we first provide a brief outline of Shibatani's nominalization theory in the next section (Section 2). Section 3 provides an overview of key information about the Tukanoan languages. Our in-depth study of gender and classifier marking begins in Section 4, where we introduce the methodology of our research, and continues in Section 5, where we examine marking patterns across various structures and usage domains in the Tuyuka language. Sections 6 and 7 present a comparative analysis of classifier-marking patterns across several Tukanoan languages, using the framework of Dynamic Functional Typology. These sections offer a glimpse of the explanatory power of the proposed framework in accounting for both language-internal and cross-linguistic variation in classifier-marking patterns. Finally, Section 8 provides a brief conclusion, along with suggestions for how the issues surrounding the gender/classifier controversy might be more effectively addressed.

2. SHIBATANI'S THEORY OF NOMINALIZATION

There are several aspects of Shibatani's (2019) nominalization theory that are innovative, as highlighted in the following review.

2.1. Cognitive underpinning

One of the most innovative aspects of Shibatani's theory of nominalization is its emphasis on the cognitive foundation of the nominalization process. Specifically, it proposes that nominalization, like other derivational processes, is fundamentally metonymic in nature, producing structures that denote nominal entities whose meanings are evoked through metonymic associations with the meaning of the base form. For example, when the base form

is the verb *run*, the derived nominal *runner* denotes¹¹ an entity intrinsically linked to the action of running—namely, a person who runs. In cases where the verb is transitive, either of the event participants closely associated with the action may be denoted by the nominalized form, such as *trainer* (agent) and *trainee* (patient). Additionally, nominalized forms can extend to denote other entities closely associated with an action or event, as seen in *diner* (facility/location) and *cleaner* (instrument/cleaning agent). The metonymic underpinning of the nominalization process imposes constraints preventing the derivation of nominal forms with meanings that are too distant from the concept of the base form. For example, it would not permit a derivation like *Londoner* to denote a casual visitor to London or *rear-ender* to denote an auto accident causing a dent in the front bumper, as such meanings lack a clear and stable metonymic link to their base forms¹².

2.2. Nominal-based nominalizations

A second innovation is the inclusion of nominal-based nominalizations to accommodate derived forms such as *Londoner* and *rear-ender* mentioned above, as well as forms like *villager*, *lifer*, *tenner*, and *quarter-pounder*. This challenges the conventional wisdom that nominalization only derives nouns from other word classes¹³. Once again, the derivational suffix *-er*, which we argue is the same suffix discussed in the previous paragraph, forms nouns denoting entities closely associated with the meanings of their base nouns—those intimately connected to village life, a life sentence, ten dollars, and a quarter pound.

The possibility of deriving new nominals from other nominals plays an important role for our discussions in this paper since some of the nominalizations effected by gender/classifier marking are based on nominals, such as nouns, numerals, and demonstratives, as in the lists of structures forming multiple-classifier environments in Section 1.3 above.

2.3. Lexical vs. grammatical nominalizations

Lexical nominalizations, such as *runner* and *villager* discussed above, are distinguished from grammatical nominalizations. The former are nouns that are likely memorized and stored in the mental lexicon, while the latter are typically phrasal units larger than a single word and are nonce expressions not retained in the brain. Many of these are referred to as participles in the traditional jargon, as in example (3) below. Here again, the bracketed nominalization denotes a person intimately associated with the act of sitting in the chair in the dark.

(3) [Sitting/Seated in the chair in the dark] was none other than my old man/woman.

¹¹ Shibatani distinguishes between “denotation” and “reference.” The former refers to the relationship between a linguistic form (e.g., *dog*) and a mental concept or simply its meaning. Reference, in contrast, is the speech act of pointing to an entity in the real or imagined world using a linguistic form, as in *the dog is barking*, uttered in a specific context to identify an entity instantiating the mental concept (denotation) of the noun *dog*.

¹² Fillmore’s earlier Case Grammar and more recent Frame Semantics are both based on the human cognitive ability to understand the world through metonymy.

¹³ Cf. “nominalization refers to the process of forming a noun from some other word-class or, especially in classical transformational grammar, the derivation of a noun phrase from an underlying clause” (Crystal 1980: 328); “operations that allow a verb to function as a noun...are called **nominalizations**, and can be described with a simple formula: $V \rightarrow N$.” (Payne 1997: 223)

Another distinction between lexical and grammatical nominalizations is that the former, as nouns, denote uniform entities, whereas grammatical nominalizations typically denote a variety of entities that satisfy the metonymic constraint mentioned above. For example, in German, as in many other European languages and Tukanian as well, “gender” marking derives both lexical nominalizations (nouns) and grammatical nominalizations. The German feminine noun (*die*) *Weiß*e (NB the initial capital letter for nouns), for instance, denotes a member of the uniform category of white women, whereas its grammatical nominalization counterpart (*die*) *wei*ße can denote a variety of white entities embodying the property of whiteness and that fall in the feminine class, such as a white hen, a white cat, a white tulip, a white flag, etc.

One of the motivations for treating grammatical nominalizations as such is their morphological connection to lexical nominalizations, the latter of which are widely recognized as instances of nominalization. This connection is evident in the so-called gerundive (or participial) endings *-ing* and *-ed* mentioned above, which also derive lexical nominalizations (nouns); e.g., *the [killing] of endangered animals*; *the [injured] were taken to the hospital*¹⁴. Such morphological links between lexical and grammatical nominalizations are widely observed, not only in ordinary nominalizations but also in classifying nominalizations involving gender/classifier marking across languages, including Amazonian languages.

A problem of including nouns in the multiple classifier domain, as seen in most studies mentioned in Section 1.3, is the mixing of lexical and grammatical nominalizations. Gender/classifier marking on nouns derives lexical nominalizations (nouns) with uniform denotations (e.g., Tuyuka *niri-go* (spouse-F.SG) ‘wife’; *ho-ñō* (banana-CLF_{STALK}) ‘banana tree’, *ho-pũ* (banana-CLF_{LEAF}) ‘banana leaf’). In contrast, marking on numerals and other structures yields grammatical nominalizations denoting a variety of things (e.g., *si’-ko* (one-F.SG) ‘one woman, one hen, one female tapir, etc.’; *pairi-ga* (big-CLF_{SPHERICAL}) ‘big ball, big head, big toe, etc.’).

2.4. Classifying nominalizations

Shibatani’s recent work (2021, 2023) centers on *classifying nominalizations*—nominal forms derived through gender or classifier marking. This is exemplified by the German examples discussed above, where the feminine suffix *-e* yields nouns denoting entities assigned to the feminine class, and by the Tuyuka examples, where classifier markers derive nouns categorized according to animacy features and other classificatory criteria. The concept of classifying nominalizations can be illuminated by contrasting them with (ordinary) non-classifying nominalizations formed by the English suffix *-er* (as seen earlier), and with the gender-based derivations illustrated by the Portuguese examples in (4) below:

¹⁴ Unlike Portuguese and some other languages, only nouns take the articles *the/a* in English.

(4) Classifying and non-classifying nominalizations

- a. English *kill* > *kill-er* (non-classifying, denoting either a masculine or feminine killer)
- b. Portuguese *assassin* > *assassin-o* (classifying, denoting a male killer)
assassin > *assassin-a* (classifying, denoting a female killer)

Paralleling these lexical classifying nominalizations, grammatical nominalizations may also be classifying, as in most of the pertinent examples in this paper. This can also be best illustrated by the comparison between ordinary, non-classifying nominalizations and classifying nominalizations. A particularly revealing case is the Tukanoan language Kotiria (aka Wanano), which we examine in some detail below due to its rich array of nominalization strategies.

(5) Kotiria classifying and non-classifying nominalizations¹⁵

- a. [[yi'i nhii-**ri**] phiri-wi/ki/pu hi-re
1SG saw-NZR big-CLF_{BUILDING}/CLF_{TREE}/CLF_{BASKET} COP-V.PERF.N1
'The one/what I saw was (a) big (house/tree/basket).'
- b. [[yi'i nhii-**wi**] phiri-wi/*ki/*pu
1SG saw- CLF_{BUILDING} big-CLF_{BUILDING}/CLF_{TREE}/CLF_{BASKET}
hi-re
COP-V.PERF.N1
'The one (house) I saw was (a) big (house).'
- c. [[yi'i nhii-**ki**] phiri-*wi/ki/*pu hi-re
1SG saw-CLF_{TREE} big-CLF_{BUILDING}/CLF_{TREE}/CLF_{BASKET} COP-V.PERF.N1
'The one (tree) I saw was (a) big (tree).'
- d. [[yi'i nhii-**pu**] phiri-*wi/*ki/pu
1SG saw-CLF_{BASKET} big-CLF_{BUILDING}/CLF_{TREE}/CLF_{BASKET}
hi-re
COP-V.PERF.N1
'The one (basket) I saw was (a) big (basket).'

Sentence (5a), which contains an ordinary, non-classifying nominalization marked by the suffix *-ri* as a subject, may refer to various types of (inanimate) entities¹⁶ and allows predicate expressions marked by different classifiers. In contrast, sentences (5b)–(5d) feature classifier-marked nominalizations as subjects, each denoting a specific range of objects that fall within a particular classifier category. As a result, these subjects restrict the predicate expressions to

¹⁵ Data from Tukanoan and other languages are based on our work, unless a specific source is cited.

¹⁶ For animate entities, speakers prefer using a classifying nominalization, marked by animate classifiers, such as M.SG (*-ki*), F.SG (*-ko*), and their plural form (*-ra*), indicating the preferential grammatical marking of more pressing information.

those marked by the same classifiers specified in the nominalized subjects. It is these classifying nominalizations marked by classifiers that are the main topic of this paper¹⁷.

2.5. Two uses of grammatical nominalizations

Shibatani’s nominalization framework recognizes two primary functions of nominalized structures, applicable to both ordinary and classifying types. The first is **NP use**, in which a nominalized structure serves as the head of a noun phrase. This is exemplified in the Kotiria sentences in (5) above, where the nominalizations function as the heads of subject noun phrases. Recall Bruil’s characterization of the NP use of nominalizations in Ecuadorian Siona in Section 1.2; “since [nominalizations] are nominal, nominalized verbs can head a noun phrase themselves. This is how these forms are predominantly used. They occur mostly as the head of a noun phrase without any head noun” (Bruil 2014: 137).

The same nominalization structures in NP use may be used as a modifier in their modification use, as in the Kotiria examples in (6) below¹⁸:

(6) Kotiria

- a. [[**yí’i nhii-ri**] wí’i] phiri-wi hi-re
 1SG saw-NZR house big-CLF COP-V.PERF.N1
 ‘The house I saw was (a) big (one/house).’
- b. [[**yí’i nhii-wi**] wí’i] phiri-wi hi-re
 1SG saw-CLF house big-CLF COP-V.PERF.N1
 ‘The house I saw was (a) big (one/house).’

In (6b) we see morphological agreement between the head noun and a modifier as a byproduct of the modification use of classifying nominalizations and the grammaticalization of the noun *wí’i* ‘house’ into a classifier in its truncated form.

The two uses of grammatical nominalizations are associated with distinct functions. In the modification use, their denotations (e.g., ‘things I saw’ or ‘things belonging to the building class that I saw’) serve a restrictive function—narrowing the interpretation of the head noun (not any kind of house but the one that I saw)—or a non-restrictive, identification function, where the nominalization helps identify the referent of the head noun without limiting its scope (the house—i.e., the one I saw—).

In NP use, nominalizations play a referring function of pointing out an entity in the speech context—just as ordinary nouns do when they head a noun phrase. What distinguishes grammatical nominalizations from ordinary nouns, however, is the nature of the entities they denote or refer to. As noted earlier, (lexical) nouns typically denote relatively uniform, well-

¹⁷ The contrasting patterns of classifying and non-classifying nominalizations illustrated here are also widely seen among East Asian languages such as Zauzou (Tibet-Burman), Hmong (Yao-Miao), and Southern Min and Yue dialects of Chinese—see Shibatani (2023).

¹⁸ However, forms used as an NP-head may require additional marking—such as an article, a demonstrative, or a classifier—and may not be directly usable as modifiers, depending on the language. In such cases, these additional elements function as NP-use markers, and the resulting structure is not compatible with use as a modifier.

defined entities. In contrast, grammatical nominalizations can refer to a wider range of entities that are metonymically linked to the meaning of the nominalized structure—in our examples, the various (building-class) entities metonymically evoked by the state of affairs of my seeing.

The classifier-marked nominalizations as in (5b)–(5d) denote a narrower range of entities (e.g., the building-class entities that I saw) than ordinary, non-classifying nominalizations like (5a), which in theory can denote anything that I saw, as shown by the various possibilities of the predicate forms. In actual discourse these denotations are narrowed down or pinpointed through the Gricean Cooperative Principle, which speakers are expected to follow to make their contributions relevant and informative. This is exactly what we do when we use ordinary metonymies, such as “Drink **a glass** a day”. In the context of a doctor prescribing a calcium supplement, *a glass* would be understood to mean the calcium drink (in a glass). However, in a discussion about the health benefits of red wine, the same phrase would be interpreted as referring to red wine.

The clarification of the metonymic basis of nominalizations and their contextual interpretations above is needed, as most—if not all—studies on Amazonian languages, as well as those dealing with other languages, attribute an anaphoric function to the NP use of grammatical nominalizations¹⁹. For instance, Aikhenvald (2000: 329), in a broader context, states that “[c]lassifiers of any type can be used as anaphoric and participant-tracking devices”. Such a view, however, is shortsighted, as classifying nominalizations can appear in contexts where true anaphors cannot—such as in news headlines or at the beginning of a conversation or narrative, where no antecedent exists for the so-called “anaphoric classifiers.” Consider the following examples from two Eastern Tukanoan languages.

(7) Kotiria (Stenzel 2004: 410, 419)

- a. ~ku-ta hi-a [~ku-iro ~bu-ro] to
 one/a COP.V.NOM one/a-NOM.SG man-SG 3SG.POSS
 ~pho'da--be're
 children-COM/INST
 wa'i~kida ko'ta-ro wa'a-a ~bakaka-pu.
 animal-PL wait-V.NOM go-ASSERT.PERF forest-LOC
 ‘Once a man went hunting with his children in the forest.’
- b. [~ku-iro] hi-ati-ga to ~ pho'da
 one/a-NOM.SG COP-IMPERF-ASSERT.PERF 3SG.POSS children
 ~ dabo-ro hiphiti-ro.
 wife-SG everyone/thing-PART
 ‘Once there was a man, his children, wife, everything.’

¹⁹ Others working on other languages often use the term “noun substitute” as in the Conklin quote in Section 1.2.

(8) Desano

- a. gia-re dẽ [yuhu-gi] dore-biri-bĩ
 1X-SPC first one-M.SG order-NEG-3M.SG
 ‘No one ordered us (to come).’ Lit. ‘One (masculine, singular entity) did not order us.’
 (Miller 1999: 124)

- b. [yuu--du] yuu ~yeku ~budu kadi ~budu
 one-CLF_{DAY} 1SG grandfather deceased Candido deceased
 wai-pudi-waa-bu
 hunt-blow-go-N3.PERF
 ‘One day, my grandfather Candido went hunting (with a blow gun).’ (Silva 2012: 279)

Kotiria sentence (7a) contains a modification use of the numeral-based classifying nominalization [~ku-iro] (one-masculine, singular). The same form appears as the head of a noun phrase in (7b), which opens a different story and lacks any antecedent to establish its referent. Comparable uses of numeral-based classifying nominalizations referring to indefinite entities are found in the Desano examples in (8)²⁰.

The same is true with classifying nominalizations derived by gender marking, as seen in Romance languages. Consider the following Portuguese news headline, where a gender-marked nominalization is used without an antecedent.

(9) Portuguese

- Acidente deixa [duas] mortas e 24 feridos
 accident leaves two.F dead and 24 injured
 ‘Accident leaves two [feminine entities] dead and 24 injured.’

A broader theory of metonymic interpretation—possibly along the line like the Gricean Cooperative Principle mentioned above—that accounts for such cases as the above would also encompass the phenomena commonly described as the anaphoric and reference-tracking functions of classifying nominalizations. What is at play here is not a grammatical operation or device such as noun substitution or antecedent–anaphor binding, but rather a discourse-pragmatic inference based on grammatical forms and the broader discourse context.

For example, in cognitive linguistics, metonymy is recognized as a pervasive conceptual mechanism that operates across various levels of language, including grammar and discourse. It functions as a “natural inference schema”, guiding discourse-pragmatic inferencing in interaction with general pragmatic principles, such as Gricean pragmatics and relevance theory—see the contributions in Panther and Thornburg (2003).

2.6. Development and distribution of classifier marking

Shibatani’s theory of Dynamic Functional Typology provides a comprehensive framework for understanding the evolution and distribution of gender and classifier marking

²⁰ For parallel cases in Indo-Aryan and East Asian languages, see Shibatani (2021).

across languages. This approach emphasizes the interplay between usage patterns and diachronic development, moving beyond static typologizing of constructions and languages. Central to this theory is a two-dimensional hierarchy (Diagram 1), integrating structural and functional dimensions, which serves as a predictive model for the emergence and spread of gender and classifier marking.

Structural Dimension: Numerals > Demonstratives > Genitives > Adjectives >
Verbal-based Nominalizations

Functional Dimension: NP Use > Modification Use

The proposed hierarchy posits that classifier/gender marking typically originates in the domain of NP use of numeral-based nominalizations and may extend, on one hand, to the modification domain, and on the other, to other grammatical structures such as demonstratives, genitives, adjectives, and verbal-based nominalizations, in a sequential manner. These constrained patterns of development form the basis for a predictive model of the diachronic development and synchronic distribution of classifier and gender marking, accounting for both language-internal and crosslinguistic variations.

This framework, which seeks diachronic explanations for variation in marking patterns, closely aligns with the concept of grammaticalization. Grammaticalization manifests both in the expansion of a form's structural and functional domains and in the increasing obligatoriness of grammatical markers. Synchronic crosslinguistic variation can thus be interpreted as reflecting different stages or rates of grammaticalization. In what follows, we examine how Shibatani's framework of Dynamic Functional Typology facilitates the crosslinguistic analysis of classifier systems in Amazonian languages.

3. TUKANOAN LANGUAGES

The Tukanoan language family, also known as Tucanoan or Tukánoan, comprises languages spoken across Colombia, Brazil, Ecuador, and Peru. According to Ramirez (1997), the family includes 11 languages, of which two are considered extinct, as in Appendix 1 (see also Barnes (1999), identifying 18 living languages). This study primarily focuses on languages within the Eastern Tukanoan branch, specifically those belonging to the Tukanid group. Occasional references are made to languages from other branches and groups, which are color-coded in Appendix A.

Several typological features stand out in Tukanoan languages. In phonology, Central and Eastern Tukanoan languages have a pervasive nasalization phenomenon that divides the lexicon into oral and nasal morphemes: in the former, all vowels are oral; in the latter, all vowels are nasal. Suffixes unspecified for nasality undergo nasalization through rightward spread, as described in Barnes and Takagi (1976) for the Tuyuka system.

Certain data drawn from published materials may indicate this distinction using a tilde to mark nasal morphemes, e.g., Desano *ba* [baa] 'to eat' vs. *~ba* [mãã] 'path' (Silva 2012: 56). In our study focusing on the grammar of classifiers, we employed broad phonetic transcription; forms like *~ba* are transcribed as [mãã], without explicit notation of the underlying morphemic

nasalization. Given our primary focus, we did not systematically analyze nasal and accentual phenomena; therefore, our transcriptions may not fully capture these suprasegmental features.

On the grammar front, Tukanoan languages have a complex system of evidential marking, conveying information on (a) the person and number of the sentence subject, (b) tense, and (c) one of the five evidential categories: (i) visual, (ii) non-visual, (iii) apparent, (iv) secondhand, and (5) assumed. Barnes (1984: 253) illustrates this by the following examples in the past tense:

- (i) *díiga apé-wi*
soccer play-EVI.PST
'He played soccer.' (I **saw** him play.)
- (ii) *díiga apé-ti*
'He played soccer.' (I **heard** the game and him, but I didn't see it or him.)
- (iii) *díiga apé-yi*
'He played soccer.' (I **have seen evidence** that he played: his distinctive shoe print on the playing field. But I did not see him play.)
- (iv) *díiga apé-yigi*
'He played soccer.' (I obtained the **information from someone else**.)
- (v) *díiga apé-hīyi*
'He played soccer.' (It is **reasonable to assume** that he did.)

Again, the evidential system has no direct bearing on our study of classifiers.

Tukanoan and many other Amazonian languages, like New Indo-Aryan and Dravidian, make extensive use of nominalized constructions in predication, often blurring the line between nominal and verbal predication. For instance, distinctions such as that between 'I am a student' and 'I study' are frequently not formally distinguished, as illustrated by the following Tukano and Barasano examples.

(10) Tukano

- a. *ati'-go nũmĩõ bue-gó ni-mõ*
this-F.SG woman study-F.SG COP-V.P.F.SG
'This girl is a student.' (Or 'This girl studies'.)
- b. *ati'-go nũmĩõ (unicamp pu) bue-gó ni-mõ*
this-F.SG woman Unicamp EMPH study-F.SG COP-V.P.F.SG
'This girl studies (at the University of Campinas).' (Or 'She is a student at Unicamp'.)

(11) Barasano (Jones and Jones 1991: 22)

- a. *riha-go yã-a-mõ*
sick-F.SG COP-PRS-3F.SG
'She is a sick person.'

- b. riha-go bahi-a-mõ
 sick-F.SG COP-PRS-3F.SG
 ‘She is sick.’

Jones and Jones (1991) offer two distinct interpretations of examples (11a.b)—their (55) and (58)—depending on the nature of the copulas involved. The copula *yã* in (11a) denotes a permanent state, whereas *bahi* in (11b) describes a temporary state, but both predicates are formally verbal-based nominalizations that literally denote a third person feminine entity. One plausible speculation about these constructions is that subject-verb agreement originates from the use of gender-/classifier-marked nominalized structures as predicates, followed by the eventual loss of the copula²¹.

Eastern Tukanoan languages are generally OV in word order with S either in initial or final position. OVS languages such as Tatuyu typically have SOV as an alternate order, except for Barasano, which is strictly OVS, the word order, O before S, frowned upon by word order typologists.

These differences in basic word order may have implications for the structure of modification constructions, particularly regarding the position of modifiers relative to their heads. Tuyuka, the primary language examined in our study, follows the SOV order and permits both prenominal and postnominal modification. However, postnominal modifiers—aside from verb-based nominalizations (i.e., relative clauses)—tend to behave more like appositive constructions. These typically require classifier-marked NP-use forms. In contrast, inanimate prenominal modifiers may appear with or without classifier marking, allowing for optionality in this position. Whether the strictly OVS language Barasano has any other marking patterns different from those of SOV languages deserves scrutiny.

Finally, nominal classifiers are a prominent typological feature of Tukanoan languages, with significant variation in the number across different branches of the family. In Western Tukanoan languages, the number of classifiers typically ranges from 17 to 28. Central Tukanoans exhibit a broader range, with reported counts between 21 and 100. Eastern Tukanoan languages, such as Tuyuka, Desano, and Barasano, possess particularly extensive classifier inventories, ranging from approximately 50 to 140 distinct forms (Barnes 1999: 219). These figures generally exclude “repeater classifiers”. In languages like Kubeo (Central Tukanoan), repeater classifiers are said to be an open class (Chacon 2022), further expanding the system’s flexibility and semantic reach. The nature of Tukanoan classifiers is further explored through the Tuyuka system at the beginning of Section 5.

4. METHODOLOGY

In addition to published works and publicly available materials such as doctoral dissertations, the present study draws on the elicitation sessions with native speakers we conducted in the Brazilian cities of Campinas and Manaus in December 2024. The primary site

²¹ Cf. Miraña (Witotoan; Seifart 2005), where a copula has dropped out: *ná:ni tsa:-:bẽ* [my.uncle come-GCM.masc.sg] ‘My uncle came.’; *ái:βẽ-ha mũhu:-ha* [burn.SUB.SCM.cover big-SCM] ‘What (house, clothes, etc.) burned down was big.’ (GCM=general CLF, SCM=specific CLF) Copulas of such constructions similarly drop out in Tukanoan languages in causal speech.

Manaus is situated in the heart of the Amazon rainforest, near the confluence of the Negro and Solimões (upper Amazon) rivers.

The initial phase of our investigation involved eliciting phrases and sentences from native speakers that were semantically equivalent to contextualized Portuguese prompts. For example, speakers were presented with Portuguese sentences embedded in specific contexts, as below, and asked to provide corresponding expressions in their language.

Context	Responses/follow-up's	
Quantos homens você vê? How many men do you see?	NP use	Mod use
	(Eu vejo) Um.	Um homem.
	(I see) One.	One man.
	si' -ki	si' -ki imi
	*si'	*si' imi
	Dois.	Dois homens.
	'Two'	'Two men'
	piã-rã	piã-rã imiã
	*piã	*piã imiã

After obtaining initial responses for singular and plural forms—such as in the Tuyuka session cited above—we systematically examined the acceptability of all logically possible combinations. This helped identify both acceptable and unacceptable constructions, clarifying constraints on classifier marking: whether it is obligatory, optional, or prohibited in the two main usage contexts—NP use and modification use. We then followed up with targeted questions on the constructions under investigation, such as numeral-based and demonstrative-based nominalizations, to better understand their syntactic behavior and semantic functions. After leaving the field sites, we continued online sessions with several speakers.

The breakdown of the hours we spent on the languages of our investigation reported in this paper is as follows:

Tuyuka (a female Tuyuka-Tukano-Portuguese speaker born in 1985)
40 hours

Tukano (a male Tukano-Tuyuka-Portuguese speaker with knowledge of Kotiria and Wa'ikhana, born in 1992; a female Tukano-Portuguese speaker with knowledge of no other indigenous languages, born in 1994; and the above-mentioned female Tuyuka-Tukano-Portuguese trilingual speaker)
20 hours

Kotiria (a female Koritia-Tukano-Portuguese trilingual speaker, born in 1989)
12 hours

Wa'ikhana (a female Wa'ikhana-Portuguese bilingual speaker, born in 1996)
5 hours

This study first examines the Eastern Tukanoan language Tuyuka in detail (Section 5) to provide a comprehensive account of classifier-marking patterns in grammatical nominalizations within a single language. Sections 6 and 7 then extend the analysis through a crosslinguistic investigation of the development and distribution of classifier marking in languages from Western, Central, and Eastern Tukanoan branches.

5. TUYUKA

5.1. Tuyuka classifiers

The Tukanoan classifier systems distinguish between (i) animate and inanimate entities. The animate classifiers differentiate between (ii) singular and plural, and within the singular category, (iii) masculine and feminine. Inanimate classifiers based on shapes and other physical properties distinguish between (iv) singular and plural. Table 2 illustrates these distinctions in Tuyuka.

Table 2. Animate and inanimate classifiers in Tuyuka

Animate M.SG si'-ki bue-gi one-M.SG study-M.SG 'one (male) student'	Animate F.SG sĩ'-kõ bue-go one-F.SG study-F.SG 'one (female) student'	Animate PL pã-rã bue-ra two-ANI.PL study-ANI.PL 'two students'
Inanimate CLF(SG) si'ka-gi yuki-gi one-CLF _{TREE/STICK} tree-CLF 'one tree'		Inanimate PL ba'parí yuki-rí four tree-INA.PL 'four trees'

Number marking is not entirely consistent. Some numerals are marked with the plural classifiers but some others are not. While plural marking on animate nouns seems to be consistent, plural marking on inanimate nouns is not. For example, *pia-gi yuki-gi* 'two trees', mark neither the numeral-CLF construction nor the noun with the inanimate plural classifier *-ri*. See Section 5.2 for further examples of number marking on numerals and the head nouns.

While animate classifiers are limited to three as in Table 2, Tuyuka has numerous inanimate classifiers, like some other Eastern Tukanoan languages such as Barasano and Tatuyu. Barnes (1990: 257ff) lists 99 Tuyuka inanimate classifiers "divided into ten semantic categories: shape [e.g., *-pólo* CYLINDICAL], collection [*-ba* PAIR], arrangement [*-betó* COIL], anatomical [*-bõ* ARM], botanical [*-gi* TREE/STICK], geographical [*-bã* PATH/RIVER], container [*-pi* DEEP.BASKET], manufactured item [*-ro/-rõ* DEVICE/MACHINE], consistency [*-kebéro* CONSISTENCY OF MUD] and time [*-to* TIME OF AN EVENT/SEASON]".

In addition to these, a subset of nouns functions as repeater classifiers, e.g., *-sope* 'DOORWAY' and *-kopé* 'HOLE.' These repeat the nouns they modify, as in *sopé* [*siká-sope*] (door [one-CLF]) 'one doorway.' Barnes (pp. 243–244) makes the point that all Tuyuka classifiers, **including repeaters**, are affixal and form a single phonological word with the root

or stem they attach to²². This affixal nature of repeaters contrasts with those in the Central Tukanoan language Kubeo, which Chacon (2022: 35) analyzes as free morphemes. If this difference holds, it may reflect varying degrees of grammaticalization of repeaters between the two languages.

Similar to the Tatuyu illustration in Table 1, Tuyuka classifiers form a morphosemantic continuum. At one end, some classifiers—such as animate classifiers and many inanimate ones—lack any morphological connection to nouns or other sources. At the other end, (full) repeater classifiers are identical in form to their corresponding nouns. Between these extremes, there are classifiers that exhibit varying degrees of phonological resemblance to their source nouns, such as *-wi* (BUILDING) and *wii* (‘house/building’), or *-ho* and *hoó* (‘banana’).

Semantically, animate classifiers, such as the masculine (*-kĩ*) and feminine (*-kõ*) markers, do not inherently link their forms to the concepts of masculinity and femininity. Similarly, many classifiers—such as *-pĩ* (RIGID) and *-ro* (FLEXIBLE)—which have lost their connection to source nouns, are like animate classifiers in this regard. In contrast, repeater classifiers maintain a clear semantic link between their category meanings and their source nouns. For example, *-kopé* (HOLE.CLASS) corresponds directly to *kopé* ‘hole’, and *-wesé* (FIELD.CLASS) to *wesé* ‘field’.

Between these two extremes of semantic and morphological transparency lie many partial repeater classifiers, such as *-pi* (DEEP.BASKET.CLASS) and *pii* ‘deep basket’, or *-di* (MEAT.CLASS) and *dii* ‘meat’, which exhibit partial morphological resemblance and a close semantic relationship to their source nouns. Within themselves, partial repeaters exhibit a grammaticalization cline, where some classifiers have generalized their category meanings beyond those of the source noun. For instance, *-pũ* (LEAF.CLASS), derived from *pũũ* ‘leaf’, embraces other leaf-like objects than leaves, such as a book. Similarly, *-wa* (GOURD.CLASS), from *waá* ‘gourd’, refers more broadly to shallow containers with a wide mouth (see Barnes 1990 for further details).

We now turn to the data and analysis of Tuyuka classifier-marking patterns, beginning with numerals and proceeding through demonstratives and genitives, following the order of the Shibatani hierarchy.

5.2. Numerals

Amazonian numeral systems are well known for their limited repertoire²³. Speakers of various Amazonian languages typically provide basic numerals only up to three or four, and at most five, often resorting to terms like ‘many’ for quantities beyond that. This pattern is evident in the Tuyuka numerals shown in (12), where the basic numerals extend up to four. Numerals for five and higher rely on the concept of ‘one hand (with five fingers)’ and references to the fingers on the other hand.

²² Miller (1999: 36) also notes that “[n]ouns in Desanao that do not have a representative classifier function as classifiers in **that they attach to the same roots and stems as a classifier to form a single phonological descriptive word**” (e.g., *iri-gobe* ‘this-HOLE’) (emphasis added). Cf. these observations with Corbett’s, informing him that: “**Classifiers are independent items**, selected largely according to semantic criteria, while **gender markers typically appear attached to agreement targets**.” (Corbett 1991: 137; emphasis added)

²³ See Frank et al. (2008) for the controversies surrounding the Amazonian numeral systems.

(12) Tuyuka numerals

si'kaga 'one'
 piaga 'two'
 ti'tiaga 'three'
 ba'pari 'four'
 si'ka mo'kañẽ 'five' (one hand—5 fingers²⁴)
 si'ka mo'kañẽ si'kaga 'six' (5 fingers plus 1 finger from the other hand)
 si'ka mo'kañẽ piaga 'seven' (5 fingers plus 2 from the other hand)

Speakers of Tukanoan languages typically report difficulty in conceptualizing numerical values independently of the objects they quantify. Instead of using numeral roots to denote abstract quantitative concepts like 'one', 'two', or 'three', they prefer expressing quantities through specific objects, such as 'one fish', 'two fish', or 'three fish', or by using numeral roots marked with classifiers. For instance, when asked how to say 'Four is larger than two', a Tukano speaker offered forms incorporating these types of information, as in (13). In Tuyuka, the numerals one through three are marked with the classifier *-ga*, which denotes three-dimensional objects.

(13) Tukano

ba'pariti-rã	(waia)	pia-rã	(waia)	yiroro	nima
four-ANI.PL	fish	two-ANI.PL	fish	more	COP

'Four (fish) is more than two (fish).'

This intriguing anecdote may support Shibatani's (2023: 43) reasoning that numerals are the first structures to attract classifiers (see Diagram 1). He hypothesizes that we all learn numerals through concrete objects representing quantities, such as images of one apple, two apples, and three apples to illustrate the numerical values of 'one', 'two', and 'three'²⁵. As a result, numerals maintain the strongest metonymic connections to physical objects, making them particularly suited for deriving classifying nominalizations.

Tuyuka classifier marking on numerals varies depending on the animate/inanimate and the singular/plural distinction. Like some other languages (e.g., Japanese²⁶), lower numerals are more consistently classifier-marked, as observed in (14) and (15).

(14) Tuyuka animates

Numeral 1 (M/F.SG marking on NUM and optionally on N)

a.	si'-ki	imi
	one-M.SG	man

'one/a man'

²⁴ The word *mo'kañẽ*, however, does not mean either hand or finger in Tuyuka, which are respectively *wamõ* and *wamõsõã*.

²⁵ Cf. the Chinese letters/characters 一, 二, 三 for one, two, and three.

²⁶ In Japanese numerals up to ten require a classifier when denoting objects; beyond ten, however, classifier marking becomes increasingly optional.

a'. si'-kõ nũmĩõ
 one-F.SG woman
 'one/a woman'

b. si'-ki kãrĩkẽ
 one-M.SG chicken
 'one/a rooster'

b'. si'-kó kãrĩkẽ(-kõ)
 one-F.SG chicken-F.SG
 'one/a hen'

Numerals 2-3 (ANI.PL marking on both NUM and N)

c. pĩã-rã kãrĩkẽ-ã
 two-ANI.PL chicken-ANI.PL
 'two chickens'

d. ĩhtĩa-rã kãrẽke-ã
 three-ANI.PL chicken-ANI.PL
 'three chickens'

Numerals 4-5 (ANI.PL marking only on N)

e. bahparĩ kãrẽke-ã
 four chicken-ANI.PL
 'four chickens'

f. sihká mohkãñe kãrẽke-ã
 one hand? chicken-ANI.PL
 'five chickens'

(15) Tuyuka inanimates

Numerals 1-3 (INA.CLF on NUM, and on some N)

a. si'ka-gĩ yukigĩ
 one-CLF_{TREE} tree.CLF
 'one/a tree'

b. pi a-gĩ yukigĩ (plural but -gĩ on N)
 two-CLF tree.CLF
 'two trees'

- c. ti'tia-gi yukigi (plural but -gi on N)
 three-CLF tree.CLF
 'three trees'

Numerals 4-5 (INA.PL on N)

- d. ba'parí *(-gi) yuki-rí²⁷
 four *(-CLF) tree-INA.PL
 'four trees'
- e. si'ká mo'kañé *(-gi) yuki-rí
 one hand? *(-CLF) tree-INA.PL
 'five trees'

5.2.1. Classifier marking of numerals in NP use and Modification use

Like other Tukanoan languages we have investigated, Tuyuka animate-classifier marking on numerals in NP use of numerals 1-3 is consistently obligatory in both NP and modification use.

(16) Tuyuka animates

	NP use	Mod use
Quantos homens/melheres você vê? 'How many men/women do you see?'	(Eu vejo) (I see) Um/Uma. 'One.'	Um(a) homem/melher. 'One man/woman.'
Masculine	*si' 'One.'	*si' imi 'One man.'
	si'-ki one-M.SG 'One.'	si'-ki imi one-M.SG man 'One man.'
	Dois. 'Two.'	Dois homens. 'Two men.'
	*piã two piã-rã two-ANI.PL	*piã imi-ã two man-ANI.PL piã-rã imi-ã two-ANI.PL man-ANI.PL

²⁷ The notation *(X) indicates that X is not optional and cannot be used in this context.

Feminine	*si'	*si'	nũmĩõ
	one	one	woman
	‘One.’	‘One woman.’	
	si’-kõ	si’-kõ	nũmĩõ
	one-F.SG	one-F.SG	woman
	‘Two.’	‘Two women.’	
	pĩã-rã	pĩã-rã	nũmĩ-ã
	two-ANI.PL	two-ANI.PL	woman-ANI.PL

Similar to the obligatory marking of animate classifiers on numerals described above, inanimate nominalizations require obligatory classifier marking when they are used as an NP head. In modification contexts, however, the marking pattern shifts: classifier marking is optional in prenominal modification PrN (the more common word order), but becomes obligatory in postnominal modification PoN (a less preferred order). Consider the following examples:

(17) Tuyuka inanimates

Quantas árvores você vê?	Uma.	Uma árvore.
‘How many trees do you see?’	‘One.’	‘One tree.’
	*si’ka	si’ka yukigi
	one	one tree
	si’ka-gi	si’ka-gi yukigi
	one-CLF _{TREE}	one-CLF tree
	Duas	Duas árvores.
	‘Two.’	‘Two trees’.
	*pia	pia yukigi *yukigi pia
	pia-gi	pia-gi yukigi yukigi pia-gi
	two-CLF	two-CLF tree

The following tables summarize classifier marking on numerals in Tuyuka.

Table 3. Tuyuka animate classifier marking on numerals

Animate NUM		
NP use	OBL	
Mod use	PrN OBL	PoN OBL

Table 4. Tuyuka inanimate classifier marking on numerals

Inanimate NUM		
NP use	OBL	
Mod use	PrN OPT	PoN OBL

Tuyuka contrasts with Tukano (Eastern Tukanoan) and Retuarã and Tanimuka (both Central Tukanoan) in that, in the latter languages, inanimate classifier marking is also obligatory in prenominal modification. These observations on classifier marking in numeral constructions already demonstrate that animate classifiers—analyzed by some as gender, distinct from inanimate classifiers—and inanimate classifiers do not constitute entirely separate or segregated systems. In Tuyuka, the two converge in NP-use and post-nominal modification contexts, where both animate and inanimate classifiers are obligatorily employed. However, the two diverge in prenominal modification; animate classifiers are obligatory, while inanimate classifiers—lagging in grammaticalization—remain optional. In contrast, Tukano, Retuarã, and Tanimuka exhibit full convergence in classifier marking on numerals, with both animate and inanimate classifiers obligatorily marked in both NP- and modification-use contexts—observe the Tukano pattern in (18):

(18) Tukano

Animate

Quantas mulheres você vê?
‘How many women do you see?’

Uma.
‘One.’
*ni’
one
ni’-kõ
one-F.SG

Uma mulher.
‘One woman.’
*ni’ numioñã
one woman
ni’-kõ numioñã
one-F.SG woman

Duas.
‘Two’
*piã
piã-rã
two-ANI.PL

Duas mulheres.
‘Two woman’
*pia numiãñã
piã-rã numiãñã
two-ANI.PL woman

Inanimate

Quantas árvores você vê?	Uma.	Uma árvore.
‘How many trees do you see?’	‘One.’	‘One tree.’
	*ni’	*ni’ yukigi
	ni’-ki	ni’-ki yukigi
	one-CLF _{TREE}	one-CLF tree
	Duas.	Duas árvores.
	‘Two’	‘Two trees.’
	*pia	*pia yukigi
	pia-gi	pia-gi yukigi
	two-CLF	two-CLF tree

5.3. Demonstratives

Tuyuka demonstratives, *a’ti-* ‘this’ and *i-* ‘that’, combine with both animate and inanimate classifiers. The morphologically regular animate masculine forms *a’ti-gi* (this-M.SG) and *i-gi* (that-M.SG) have lexicalized as *ã’nĩ* ‘he near the speaker’ and *ĩnĩ* ‘he away from the speaker’, while the feminine forms retain the regular morphological makeup as *a’ti-go* (this-F.SG) and *i-go* (that-F.SG) ²⁸. Their plural forms follow a regular pattern: as *ã’nĩ-ã* ‘these near the speaker’ and *ĩnĩ-ã* ‘those away from the speaker’.

5.3.1. Classifier marking on demonstratives in NP use and Modification use

Animate-classifier marking on demonstratives are already shown above; and it is obligatory in both NP and modification use, as shown in (19):

(19) Tuyuka animate demonstratives

Qual deles é seu filho?	Este.	Este garoto.
‘Which one of them is your son?’	‘This.’	‘This boy.’
	ã’nĩ	ã’nĩ ma’ki ma’ki ani
	this.M.SG	this.M.SG child
	*a’ti	*a’ti ma’ki
	*a’ti-gi	*a’ti-gi ma’ki
	this-M.SG	this-M.SG child
Qual deles é sua galinha?	Esta.	Esta galinha.
‘Which one of them is your hen?’	‘This.’	‘This hen.’
	*a’ti	*a’ti kārĩkē
	this	this chicken
	a’ti-go	a’ti-go kārĩkē kārĩkē a’ti-go
	this-F.SG	this-F.SG chicken

²⁸ The morphologically regular masculine forms are not available in Barasano either, according to Jones and Jones (1991).

(20) Tuyuka inanimate demonstratives

De qual casa você gosta?	Esta.	Esta casa.
Which house do you like?	‘This.’	‘This house.’
	*a’ti	a’ti wi’i *wii a’ti
	this	‘this house’
	a’ti-wi	a’ti-wi wii wii a’ti-wi
	this-CLF	this-CLF house

The classifier-marking patterns for demonstratives are the same as those for numerals. Animate-classifier marking is obligatory in both NP- and modification-use contexts. Inanimate-classifier marking, on the other hand, remain optional in prenominal modification; NP-use and postnominal modification call for obligatory marking. Tukano shows the similar patterns.

5.4. Genitives

Eastern Tukanoan genitives, as illustrated by the Tuyuka first-person possessor forms in (21), can be categorized into four morphological patterns based on the type of possessum: humans (21a–c), higher animals (21d–e), inanimates (21f), and plurals (21g–h).

(21) Tuyuka genitives

Humans: *yí’i* ‘1st person singular’

a. *yí’i ma’ki*

1SG son

‘my son’

b. *yí’i ma’kō*

1SG daughter

‘my daughter’

c. *yí’i buego*

1SG female.student

‘my female student’

Animals: *yáa*-M/F.SG

d. *yáa-gi*

1SG.GEN-M.SG

‘my rooster’

kārĩkẽ

chicken

e. *yáa-go*

1SG.GEN-F.SG

‘my hen’

kārĩkẽ

chicken

Inanimates: *yáa*-CLF

- f. *yáa-gi* *yukigi*
 1SG.GEN-CLF tree
 ‘my tree’

Plurals: *yi’i* (humans), *yé* (inanimates)

- g. *yi’i* *pona*
 1SG children
 ‘my children’
- h. *yé* *yukigi-ri*
 1SG.GEN.INA.PL tree-PL
 ‘my trees’

For human possessums, as in (21a–e) and (21g), the possessor forms in Tuyuka are identical with the personal pronouns set out in Table 5 and are not marked with classifiers; compare (21a) and (21b). For non-humans (animals and inanimates), possessor forms involve the nominalizer *yáa* for N-based nominalizations (Shibatani and Chung 2018). The resulting form for the first-person pronoun, *yi’i-yáa*, is reduced to *yáa*, as in (21d)–(21f), while other personal pronouns and nouns retain their basic forms, as *mĩ’ĩ-yá* ‘yours’, *kĩ’ĩ-yáa* ‘his’, *Tom-yáa* ‘Tom’s’. It is these forms for animals and inanimates that are classifier-marked, as exemplified in the next section.

Table 5. Tuyuka personal pronouns

1 st person	Singular	<i>yi’i</i>
	Plural	Exclusive <i>ĩ’sã</i>
		Inclusive <i>mã’ni</i>
2 nd person	Singular	<i>mĩ’ĩ</i>
	Plural	<i>’mãã</i>
3 rd person	Singular	Masculine <i>kĩ’ĩ</i>
		Feminine <i>ko’o</i>
	Plural	<i>’kĩã</i>

5.4.1. Tuyuka genitives in NP use and modification use

Humans		NP use	Modification use
(22) De quem é este filho? 'Whose is this son?'		Meu.	Meu filho.
		'Mine.'	'My son'
		*yi'i ²⁹	yi'i ma'ki ma'ki yi'i
		1SG	1SG son
		*yáa-gi	*yáa-gi ma'ki
De quem é esta filha? 'Whose is this daughter?'		Minha.	Minha filha.
		'Mine.'	'My daughter'
		*yi'i	yi'i ma'kō ma'kō yi'i
		1SG	1SG daughter
		*yáa-go	*yáa-go ma'kō
Human PL			
De quem são os filhos? 'Whose are these children?'		Meus.	Meus filhos.
		'Mine.'	'My children'
		*yi'i	yi'i pōnā pōnā yi'i
		1SG	1SG children
Animals De quem é esta galinha? 'Whose is this hen?'		Minha.	Minha galinha.
		'Mine.'	'My hen'
		*yáa	yáa kārīkē *kārīkē yáa
		yáa-go	yáa-go kārīkē kārīkē yáa-go
		1SG-F.SG	1SG-F.SG chicken
Animal PL De quem são as galinhas? 'Whose are these hens?'		Minhas.	Minhas galinhas.
		'Mine.'	'My hens.'
		*yáa	*yáa kārīkē
		yáa-ra	yáa-ra kārīkēã kārīkēã yáa-ra
		1SG.GEN-ANI.PL	
Inanimates		NP use	Modification use
(23) De quem é esta árvore? 'Whose is this tree?'		Minha.	Minha árvore.
		'Mine.'	'My tree.'
		*yáa	yáa yukigi *yukigi yáa
		1SG.GEN	1SG.GEN tree
		yáa-gi	yáa-gi yukigi yukigi yáa-gi
		1SG.GEN-CLF	1SG.GEN-CLF tree

²⁹ For the forms that do not take the *yáa* nominalizer, there is no way to form the equivalents of the English expressions like *mine*, *yours*, and *hers*. The best speakers can do is to express the possessums (overtly) as in the modification-use forms.

Inanimate PL

De quem são as árvores?	Minhas.	Minhas árvores.
‘Whose are these trees?’	‘Mine.’	‘My trees.’
	yé	yé yukigi yukigi yé
	1SG.INA.PL	1SG.INA.PL tree

As observed above, human genitives do not show classifier marking. Animal and inanimate genitives require classifier marking in NP use and postnominal modification constructions, while in prenominal modification, marking is optional. These marking patterns are consistent with numerals and demonstratives.

Among our Tukano consultants we observed divergent judgments regarding classifier marking in genitive constructions. The husband of a couple, fluent in Tuyuka and possessing some knowledge of Kotiria and Wa’ikhana, shows classifier marking patterns akin to those in Tuyuka. Conversely, his wife, who is two years younger and who speaks no other indigenous languages, consistently requires classifier marking in prenominal modification. Her usage aligns genitive marking patterns with those of numerals, wherein classifier marking is obligatory for both animate and inanimate nouns across both NP- and modification-use contexts.

5.5. Tukanoan “adjectives”

Barnes (1990) does not recognize property concept words (PCWs) or descriptive words as a distinct class of adjectives. Instead, she demonstrates that PCWs (e.g., *pai-* ‘be big’) and activity verbs (e.g., *ati-a-* ‘come-recently’) exhibit similar patterns in their interaction with the nominalizer *-ri* and inanimate-classifier marking. For instance, *pai-ri-sawero* (big-NZR-CLF_{EAR.SHAPED}) means ‘a big thing in the class of ear-shaped things’ and *ati-a-ri-bã* (come-recently-NZR-CLF_{PATH}) denotes ‘a path-class thing that one had recently come on’. Both Barasano (Jones and Jones 1991) and Desano (Miller 1999) are described as verb-adjective languages, meaning that “there is a single set of roots which may take either regular verbal inflection and serve as predicates, or may take nominalizers, participials, and classifiers, and serve as modifiers in noun phrases” (Jones and Jones 1991: 63).

Despite these general characterizations of Tukanoan PCWs, Barnes (1999) notes that Tuyuka roots for ‘old’ and ‘young’ take the same suffixes as nouns, indicating that Tukanoan PCWs are not entirely uniform. Furthermore, even those PCWs that behave like activity verbs do not take the full range of verb suffixes. Several Tukanoan studies also recognize a small group of PCWs that behave differently from (stative) verbs. Chacon (2012: 304ff), for example, identifies six PCWs in Kubeo (Central Tukanoan) that, among other properties, can directly modify a head noun without involving the nominalizer *-ri*, which is required of other property concept words (stative verbs) and (activity) verbs. They are: ‘big’, ‘small’, ‘new, young’, ‘old, great’, ‘good, beautiful’, and ‘bad, ugly’. Similarly, for Máihĩki (Western Tukanoan), Farmer (2015: 38, 154) identifies 11 “adjectival roots” that take neither nominal morphology (e.g., plural marking) nor verbal inflection and that can be directly suffixed with a classifier. The list includes: ‘white/shiny’, ‘red’, ‘yellow’, ‘black/dark’, ‘green/young’, ‘big’, ‘small’, ‘short’, ‘long/tall’, ‘good’, ‘new’.

The primary purpose of setting up a “Tukanoan adjectives” section in this paper is twofold: (i) to align this study with research on languages that have a distinct adjective class and (ii) to examine whether some of the words singled out by Barnes, Chacon, and Farmer differ from other PCWs in their use of the nominalizer *-ri* and classifier marking, and in the interaction between the two, which are major issues in the study of Tukanoan PCW- and verbal-based nominalizations, as detailed below.

5.5.1. Tuyuka PCWs in NP use and modification use

In studying PCW- and verbal-based nominalizations in Tukanoan languages, we must consider the tense/aspect distinction, which correlates with the use of the nominalizer *-ri*. As in Barnes and Malone’s (2000) summary table (see Appendix B), animate nominalizations appear to have lost the nominalizer *-ri* in the present/imperfective tense/aspect, while inanimate nominalizations retain it in all tense/aspect categories, similar to the animate nominalizations in the past, recent past, and future tenses. We shall discuss the idea on the loss of nominalizer *-ri* for the animate nominalizations in the present tense in Section 6.

In the following data on PCW-based **animate** nominalizations in (24), the form with the nominalizer *-ri* (*pai-ri-go* [big-NZR-F.SG] ‘a/the big feminine one’) receives the past interpretation, while the one without it (*pai-go* [big-F.SG] ‘a/the big feminine one’) receives the present interpretation.

(24) Animate CLFs	NP use	Modification use
Qual é a sua galinha? ‘Which is your hen?’	A grande. ‘The big one.’	A galinha grande. ‘The big hen.’
Present tense		
*pai ‘big’	*pai kārīkē big chicken	*kārīkē pai
*pai-ri big-NZR	*pai-ri kārīkē big-NZR chicken	*kārīkē pai-ri
*pai-ri-go big-NZR-F.SG	*pai-ri-go big-NZR-F.SG	kārīkē *kārīkē pai-ri-go chicken
pai-go big-F.SG	pai-go kārīkē big-F.SG chicken	kārīkē pai-go
Past tense		
*pai ‘big’	*pai kārīkē big chicken	*kārīkē pai
*pai-ri big-NZR	*pai-ri kārīkē big-NZR chicken	*kārīkē pai-ri
pai-ri-go big-NRZ-3F.SG	pai-ri-go big-NZR-3F.SG	kārīkē kārīkē pai-ri-go chicken
*pai-go big-F.SG	*pai-go kārīkē big-F.SG chicken	*kārīkē pai-go

A molhada.	A galinha molhada.	
‘The wet one.’	‘The wet hen.’	
*pua	*pua kārīkē	*kārīkē pua
wet	wet chicken	
*pua-ri	*pua-ri kārīkē	*kārīkē pua-ri
wet-NZR	wet-NZR chicken	
*pua-ri-go	*pua-ri-go kārīkē	*kārīkē pua-ri-go
wet-NZR-F.SG	wet-NZR-F.SG chicken	
pua-go	pua-go kārīkē	kārīkē pua-go
wet-F.SG	wet-F.SG hen	
‘The wet one.’	‘The wet hen.’	
Past tense		
*pua	*pua kārīkē	*kārīkē pua
wet	wet chicken	
*pua-ri	*pua-ri kārīkē	*kārīkē pua-ri
wet-NZR	wet-NZR chicken	
pua-ri-go	pua-ri-go kārīkē	kārīkē pua-ri-go
wet-NZR-F.SG	wet-NZR-F.SG	
‘The one that was wet.’	‘The hen that was wet.’	
*pua-go	*pua-go kārīkē	*kārīkē pua-go
wet-F.SG	wet-F.SG chicken	

A first point of the Tuyuka PCW marking-patterns to recognize is that the small class of non-stative verb PCWs identified by Barnes, Chacon, and Farmer is not distinguished in nominalizer marking for either animate or inanimate forms. Both stative verb PCWs (e.g., “wet”, “strong”, “quick”) and what are recognized as non-stative verb PCWs (e.g., “big”, “good”, “old”) take the *-ri* nominalizer, which is unexpected for non-stative verb PCWs. Additionally, both types of PCW roots allow direct (without *-ri*) animate-classifier marking for the present tense, and both disallow direct inanimate classifier marking, as revealed in the comparison of the animate marking patterns in (24) and those of inanimate marking patterns in (25) below³⁰.

³⁰ Both Chacon and Farmer implicate that stative verb PCWs do not take classifiers directly. We examined other PCW roots belonging to the two types as well and found that they are consistent with the patterns of the two PCW examples used in (24) and (25).

(25) Inanimate CLFs	NP use	Modification use
Qual é a sua árvore?	A grande.	A árvore grande.
‘Which is your tree?’	‘The big one.’	‘The big tree.’

Present/Past tense

*pai	*pai yukigi	*yukigi pai
big	big tree	
*pai-ri	pai-ri yukigi	*yukigi pai-ri
big-NZR	big-NZR tree	

pai-ri-gi	pai-ri-gi yukigi	yukigi pai-ri-gi
big-NZR-CLF	big-NZR-CLF tree	
*pai-gi	*pai-gi yukigi	*yukigi pai-gi
big-CLF	big-CLF tree	

A molhada.	A árvore molhada.
‘The wet one.’	‘The wet tree.’

Present/Past tense

*pua	*pua yukigi	*yukigi pua
wet	wet tree	
*pua-ri	pua-ri yukigi	*yukigi pua-ri
wet-NZR	wet-NZR tree	

pua-ri-gi	pua-ri-gi yukigi	yukigi pua-ri-gi
wet-NZR-CLF	wet-NZR-CLF tree	
*pua-gi	*pua-gi yukigi	*yukigi pua-gi
wet-CLF	wet-CLF tree	

The inanimate PCW nominalizations marked by *-ri*, in principle, allow both present and past interpretations according to Barnes and Malone’s summary table (Appendix B). The predominant interpretation for forms like *pai-ri-gi* [big-NZR-CLF_{TREE}] ‘a/the big tree-class one’ and *pua-ri-gi* [wet-NZR-CLF_{TREE}] ‘a/the wet tree-class one’ appears to be a present-tense reading. If demanded for the past reading, our Tuyuka consultant offered the nominalization based on the PCW predicate form involving the copula *ni*, as in (26) below³¹.

³¹ At one time, our Tuyuka consultant tried to distinguish the present/past readings in terms of the difference in the accent placement, as below:

yuhkigi añu-**ri**-gi ‘A árvore que é boa/The tree that is beautiful.’ (present).

yuhkigi añú-**ri**-gi ‘A árvore que era boa/The tree that was beautiful (past)

But, she opted for the predicative PCW form for the past reading at another time.

(26) Tuyuka nominalizations of PCW predicates

- a. wĩmãgõ [pai-gó ni-ri-gó]
 girl big-F.SG COP-PST.NZR-F.SG
 ‘a/the girl who was big’ Lit. ‘a/the girl that was a big feminine thing’

- b. yukigi [pai-ri-gi ni-ri-gi]
 tree big-PST.NZR-CLF_{TREE} COP-PST.NZR-CLF_{TREE}
 ‘a/the tree that was big’ Lit. ‘a/the tree that was a big tree-class thing’

Classifier marking patterns of PCW-based nominalizations are again like the other structures studied above: animate-classifier marking is obligatory in their NP and modification use; inanimate-classifier marking, on the other hand, is obligatory in both NP use and postnominal modification use, but is optional in prenominal modification.

Tukano speakers generally prefer inanimate classifiers to follow the patterns of animate classifiers, marking PCW roots directly with a classifier (without an intervention of the nominalizer *-ri*) except for some adjectival forms such as *pahi-ri(-wi)* [big-NZR(-CL)], which contains *-ri* and can modify (see the related discussion on V-based nominalizations in Section 7).

The following summary format provides a clearer presentation of the patterns of convergence (indicated by square brackets) and divergence between animate and inanimate marking.

Present tense

NP use

Animate	
*-Ø, *-ri, *-ri-M/F, -M/F	CLF OBL
Inanimate	
*-Ø, *-ri, -ri-CLF, *-CLF	CLF OBL

PrN modification

Animate	
*-Ø, *-ri, *-ri-F/M, -M/F	CLF OBL
Inanimate	
*-Ø, -ri, -ri-CLF, *-CLF	CLF OPT

PoN modification

Animate	
*-Ø, *-ri, *-ri-F/M, -M/F	CLF OBL
Inanimate	
-Ø, --ri, ri-CLF, *-CLF	CLF OBL

Past tense

NP use

Animate	
*-Ø, *-ri, -ri-F/M, *-M/F	CLF OBL
Inanimate	
*-Ø, *-ri, -ri-CLF, *-CLF	CLF OBL

PrN modification

Animate	
*-Ø, *-ri, -ri-F/M, *-M/F	CLF OBL
Inanimate	
*-Ø, -ri, -ri-CLF, *-CLF	CLF OPT

PoN modification

Animate	
*-Ø, *-ri, -ri-F/M, *-M/F	CLF OBL
Inanimate	
*-Ø, *-ri, -ri-CLF, *-CLF	CLF OBL

5.6. Tuyuka V-based nominalizations

Verbal-based nominalizations in Tuyuka basically exhibit the classifier-marking patterns of PCWs discussed above. In investigating V-based nominalizations, we first elicited baseline structures to identify formal similarities and differences between finite and nominalized verb forms, as illustrated in the examples in (27) below.

5.6.1. Tuyuka V-based nominalizations in NP use and modification use

(27) Animates

Present tense

Baseline structures

- a. yí'i ññã-ã si'-kõ nũmĩõ
 1SG see-V.P.O one-F.SG woman
 'I see a woman.'
- b. nũmĩõ pai-go ni-yo
 woman big-F.SG COP-V.P.3FS
 'The woman is big.'

Animate present NP use

- a. *[[yí'i ññã(-ã)] [nũmĩõ] ni-yo
 1SG see(-V.P.O) woman COP-V.P.FS
 'The one I see is a woman.'

- b. *[[yɪ'i ñã-ri] [nũmĩõ] ni-yo
 1SG see-NZR woman COP-V.P.FS
 'The one I see is a woman.'
- c. *[[yɪ'i ñã-ri-gõ] [nũmĩõ] ni-yo
 1SG see -NZR-3.SG woman COP-V.P.FS
 'The one that I see is a woman.'
- d. [[yɪ'i ñã-gõ] [nũmĩõ] ni-yo
 1SG see-F.SG woman COP-V.P.FS
 'The one I see is a woman.'

Animate present prenominal modification

- e. *[[yɪ'i ñã(-ã)] nũmĩõ] pai-go ni-yo
 1SG see(-V.P.O) woman big-F.SG COP-V.P.FS
 'The woman I see is big.'
- f. *[[yɪ'i ñã-ri] nũmĩõ] pai-go ni-yo
 1SG see-NZR woman big-F.SG COP-V.P.FS
 'The woman I see is big.'
- g. *[[yɪ'i ñã-ri-go] nũmĩõ] pai-go ni-yo
 1SG see-NZR-3F.SG woman big-F.SG COP-V.P.FS
 'The woman I see is big.'
- h. [[yɪ'i ñã-go] nũmĩõ] pai-go ni-yo
 1SG see-3F.SG woman big-F.SG COP-V.P.FS
 'The woman I see is big.'

Animate present postnominal modification

- i. *[nũmĩõ [yɪ'i ñã(-ã)]] pai-go ni-yo
 woman 1SG see(-V.P.O) big-F.SG COP-V.P.FS
 'The woman I see is big.'
- j. *[nũmĩõ [yɪ'i ñã-ri]] pai-go ni-yo
 woman 1SG see-NZR big-F.SG COP-V.P.FS
 'The woman I see is big.'
- k. *[nũmĩõ [yɪ'i ñã-ri-go]] pai-go ni-yo
 woman 1SG see-NZR-3F.SG big-F.SG COP-V.P.FS
 'The woman I see is big.'

1. [nũmĩõ [yĩ'i ãĩã-go]] pai-go ni-yo
 woman 1SG see-F.SG big-F.SG COP-V.P.3FS
 'The woman I see is big.'

Past tense

Baseline structures

- a. yĩ'i ãĩã-wĩ si'-kõ nũmĩõ
 1SG see-V.PA.O one-F.SG woman
 'I saw a woman.'
- b. nũmĩõ pai-go ni-awõ
 woman big-F.SG COP-V.PA.FS
 'The woman was big.'

Animate past NP use

- a. *[yĩ'i ãĩã(-wĩ)] nũmĩõ ni-awõ
 1SG see(-V.PA.O) woman COP-V.PA.FS
 'The one I saw was a woman.'
- b. *[yĩ'i ãĩã-ri] nũmĩõ ni-awõ
 1SG see-NZR woman COP-V.PA.FS
 'The one I saw was a woman.'
- c. [yĩ'i ãĩã-ri-gõ] nũmĩõ ni-awõ
 1SG see -NZR-F.SG woman COP-V.PA.FS
 'The one I saw was a woman.'
- d. *[yĩ'i ãĩã-gõ] [nũmĩõ] ni-awõ
 1SG see-F.SG woman COP-V.PA.FS
 'The one I saw was a woman.'

Animate past pronominal modification

- e. *[[yĩ'i ãĩã(-wĩ)] nũmĩõ] pai-go ni-awõ
 1SG see(-V.PA.O) woman big-F.SG COP-V.PA.FS
 'The woman I saw was big.'
- f. *[[yĩ'i ãĩã-ri] nũmĩõ] pai-go ni-awõ
 1SG see-NZR woman big-F.SG COP-V.PA.FS
 'The woman I saw was big.'
- g. [[yĩ'i ãĩã-ri-gõ] nũmĩõ] pai-go ni-awõ
 1SG see-NZR-F.SG woman big-F.SG COP-V.PA.FS
 'The woman I saw was big.'

- h. *[yí'i ññã-go] nũmĩõ pai-go ni-awõ
 1SG see-F.SG woman big-F.SG COP-V.PA.FS
 'The woman I saw was big.'

Animate past postnominal modification

- i. *[nũmĩõ [yí'i ññã(-wi)]] pai-go ni-awõ
 woman 1SG see(-V.PA.O) big-F.SG COP-V.PA.FS
 'The woman I saw was big.'
- j. *[nũmĩõ [yí'i ññã-ri]] pai-go ni-awõ
 woman 1SG see-NZR big-F.SG COP-V.PA.FS
 'The woman I saw was was big.'
- k. [nũmĩõ [yí'i ññã-ri-go]] pai-go ni-awõ
 woman 1SG see-NZR-F.SG big-F.SG COP-V.PA.FS
 'The woman I saw was big.'
- l. *[nũmĩõ [yí'i ññã-go]] pai-go ni-awõ
 woman 1SG see-F.SG big-F.SG COP-V.PA.FS
 'The woman I saw was big.'

(28) Inanimates

Present tense

Baseline structures

- a. yí'i ññã-ã si'-ka wi'i
 1SG see-V.P.O one-CLF house
 'I see a house.'
- b. wi'i pairi-wi ni-a
 house big-CLF COP-V.P.O
 'The house is big.'

Inanimate present NP use

- a. *[yí'i ññã(-ã)] pairi-wi ni-a
 1SG see(-V.P.O) big-CLF_{BUILDING} COP-V.P.O
 'The one I see is big.' Lit. 'The one I see is a big building-class thing.'
- b. *[yí'i ññã-ri]] pairi-wi ni-a
 1SG see-NZR big-CLF COP-V.P.O
 'The one I see is big.'

- c. [yí'í ññã-ri-wi] pairi-wi ni-a
 1SG see-NZR-CLF big-CLF COP-V.P.O
 'The one (building) I see is big.'
- d. *[yí'í ññã-wi] pairi-wi ni-a
 1SG see-CLF big-CLF COP-V.P.O
 'The one (building) I see is big.'

Inanimate present prenominal modification

- e. *[yí'í ññã(-ã)] wi'í pairi-wi ni-a
 1SG see(-V.P.O) house big-CLF_{BUILDING} COP-V.P.O
 'The house I see is big.' Lit. 'The house I see is a big building-class thing.'
- f. [[yí'í ññã-ri] wi'í] pairi-wi ni-a
 1SG see-NZR house big-CLF COP-V.P.O
 'The house I see is big.'
- g. [[yí'í ññã-ri-wi] wi'í] pairi-wi ni-a
 1SG see-NZR-CLF house big-CLF COP-V.P.O
 'The house I see is big.'
- h. *[yí'í ññã-wi] wi'í pairi-wi ni-a
 1SG see-CLF house big-CLF COP-V.P.O
 'The house I see is big.'

Inanimate present postnominal modification

- i. *[wi'í [yí'í ññã(-ã)]] pairi-wi ni-a
 house 1SG see(-V.P.O) big-CLF_{BUILDING} COP-V.P.O
 'The house I see is big.' Lit. 'The house I see is a big building-class thing.'
- j. *[wi'í [yí'í ññã-ri]] pairi-wi ni-a
 house 1SG see-NZR big-CLF COP-V.P.O
 'The house I see is big.'
- k. [wi'í [yí'í ññã-ri-wi]] pairi-wi ni-a
 house 1SG see-NZR-CLF big-CLF COP-V.P.O
 'The house I see is big.'
- l. *[wi'í [yí'í ññã-wi]] pairi-wi ni-a
 house 1SG see-CLF big-CLF COP-V.P.O
 'The house I see is big.'

Past tense

Baseline structures

- a. *yì'i ññ-wì si'-ka wi'i*
 1SG see-V.PA.O one-CLF house
 'I saw a house.'
- b. *Wi'i pairi-wi ni-awì*
 house big-CLF COP-V.PA.O
 'The house was big.'

Inanimate past NP use³²

- a. **[yì'i ññ(-wì)] pairi-wi ni-awì*
 1SG see(-V.PA.O) big-CLF_{BUILDING} COP-V.PA.O
 'The one (building) I saw was big.' Lit. 'The one I saw was a big building-class thing.'
- b. **[yì'i ññ-ri] pairi-wi ni-awì*
 1SG see-NZR big-CLF COP-V.PA.O
 'The one I saw was big.'
- c. *[yì'i ññ-ri-wi] pairi-wi ni-awì*
 1SG see-NZR-CLF big-CLF COP-V.PA.O
 'The one (building) I saw was big.'
- d. **[yì'i ññ-wì] pairi-wi sni-awì*
 1SG see-CLF big-CLF COP-V.PA.O
 'The one (building) I saw was big.'

Inanimate past prenominal modification

- e. **[[yì'i ññ(-wì)] wi'i] pairi-wi ni-awì*
 1SG see(-V.PA.O) house big-CLF_{BUILDING} COP-V.PA.O
 'The house I saw was big.' Lit. 'The house I saw was a big building-class thing.'
- f. *[[yì'i ññ-ri] wi'i] pairi-wi ni-awì*
 1SG see-NZR house big-CLF COP-V.PA.O
 'The house I saw was big.'
- g. *[[yì'i ññ-ri-wi] wi'i] pairi-wi ni-awì*
 1SG see-NZR-CLF house big-CLF COP-V.PA.O
 'The house I saw was big.'

³²All the nominalization structures here are identical in form with the present-tense forms, the past-tense reading being forced by the past copula form (*ni-awì*) of the main clause.

- h. *[wi'i ññã-wi] wi'i] pairi-wi ni-awi
 1SG see-CLF house big-CLF COP-V.PA.O
 'The house I saw was big.'

Inanimate past postnominal modification

- i. *[wi'i [yi'i ññã(-wi)]] pairi-wi ni-awi
 house 1SG see(-V.PA.O) big-CLF_{BUILDING} COP-V.PA.O
 'The house I saw was big.' Lit. 'The house I saw was a big building-class thing.'
- j. *[wi'i [yi'i ññã-ri]] pairi-wi ni-awi
 house 1SG see-NZR big-CLF COP-V.PA.O
 'The house I saw was big.'
- k. [wi'i [yi'i ññã-ri-wi]] pairi-wi ni-awi
 house 1SG see-NZR-CLF big-CLF COP-V.PA.O
 'The house I saw was big.'
- l. *[wi'i [yi'i ññã-wi]] pairi-wi ni-awi
 house 1SG see-CLF big-CLF COP-V.PA.O
 'The house I saw was big.'

Consistent with the characterization of Tukanoan languages as verb–adjective languages, the marking patterns of verb-based nominalizations align with those of adjective-based nominalizations discussed in the previous section. These also align with the marking patterns of the other structures studied previously: (i) animate classifier marking is obligatory in both NP and modification uses, and (ii) inanimate classifier marking is obligatory in NP use and postnominal modification, but optional in prenominal modification. We will further elaborate on Tuyuka verb-based nominalizations in comparison with Kotiria and Tukano marking patterns on V-based nominalizations in Section 7.

5.7. Summary of Tuyuka classifier marking

The following tables summarize the overall classifier-marking patterns in Tuyuka.

Table 6. Tuyuka animate nominalizations

	NUM		DEM		GEN		ADJ		V-based	
NP use	OBL		OBL		OBL		OBL		OBL	
Mod use	PrN	PoN	PrN	PoN	PrN	PoN	PrN	PoN	PrN	PoN
	OBL	OBL	OBL	OBL	OBL	OBL	OBL	OBL	OBL	OBL

Table 7. Tuyuka inanimate nominalizations

	NUM		DEM		GEN		ADJ		V-based	
NP use	OBL		OBL		OBL		OBL		OBL	
Mod use	PrN	PoN	PrN	PoN	PrN	PoN	PrN	PoN	PrN	PoN
	OPT	OBL	OPT	OBL	OPT	OBL	OPT	OBL	OPT	OBL

(OBL=marking obligatory, OPT=marking optional, PrN=prenominal modification, PoN=postnominal modification)

The findings presented above yield three significant observations regarding classifier marking in Tuyuka:

1. Greater Grammaticalization of Animate Classifiers

Classifier marking for animate referents is more grammaticalized than for inanimate ones. This is evidenced by its obligatory presence across both structural and functional dimensions, aligning with the Shibatani hierarchy. Such obligatoriness is characteristic of highly grammaticalized systems, where certain morphological markers become entrenched and consistently applied.

2. Optional Inanimate Classifier-Marking Suggests Secondary Development

In contrast, the optional nature of classifier marking for inanimate nominalizations in prenominal modification contexts suggests a less advanced stage of grammaticalization. This pattern implies two key points:

- The NP-use context, which lacks agreement-controlling head nouns, serves as the primary locus for classifier marking.
- Optional classifier marking in prenominal modification contexts likely represents a secondary development.

This asymmetry indicates that classifier marking for inanimate referents is less fully grammaticalized compared to the consistently obligatory pattern observed with animate classifiers across all contexts.

3. Animate classifiers—often referred to as “gender” by some—and inanimate classifiers do not constitute two entirely separate systems, as the “gender vs. classifier” dichotomy might imply. Instead, their marking patterns exhibit substantial overlap and convergence, as demonstrated more clearly below.

In relation to Point 3, while Tables (6) and (7) effectively present the overall marking patterns of classifying nominalizations, they are less effective in capturing the nuanced similarities and differences between animate and inanimate marking patterns. A more revealing summary is provided below, where possible, impossible, and **shared patterns or convergence** are clearly delineated.

Table 8. Marking on NUM, DEM, GEN

a. NP use	b. Prenominal modification	c. Postnominal modification
Animate *-Ø, -M/F	Animate *-Ø, -M/F	Animate *-Ø, -M/F
Inanimate *-Ø, -CLF	Inanimate -Ø, -CLF	Inanimate *-Ø, -CLF

Table 9. Marking on Adjective/V-based nominalizations

Present tense

a. NP use	b. PrN modification	c. PoN modification
Animate *-Ø, *-ri, *-ri-M/F, -M/F	Animate *-Ø, *-ri, *-ri-F/M, -M/F	Animate *-Ø, *-ri, *-ri-F/M, -M/F
Inanimate *-Ø, *-ri, -ri-CLF, *-CLF	Inanimate *-Ø, -ri, -ri-CLF, *-CLF	Inanimate *-Ø, *-ri, ri-CLF, *-CLF

Past tense

d. NP use	e. PrN modification	f. PoN modification
Animate *-Ø, *-ri, -ri-M/F, *-M/F	Animate *-Ø, *-ri, -ri-M/F, *-M/F	Animate *-Ø, *-ri, ri-M/F, *-M/F
Inanimate *-Ø, *-ri, -ri-CLF, *-CLF	Inanimate *-Ø, -ri, -ri-CLF, *-CLF	Inanimate *-Ø, *-ri, -ri-CLF, *-CLF

Summary Tables (8) and (9) demonstrate that animate and inanimate classifier marking-patterns overlap across all usage categories. Especially compelling is the near-perfect parallelism in the marking of adjective- and verb-based nominalizations in the past tense. Despite the potential variability introduced by the nominalizer *-ri*, animate and inanimate forms exhibit strikingly similar patterns.

As for NUM, DEM, and GEN, both animate and inanimate classifiers show identical patterns in NP use and postnominal modification (Tables 8a and 8c), differing minimally in prenominal modification (Table 8b). For A/V-based nominalizations in the past tense, the convergence continues in NP- and postnominal-modification contexts (Tables 9d and 9f), with only slight differences in prenominal use (Table 9e). More pronounced divergence emerges in the present tense, where animate classifiers lose the nominalizer *-ri*, leading to distinct marking patterns (Tables 9a–9c). The rise and fall of *-ri* will be discussed further in Section 7.

One limitation of the Tuyuka data, for our purposes, is that they do not reveal the role of structural hierarchy within the Shibatani framework. This is due to the uniformity of classifier marking across all structural contexts—from numerals to V-based nominalizations (see Tables 6 and 7). To underscore the significance of the structural hierarchy, we now turn to other Tukanano languages that display distributional patterns diverging from those found in Tuyuka. Even a brief examination of these languages highlights not only the importance of the structural hierarchy, but also provides further—and even more compelling—evidence of

parallelism and convergence between animate- and inanimate-classifier marking, extending beyond what Tuyuka alone illustrates.

6. CLASSIFYING NOMINALIZATIONS IN OTHER TUKANOAN LANGUAGES

This section provides a crosslinguistic overview of the marking patterns of classifying nominalizations, focusing on six Tukanoan languages that vary in their genetic relationship to Tuyuka: Sekoya from the Western branch; Retuarã and Tanimuka from the putative Central branch; and Kotiria (Wanano), Wa'ikhana (Puratapuyu), and Tukano from the Eastern branch—all of which are closely related to Tuyuka.

We begin our discussion with languages that provide evidence for the left edge of the Shibatani hierarchy, where structures most likely to first attract nominalization markers are arranged. That is, the hierarchy predicts that the grammatical use of nominalization markers—including gender and classifiers—initially emerge in numeral-based structures in NP use³³. From there, they gradually extend both to other structures (demonstratives, genitives, etc.) and to the use of nominalized forms in modification contexts.

6.1. Sekoya

Vallejos (2021) is among the most relevant studies on Amazonian classifiers for our research. The author observes that “Sekoya does not exhibit grammatical agreement... Neither shape nor gender markers are required in modifiers within the noun phrase, whether these are demonstratives, adjectives, or numerals” (p. 424). However, classifier marking in Sekoya does occur, primarily in NP-use contexts involving grammatical nominalizations based on numerals, demonstratives, and others. Observe:

(29) Sekoya (Vallejos 2021: 424, 437, 441)

NP use

a. Javin-pi toatsõ-**kõ**-waʔi-re dʒe-**ki**-pi teʔ-i-re
 Javin-NOM three-**F**-PL:ANI-ACC DEM.other-**M**-NOM one-M-ACC
 hānehẽ-tse wa-ãʔhiʔi
 like-that-RES kill-REM-3SG:M.DE
 ‘Javin killed three [peccaries], the other [hunter] only one.’

b. dʒe-dʒ-ki-pi kahetse-pāi ki-a-**pi**-ã
 teach-NZR-M-NOM four-COLL plastic-**CLF.SPH**-PL:INA
 paʔa-ki-na toatsõ-**pi**-ã ña-wi-
 have-DEP:SG:M-DS three-CLF.SPH-PL:INA steal-N3SG.PFV.DE
 ‘The teacher had four soccer balls; three [soccer balls] were stolen.’

³³ NB: we are restricting our discussion in this paper primarily to grammatical nominalizations. It is likely that nominalization, including classifying nominalizations by gender/classifiers, starts in the lexical domain, where nouns—lexical items—are derived. There are many Amazonian languages (e.g., Arawakan language Apeña) in which classifying nominalization is largely restricted to lexical nominalizations.

Modification use

- a. hã dʒari dʒo-wi
 DEM.PROX small canoe-CLF.CONT
 ‘that small canoe’
- b. kadʒa dʒari dʒo-wi-ã
 two small canoe-CLF.CONT-PL:INA
 ‘two small canoes’

What is particularly significant for our study is Vallejos’s further observation that classifier marking is beginning to emerge in modification contexts, following the pattern predicted by the Shibatani hierarchy—specifically, that elements at the high end of the hierarchy, such as numerals and demonstratives on the left edge, are the most susceptible to classifier marking (see Diagram 1). This is what Vallejos has to say: “the numeral ‘one’ can sometimes be suffixed with gender morphology [i.e., animate classifiers] in adnominal constructions [i.e., in the modification contexts]... This observation applies, **although more rarely**, to some demonstratives” (p. 445; emphasis added). Below are examples illustrating these patterns of marking³⁴.

(30) Sekoya (Vallejos 2021: 447, 448)

- a. [teʔ-e ãmi-pi] idʒa-ire tsiwa-ki
 one-GEN man-NOM 3SG-dog-ACC happy-DEP:SG:M
 paʔa-hi, teʔ ãmi-pi
 have-SG:M:IPFV:DE one man-NOM
 ‘A man has a dog and is happy, and the man ...’
- b. [idʒe wiʔe] deo wiʔe-a-ʔʔ
 DEM.PROX-CLF.GEN house beautiful house-COP-3SG:INA:DE
 ‘This house is a beautiful house.’

The occurrence of classifier marking in Sekoya modification contexts by animate classifiers and the generic classifier resonates their higher frequency of occurrence, as noted by Vallejos; “in the corpus, one of the most common environments for the gender markers [i.e., animate classifiers] *-ko* [feminine]/*-ki* [masculine] is suffixed to numerals and demonstratives [in NP-use contexts]” (p. 437)

To summarize, the classifier marking patterns in Sekoya highlight NP-use contexts as the primary environment for obligatory marking, consistent with the Shibatani hierarchy. The emergence of optional marking in modification contexts—with the numeral ‘one’ and the proximate demonstrative—further supports this hierarchy. Marking is more frequent with the highest-ranked numeral (‘one’) and less so with the second-ranked (proximal) demonstrative. Moreover, the use of animate classifiers in modification contexts suggests that they

³⁴ Vallejos notes that both animate classifiers and the generic classifier *-(dʒ)e* mark the modifying numeral “one” and proximate demonstrative, but offers only examples marked by the generic classifier.

grammaticalize more rapidly—appearing in a broader range of contexts—than inanimate classifiers.

6.2. Retuarã and Tanimuka

According to Strom (1992: 1), “the Retuarã and Tanimuka, located in southeast Colombia, consider themselves separate groups yet speak the same language, which belongs to the central branch of Tucanoan.” However, he acknowledges “some grammatical variations” arising from the extensive dispersion of villages and their interactions with neighboring languages. Strom’s Retuarã data, based on speakers from the lower Aparoris River, and Eraso’s (2015) Tanimuka data from the Mirití River will be discussed below.

As for Retuarã, Strom observes that “[c]lassifiers occur **obligatorily bound to numerals** and **optionally to adjectives and demonstratives**” (p. 54; emphasis added), giving examples like the ones below:

(31) Retuarã modification forms (Strom 1992: 55, 56)

- a. ɪ̃rã-**ika** pōʔĩbã-**h̃** (animate CLF)
 one-M.SG man-M.SG
 ‘one man’
- b. bāēkaraka-**to** puyũā (general CLF)
 three-CLF_{2DIMEN} leaf
 ‘three leaves’
- c. ɪ̃rã-**pāũ** pāũā (specific CLF)
 one-CLF_{HAMMOCK} hammock
 ‘one hammock’

Concerning the NP use of classifying nominalizations, Strom makes the incisive observation as follows; “Classifiers can also occur without a head. Thus, they can have referential capacity like the head of a noun phrase.” (p. 55) Observe:

(32) Retuarã (Strom 1992: 56)

- a. ĩsi-rībī yiha-pe’-rape
 that-CLF_{DAY} 1PL-reurn-PST
 ‘That day we returned.’
- b. hī’bī-ri-bi dā-e’e-rape
 green-INDEF-CLF_{1DIMEN} 3PL-get-PST
 ‘They got (a) green one(s) (sticks).’

Strom’s observation that “classifiers occur ... optionally with adjectives and demonstratives” should be taken with caution, in light of Eraso’s (2015: 329–330) account of Tanimuka, where the distribution pattern of classifiers is summarized as follows: “Los

clasificadores del tanimuka, como se dijo, son del tipo numeral. Se **sufijan de manera obligatoria** en palabras que denotan cantidad (**numerales**, cuantificadores, interrogativos decantidad), **en demostrativos con función deíctica** y son opcionales en adjetivos. No aparecen sufijados ni a los nombres ni al verbo” (As mentioned, the classifiers in Tanimuka are of the numeral type. **They are obligatorily suffixed** to words that denote quantity (**numerals**, quantifiers, interrogative words related to quantity), **to demonstratives** with a deictic function, and are **optional in adjectives**. They do not appear suffixed to nouns or verbs—our translation; emphasis added.

The obligatory classifier marking on demonstratives in Tanimuka, as discussed by Eraso, appears to reflect a nominal classification based on the human/non-human distinction. It differentiates among “human masculine/M”, “human feminine/F”, and “neutral or non-human/N” categories (Eraso 2015: 325), as illustrated in the following examples:

(33) Tanimuka (Eraso 2015: 325)

- a. i'-kí ~ibirih-í hó'ba-kí a'-jú-kí (Human masculine/M)
 DEM-M man-M big-M go-PRS-M
 ‘This big man goes.’
- b. i'-kó ~rob-ó boi-kó a'-jú-kó (Human feminine/F)
 DEM-F woman-F white-F go-PRS-F
 ‘This white woman goes’
- c. i'-ká ~jabá-ká hó'ba-ká a'-jú-á (Non-human/N)
 DEM-N deer-N big-N go-PRS-N
 ‘This big deer goes.’
- d. i'-ká hotó-a hó'ba-ká tiri-ko'o-ka (Non-human/N)
 DEM-N pot- N big-N break-PST.DISTANT-N
 ‘This big pot broke.’

The classification based on the human/non-human distinction, as illustrated above, is referred to as “noun classes” by Strom and as “gender” by Eraso. Eraso adopts this label based on the assumption that the marking is driven by agreement between the head noun and its dependent modifiers, such as numerals and demonstratives. Eraso (2015) describes this situation as follows: “El sistema de género en tanimuka es un sistema de concordancia gramatical, característica que marca la diferencia principal entre éste y el sistema de clasificadores (The gender system in Tanimuka is a grammatical agreement system, a characteristic that marks the main difference between it and the classifier system³⁵—our translation) (p. 322)³⁶. However, the agreement involving so-called “gender” classes does not

³⁵ Agreement between inanimate classifiers and head nouns is observed in numeral modifiers—see examples (35c, d).

³⁶ See further: “El sistema de género, en tanimuka, presenta concordancia en el ámbito del sintagma nominal, donde se marcan cabeza y dependientes del SN, y en el ámbito de la construcción sujeto-verbo” (The gender

appear to be entirely obligatory. Consider the following examples involving what Eraso calls “gender”—the “neuter/non-human (N) gender”. While agreement is seen in (34a), it is clearly optional with an adjectival modifier as in (34b)–(34d), and can be entirely absent even in an NP-use context in (34e):

(34) Tanimuka (Eraso 2015: 203, 343, 357)

- a. hó'ba-ó-**ká** biá-Ø
 big-CLF_{ROUND}-**N** chili-**N**
 ‘big round chili’
- b. hó'ba-ó biá-Ø
 big-CLF_{ROUND} chili-**N**
 ‘big round chili’
- c. hó'ba-~hukí ~kára-ka
 grande-CLF_{PLANT} caimo-**N**
 ‘big caimo tree’
- d. hó'bà-~kù'bú ~kùbú-á
 big-CLF_{CANOE} canoe-**N**
 ‘big canoe’
- e. jí-ré bi-~ihí-be hó'ba-ó
 1-DEIC 2S-give-IMP big-CLF_{ROUND}
 ‘Give me the big round one (referring to the fruit).’

Consider the above in light of Eraso’s observation on classifiers: “**the classifiers** in Tanimuka are of the numeral type ... and **are optional in adjectives**” (2015: 329–330). In the examples in (34), the “neutral/non-human gender (N)” behaves much like inanimate classifiers. What appears to be obligatory are the animate classifiers—namely, “masculine singular,” “feminine singular,” and “animate plural”—a pattern closely paralleling our findings on Tuyuka. Unfortunately, neither Strom nor Eraso provides a detailed account of the relationship—despite considerable morphological overlap—between what they refer to as “noun classes” and “gender”, on the one hand, and the animate classifiers within the (numeral) classifier system, on the other hand, which are separately recognized for Retuarã (Strom 1992: 54) and Tanimuka (Eraso 2015: 330ff). A more integrated approach, like the one advocated in this paper, would collapse the human “noun class/gender” together with the animate classifiers as one set consisting of **M.SG** (-ika/-ki), **F.SG** (-ko), and **ANI.PL** (-rã), corresponding to the animate classifiers in Tuyuka and others. And then recognize a non-human/inanimate, **N** class (-ka), which seems to share characteristics of both animate and inanimate classifiers³⁷.

system in Tanimuka exhibits agreement in the realm of the noun phrase, where the head and dependents of the noun phrase are marked, and in the area of the subject-verb construction—our translation.) (p. 325)

³⁷ Recall that we must set up a human class in describing the marking pattern of genitives in Tuyuka (see (21)).

Whatever formal differences, as observed between Strom’s and Eraso’s “gender” **M** class (*-ki*) and the animate **M.SG** (*-ika*), can be attributed to the differences in syntactic environments; e.g., *-ika* with numerals and *-ki* with demonstratives, adjectives, and verb agreement. Also, see the variation between *-a* and *-ka* for the N-class morpheme in verb agreement (33)³⁸.

With the reorganization of the animate classifiers for Retuarã and Tanimuka for our purposes, let us examine Strom’s and Eraso’s observations on the classifier marking in these languages. Assuming that NP-use forms all require animate or inanimate classifiers as in Sekoya, Tuyuka, and some other Tukanoan languages, our focus below will be on the modification use of classifying nominalizations in Retuarã and Tanimuka.

When used with numerals, both animate and inanimate classifiers appear to be obligatory as illustrated in the examples below:

(35) Retuarã and Tanimuka

- a. *ĩ’ã-ika* *yaiwẽkoa*
 one-M.SG dog
 ‘one dog’ (Strom 1992: 50)

- b. *ĩpa-korã* *rõbĩha*
 two-F.ANI.PL women
 ‘two women’ (Strom 1992: 52)

- c. *~ípha~kù’bú* *~kùbú-á*
 two-CLF_{CANOE} canoe-N
 ‘two canoes’ (Eraso 2015: 357)

- d. *i’~rá-tí* *~jaké-á*
 one-CLF_{STRING} bead-N
 ‘one bead necklace’ (Eraso 2015: 380)

The data above indicate that numeral-based nominalizations in modification use may not show agreement for Eraso’s gender class N (non-human)³⁹, whereas both animate classifiers and **inanimate classifiers show agreement**—the head noun “canoe” calls for the canoe classifier and the head noun “bead necklace” invokes the string classifier⁴⁰.

As for demonstratives and adjectives, Strom and Eraso disagree in their observations: the former says classifier marking is optional in both, whereas Eraso claims that the marking is obligatory with demonstratives but optional with adjectives. Let us start with adjectives on which the two researchers agree.

³⁸ Such formal variations of form are widely seen in Amazonian languages (see Aikhenvald 2007).

³⁹ We failed to find examples where numerals are marked by the N-class *-(k)a* morpheme in the literatures cited here.

⁴⁰ NB: (35c), where morphological agreement involving a repeater classifier is also seen, which is a common place observation in Tukanoan languages. Recall that Steele (1978: 610) defines “agreement” as “some syntactic covariance between a semantic or formal property of one element and a formal property of another”.

(36) Retuară (Strom 1992: 51, 53, 56)

- Eraso's richer examples of adjectival nominalizations in Tanimuka include inanimate classifier marking in modification contexts, as in (37) below:

a. hó'ba-kó ~be'r-ó
big-F child-F
'big girl'

- 69

- f. hó'ba-ó-ká ~kára-ka
 big-CLF_{ROUND}-N caimo-N
 'big (round) caimo (fruit)'
- g. hó'ba-~hukí ~kára-ka
 big-CLF_{PLANT} caimo-N
 'big (plant of) caimo'
- h. hó'bà-~kù'bú ~kùbú-á
 big-CLF_{CANOE} canoe-N
 'big canoe'

The interesting patterns in the Tanimuka data above are seen in (37e)–(37h), where the non-human classifier (N) may mark a modifier alone (37e) without further marking, like animate classifiers (37a)–(37c), and like inanimate classifiers (37g) and (37h). The non-human classifier may also cooccur in combination with an inanimate classifier (37e, f), showing its unique characteristic separating it from animate classifiers, which apparently cannot combine with an inanimate classifier.

The non-human classifier (N) alternates with inanimate classifiers, as shown in the comparison between (37d) and (37f-g). This suggests that the non-human classifier functions as a general inanimate classifier. For instance, the sentence “I want *hó'ba-ka*” would mean “I want a big non-human thing,” contrasting with “I want *hó'ba-~kù'bú*”, which specifically denotes a canoe-class object. It is likely this alternating pattern that leads both Strom and Eraso to claim that (inanimate-) classifier marking is optional when used with adjectives.

If the above interpretation of Strom's and Eraso's understanding of inanimate classifier marking on adjectives in modification contexts—that it is optional—is correct, then we must conclude that (inanimate-) classifier marking on demonstratives in such contexts is optional as well, contrary to Eraso's claim that it is obligatory (Eraso 2015: 329–330). Observe the following data from the works by Strom and Eraso.

(38) Retuarã and Tanimuka

Animate and non-human (Eraso 2015: 325, 482)

- a. i'-kí ~ibirih-í
 DEM-M man-M
 'this man'
- b. i'-kó ~rob-ó
 DEM-F woman-F
 'this woman'
- c. í'-rã rōmi-há
 DEM-PL woman-COLL
 'these women'

- d. i'-ká hotó-a
 DEM-N pot- N
 'this pot'
- e. i'-ká radio
 DEM-N radio
 'this radio' (Strom 1992: 52)

Inanimate (Eraso 2015: 250)

- f. i-ó
 this-CLF_{ROUND}
 'this (round) one (e.g., caimo fruit “star apple”)'
- g. i-bí
 this-CLF_{LONG}
 'this (long) one (e.g., chili pepper)'
- h. í~hùkí
 this-CLF_{PLANT}
 'this (plant kind) one (e.g., caimo tree)'
- i. i'si-to yi-re bĩ-re bĩ-ĩhĩ-be
 that-CLF_{1DIMEN} 1SG-TERM 2SG-give-IMP
 'Give me that (flat thing, i.e., a piece of clothes).' (Strom 1992: 52)

Unfortunately, neither Strom nor Eraso provides examples in which demonstratives marked by inanimate classifiers used as a modifier. The following form (39b), therefore, is hypothesized by us in order to show a likely alternation between a demonstrative marked by a non-human classifier (N) and the one marked by an inanimate classifier.

(39) Tanimuka

- a. i'-ká hotó-a
 this-N pot-N
 'this pot' (Eraso 2015: 325)
- b. í~ hòtó kárí.hótó-a (hypothetical⁴¹)
 this-CLF_{POT} tucupi.pot-N
 'this tucupi pot'

⁴¹ This hypothetical form is based on Eraso's example with a numeral:
 ~í'ra-hòtó kárí.hótó-a (Eraso 2015: 352)
 one-CLF_{POT} tucupi.pot-N
 'one tucupi pot'

This highly likely alternation pattern is just like the similar alternation between non-human, N-marked adjective and the one marked by an inanimate classifier (see (37)), where we concluded that inanimate classifier marking with a modifying adjective is optional, as an inanimate head now does not necessarily invoke (inanimate-)classifier marked adjective. This conclusion carries over to demonstratives, to the alternation pattern seen in (39) immediately above.

In fact, there are examples that more straightforwardly show that a modifying demonstrative can be totally free of classifier marking. This concerns the distal demonstrative *i'si* 'that', which may be marked by a classifier, as in (38i) above, and in (40) below:

(40) Tanimuka (Eraso 2015: 220, 246, 250)

- a. *i'sí-á b-e'ê.ra'á-be*
that-N 2SG-bring-IMP
'Bring that (non-human thing)!'
- b. *í'si-a jaφu-a*
that-N stick-N
'that stick'
- c. *i'si-ó*
that -CLF_{ROUND}
'that (round) one'

The distal demonstrative *i'si* can occur without a classifier in modification contexts, as clearly shown by the examples in (41) from both Retuarã and Tanimuka⁴².

(41) Retuarã and Tanimuka

- a. *i'si pisadaka*⁴³
that cat
'that cat' (Strom 1992: 53)
- b. *i'sí φi'í-á ko-riká*⁴⁴
DEM.DISTAL basket-N 3F-GEN
'That basket is hers.' (Eraso 2015: 233)

⁴² It is most likely that the distal demonstrative *i'si* also calls for animate classifiers in both NP- and modification-use contexts like the proximate demonstrative in (38). That the distal demonstrative in these languages can occur without any classifier in modification contexts indicates a slower pace of grammaticalization of a distal demonstrative, similar to the distal demonstrative in Sekoya, which does not take a classifier in modification contexts, unlike the proximate demonstrative, which occasionally is marked by a classifier even in modification contexts.

⁴³ The non-human classifier *-ka* should be able to optionally mark the demonstrative here.

⁴⁴ The demonstrative here should be able to be optionally marked by either the non-human classifier *-ka* or by the basket repeater classifier *-φi'í*.

Now, returning to our focus on the development and distribution of classifier marking, Retuarã and Tanimuka represent a more advanced stage than that of Sekoya. In Sekoya, **optional classifier marking occurs in modification contexts (i) with the numeral ‘one,’ and (ii) only marginally with the proximate demonstrative.** By contrast, Retuarã and Tanimuka, in the same contexts, show (i) **obligatory classifier marking with numerals and (ii) optional marking with demonstratives and adjectives**—consistent with the developmental trajectory and crosslinguistic distribution predicted by the Shibatani hierarchy, where adjectives are claimed to lag behind numerals and demonstratives in acquiring classifier marking. Additionally, the inventory of classifiers used with modifiers expands significantly from Sekoya to Retuarã/Tanimuka. In Sekoya, such marking is restricted to animate and general classifiers, whereas in Retuarã and Tanimuka, specific inanimate classifiers are also recruited. Assuming that marking in NP-use contexts is obligatory, as in Tuyuka and the other Tukanooan languages we studied, the situations in Sekoya and Retuarã/Tanimuka can be summarized as in Tables 10 and 11.

Table 10. Sekoya marking pattern

	NUM	DEM	GEN	ADJ
NP use	OBL	OBL	OBL	OBL
Mod use	OPT	^MOPT	NO	NO

(^MOPT = marginal)

Table 11. Retuarã/Tanimuka

	NUM	DEM	GEN	ADJ
NP use	OBL	OBL	OBL/NO	OBL
Mod use	OBL	OPT	?/NO	OPT

The Retuarã/Tanimuka genitives are interesting in showing several things. First, the genitive marking pattern violates the Shibatani hierarchy, showing that GEN higher in the hierarchy appears to be marked to a lesser extent than the lower ranked ADJ. This calls for some explanation. Secondly, the genitive morpheme *-rika* shows a highly transparent connection with the possessive verb *rika*. This morphological transparency offers a potential clue to the exceptional behavior of genitives in Retuarã/Tanimuka. The genitive marker *-rika* in question appears in NP-use contexts, as illustrated in examples (42) below:

(42) Retuarã and Tanimuka

- a. hũã-**rika** i'ka yaiwẽkoa
 John-GEN this dog
 ‘This dog is **John’s**.’ (Strom 1992: 66)
- b. i'sí φi'i-á ko-**riká**
 DEM basket-N 3F.SG-GEN
 ‘That basket is **hers**.’ (Eraso 2015: 233)

When the genitive forms with *-rika* denote animate entities, they require animate-classifier marking, as in (43) below:

(43) Retuarã

marco-**ri-ko** ko-ĩbẽ
marco-**GEN-F.S** 3SG-be
‘She is **Marco’s**’. (Strom 1992: 67)

It is not made clear whether such genitives marked by an animate classifier can be used as a modifier. Genitive modifiers exemplified by both Strom and Eraso have no classifier, as seen in (44):

(44) Retuarã and Tanimuka

a. hũã-re **yi**-sara ki-**rika**-yu
Juan-TERM 1SG-machete 3M.SG-**have**-PRS
‘Juan has my machete’. (Strom 1992: 131)

b. ~**bi**-~tíbí
2SG-husband
‘your husband’ (Eraso 2015: 221)

Strom’s example (44a) shows that the genitive marker *-rika* originated from the possessive verb *rika* ‘to have’ (Strom 1992: 131-132). This grammaticalization likely occurred relatively recently, possibly as a replacement for the older genitive marker, an equivalent of *yaa* found in Eastern Tukanoan languages (see (12)). Because of the recency of this innovation, the genitive constructions in Retuarã/Tanimuka diverge from the typical genitive marking patterns observed in Eastern Tukanoan and many other languages. A comparable case of exceptional genitive behavior is discussed in Shibatani (2013) in relation to Thai.

Having examined the development of classifier marking at the higher end of the Shibatani hierarchy, we now turn to crosslinguistic patterns at the lower end—the right edge—of the hierarchy. This domain hosting V-based nominalizations is expected to exhibit less developed marking patterns (e.g., unmarked or optionally marked forms), at least in some languages, compared to structures found higher in the hierarchy. We begin by examining V-based nominalizations (aka participles, relative clauses) in Kotiria, then turn to Tukano, and compare both with Tuyuka data to follow up on the preceding discussion of Tuyuka adjective/verb-based nominalizations. These relatively closely related Eastern Tukanoan languages reveal intriguing patterns in the development of classifier marking at the lower end of the Shibatani hierarchy.

6.3. Kotiria

Also known as Wanano, Kotiria presents a remarkable case of forming V-based nominalizations from finite verbs in both the present and past tenses⁴⁵. Such forms may be marked by a nominalizer and/or a classifier, but needn't be in NP- or modification-use contexts, reflecting a less developed pattern of classifier marking compared with numerals and demonstratives of the language, which typically require classifier marking, at least in NP-use contexts. First observe these facts below:

(45) Kotiria

Imperfective/Present tense⁴⁶

Baseline structures

- a. yï nũmĩrõ nhi-hã
 1SG woman see-V.IMP.1
 'I see a woman.'
- b. a'ri-go nũmĩrõ phiri-koro hi-ra
 that-F.SG woman big-F.SG COP-V.IMP.N1
 'That woman is big.'

V-based nominalization

NP use

- a. [yï'i nhi-ha] nũmĩrõ hi-ra
 1SG see-V.IMP.1 woman COP-V.IMP.N1
 'The one I see is a woman.'

Modification use

- b. [[nũmĩrõ [yï'i nhi-ha]] phiri-koro hi-ra
 woman 1SG see-V.IMP.1 big-F.SG COP-V.IMP.N1
 'The woman I see is big.'

Past/Perfective baseline structures

- a. yï nũmĩrõ nhi-i
 1SG woman see-V.PERF.1
 'I saw a woman.'
- b. a'ri-go nũmĩrõ phiri-koro hi-re
 that-F.SG woman big-F.SG COP-V.PERF.N1
 'That woman was big.'

⁴⁵ See Stenzel (2013) for the rationale behind shifting the language name from "Wanano" to "Kotiria".

⁴⁶ In the glosses below, we follow Stenzel (2004) and use IMPERF and PERF for the present and past readings.

NP use

- c. [yí'i **nhi-i**] nũmĩrõ hi-re
1SG see-V.PERF.1 woman COP-V.PERF.N1
'The one I saw was a woman.'

Modification use

- d. [[nũmĩrõ [yí'i **nhi-i**]] phiri-koro hi-re
woman 1SG see-V.PERF.1 big-F.SG COP-V.PERF.N1
'The woman I saw was big.'

With this background information on V-based nominalizations in Kotiria in mind, let us examine the patterns of occurrence of the nominalizer (*-ri*) and classifiers. All the examples below are in the past tense/perfective aspect.

(46) Kotiria **Animate** V-based nominalizations (see the past/perfective baseline structures above)

NP use

- a. [yí'i **nhi-i**] nũmĩrõ hi-re
1SG s see-V.PERF.1 woman COP-V.PERF.N1
'The one I saw was a woman.'
- b. *[yí'i **nhi-i-ri**] nũmĩrõ hi-re
1SG see-V.PERF.1-**NZR** woman COP-V.PERF.N1
'The one I saw was a woman.'
- c. [yí'i **nhii-ri-koro**] nũmĩrõ hi-re
1SG see-V.PERF.1-**NZR-F.SG** woman COP-V.PERF.N1
'The (feminine) one I saw was a woman.'
- d. yí'i **nhii-koro**] nũmĩrõ hi-re
1SG see-V.PERF.1-**F.SG** woman COP-V.PERF.N1
'The (femine) one I saw was a woman.'

Modification use

- e. [[yí'i **nhi-i**]]⁴⁷ nũmĩrõ] phiri-koro hi-re
1SG see-V.PERF.1 woman big-F.SG COP-VERF.P.N1
'The woman I saw was big.'
- f. [[yí'i **nhi-i-ri**] nũmĩrõ] phiri-koro hi-re
1SG see-V.PERF.1-**NZR** woman big-F.SG COP-V.PERF.N1
'The woman I saw was big.'

⁴⁷ All the modification constructions here can also use the verb root form *nhi* 'see' instead of *nhi-i* marked for the past/perfective aspect.

f'. *[nũmĩrõ [yi'i nhii-**ri**]] phiri-koro hi-re⁴⁸
 woman 1SG saw-NZR big-F.SG COP-V.PERF.N1
 'The woman I saw was big.'

g. [[yi'i nhi-i-**ri-koro**] nũmĩrõ] phiri-koro hi-re
 1SG see-V.PERF.1-NZR-F.SG woman big-F.SG COP-V.PERF.N1
 'The woman I saw was big.'

g'. *[nũmĩrõ [yi'i nhii-**ri-koro**]] phiri-koro hi-re
 woman 1SG saw-NZR-F.SG big-F.SG COP-V.PERF.N1
 'The woman I saw was big.'

h. [[yi'i nhi-i-**koro**] nũmĩrõ] phiri-koro hi-re
 1SG see-V.PERF.1-F.SG woman big-F.SG COP-V.PERF.N1
 'The woman I saw was big.'

(47) Kotiria **inanimate** V-based nominalizations

Baseline structures

a. yi'i wi'i nhi-i
 1SG house see-V.PEF.1
 'I saw a house.'

b. a'ri wi'i phiri-wi hi-re
 that house big-CLF_{BUILDING} COP-V.PERF.N1
 'That house was big.'

NP use

a. *[yi'i nhi-i] wi'i hi-re
 1SG see-V.PERF.1 house COP-V.PERF.N1
 'The one I saw was a house.'

b. [yi'i nhi-i-**ri**] wi'i hi-re
 1SG see-V.PERF.1-NZR house COP-V.PERF.N1
 'The one I saw was a house.'

c. [yi'i nhi-i-**ri-wi**] wi'i hi-re
 1SG see-V.PERF.1-NZR-CLF house COP-V.PERF.N1
 'The one I saw was a house.'

⁴⁸ (46f') and (46g') are ungrammatical as postnominal modification; all other forms are possible as postnominal modifiers.

- d. [yi'i nhi-i-**wi**] wi'i hi-re
 1SG see-V.PERF.1-CLF house COP-V.PERF.N1
 'The one I saw was a house.'

Modification use⁴⁹

- e. [[yi'i **nhi-i**] wi'i] phiri-wi hi-re
 1SG see-V.PERF.1 house big-CLF COP-V.PERF.N1
 'The house I saw was a house.'
- f. [[yi'i nhi-i-**ri**] wi'i] phiri-wi hi-re
 1SG see-V.PERF.1-**NZR** house big-CLF COP-V.PERF.N1
 'The house I saw was big.'
- g. [[yi'i nhi-i-**ri-wi**] wi'i] phiri-wi hi-re
 1SG see-V.PERF.1-**NZR-CLF** house big-CLF COP-V.PERF.N1
 'The house I saw was big.'
- h. [[yi'i nhi-i-**wi**] wi'i] phiri-wi hi-re
 1SG see-V.PERF.1-**CLF** house big-CLF COP-V.PERF.N1
 'The house I saw was big.'

The Kotiria marking patterns on V-based nominalizations are summarized in Tables 12 and 13 below:

Table 12. NP-use of V-based nominalizations in Kotiria

Animate	-Ø, *-ri, -ri-M/F, -M/F
Inanimate	*-Ø, -ri, -ri-CLF, -CLF

Table 13. Prenominal modification use of V-based nominalizations in Kotiria

Animate	-Ø, -ri, -ri-M/F, -M/F
Inanimate	-Ø, -ri, -ri-CLF, -CLF

⁴⁹ All the prenominal modifiers below can also be postnominal modifiers.

Table 14. Postnominal modification use of V-based nominalizations in Kotiria

Animate	-Ø, *-ri, *-ri-M/F	-M/F
Inanimate	-Ø, -ri, -ri-CLF,	-CLF

These marking patterns largely confirm Stenzel’s summaries for Kotiria.

Table 15. Kotiria making patterns according to Stenzel (2004)

Animates (Stenzel 2004: 150)

Vroot+	nominalizing suffix -ri +	noun classification morphology: -(ko)ro , (feminine/masculine singular) ~da (plural)

Inanimates (Stenzel 2004: 153)

Vroot+	(nominalizing suffix -ri +)	noun classification morphology: -ro_p/classifier (singular) or ~ri (plural)

Our findings reveal a broader range of patterns than those identified in Stenzel’s summaries, particularly in relation to the use of the nominalizer *-ri*. Contrary to the implications of Stenzel’s summary tables, our data indicate that animate-classifier marking more strongly facilitates the loss of *-ri* than inanimate-classifier marking does (see Tables 12 and 14, and the discussion of other Tukanoan languages in Section 7). Nonetheless, both sets of findings reveal an even greater degree of convergence between inanimate and animate classifier marking than is evident in Tuyuka, thereby offering stronger support for a unified analysis of the two.

Before concluding this section, it is worth noting that Wa’ikhana (aka Piratapuyo), a close relative of Kotiria, displays patterns of classifier marking on V-based nominalizations that mirror some of those discussed above, once again demonstrating a high degree of parallelism between animate and inanimate classifier marking. Observe the marking patterns in (48) below, where the distributions of animate- and inanimate-classifier marking show a perfect parallelism.

(48) Wa’ikhana

Baseline structure

ũmũnõ/nũmĩnõ papera rũã-ri (ri-dĩ)⁵⁰
man/woman book burn-PST/NZR COP-IMPERF
‘The man/woman burned a book.’

⁵⁰ The fact that this sentence can be concluded with a copula indicates that the V-based nominalization marked by past/NMZR *-ri* is being insubordinated and becoming a main sentence by itself.

NP use of V-based nominalizations

- a. [papera rũa-ri-**kiro**/rũa-ri-**koró**] yĩ'i pa'ki/pa'kó ri-dí
 book burn-NZR-**M.SG**/burn-NZR-**F.SG** 1SG father/mother COP-IMPERF
 'The one that burned the book is my father/mother.'
- b. [yĩ'i pa'ki tikiró riã-ri-**pun**] insēbi'ti
 1SG father he burn-NZR-CLF_{LEAF} thick
 'The one that my father, he, burned was thick.'

Modification use of V-based nominalizations

- a'. [ũmũnõ/nũmĩnõ [papera rũa-ri-**kiro**/rũa-ri-**koró**]
 man/woman book burn-NZR-**M.SG**/burn-NZR-**F.SG**
 yĩ'i pa'ki/pa'kó ri-dí
 1SG father/mother COP-IMPERF
 'The man/woman that burned the book is my father/mother.'
- b'. papera [yĩ'i pa'ki tikiró riã-ri-**pun**] insēbi'ti
 book 1SG father he burn-NZR-**CLF**_{LEAF} thick
 'The book that my father, he, burned was thick.'

6.4 Tukano

What appears to be the most critical point of comparison of V-based nominalizations across Tukanoan languages is the rise and decline of the marking pattern of: **V-ri-ANI.CLF** and **V-ri-INA.CLF**, both of which are preserved in Kotiria and Wa'ikhana, along with other marking patterns. What we saw in Tuyuka was the loss of *-ri* before animate classifiers in the present/imperfective context. In Tukano the nominalizer is lost in both present/imperfective and past/perfective contexts, with some exceptions. Consider the following marking patterns in (49):

(49) Tukano

Animate

Present baseline structures

- a. yĩ'i ññã-a ni-kō nũmĩõ
 1SG see-V.P.N3M/F one-F.SG woman
 'I see a woman.'
- b. nũmĩõ pa'i-go ni-mõ
 woman big-F.SG COP-V.P.F.SG
 'The woman is big.'

NP use

- a. *[[yɪ'i ññã-ã] pa'i-go nũmĩõ ni-mõ
1SG see-V.P.N3M/F big-F.SG woman COP-V.P.F.SG
'What I see is a big woman.'
- b. *[[yɪ'i ññã(-ã)-ri] p a'i-go nũmĩõ ni-mõ
1SG see-NZR big-F.SG woman COP-V.P.F.SG
'What I see is a big woman.'
- c. *[[yɪ'i ññã(-ã)-ri-go] pa'i-go nũmĩõ ni-mõ
1SG see-NZR-F.SG big-F.SG woman COP-V.P.F.SG
'What I see is a big woman.'
- d. [[yɪ'i ññã(-ã)-gõ] pa'i-go nũmĩõ ni-mõ
1SG see-F.SG big-F.SG woman COP-V.P.F.SG
'What I see is a big woman.'

Modification use

- e. *[[yɪ'i ññã-ã] nũmĩõ] pa'i-go ni-mõ
1SG see-V.P.N3M/F woman big-F.SG COP-V.P.F.SG
'The woman I see is big.'
- f. *[[yɪ'i ãñã(-ã)-ri] nũmĩõ] pa'i-go ni-mõ
1SG s ee-NZR woman big-F.SG COP-V.P.F.SG
'The woman I see is big.'
- g. *[[yɪ'i ãñã(-ã)-ri-go] nũmĩõ] pa'i-go ni-mõ
1SG see-NZR-F.SG woman big-F.SG COP-V.P.F.SG
'The woman I see is big.'
- h. [[yɪ'i ññã(-ã)-gõ] nũmĩõ] pa'i-go ni-mõ
1SG see-F.SG woman big-F.SG COP-V.P.F.SG
'The woman I see is big.'

Past baseline structures

- a. yɪ'i ññã-api⁵¹ ni-kõ nũmĩõ
1SG see-V.PS.N3M/F one-F.SG woman
'I saw a woman.'

⁵¹The past *ññã-api-ri* form is ungrammatical as the present *ññã-ã-ri* form. They are not listed in the data below.

- b. nũmĩõ pa'i-go ni-amõ
 woman big-F.SG COP-V.PS.F.SG
 'The woman was big.'

NP use

- a. *[yĩ'i ññã-ãpĩ] pa'i-go nũmĩõ ni-amõ
 1SG see-V.P.N3M/F big-F.SG woman COP-V.PS.F.SG
 'The one I saw was a big woman.'
- b. [[yĩ'i ññã-'ko⁵²] pa'i-go nũmĩõ ni-amõ
 1SG see-PST.F.SG big-F.SG woman COP-V.PS.F.SG
 'The one I saw was a big woman.'

Modification use

- c. *[[yĩ'i ññã-ãpĩ] nũmĩõ] pa'i-go ni-amõ
 1SG see-V.P.N3M/F woman big-F.SG COP-V.PS.F.SG
 'The woman I saw was big.'
- d. [[yĩ'i ññã-'ko] nũmĩõ] pa'i-go ni-amõ
 1SG see-PST.F.SG woman big-F.SG COP-V.PS.F.SG
 'The woman I saw was big.'

Inanimate

Present

Baseline structures

- a. yĩ'i ññã-ã nika-gi yukigi
 1SG see-V.P.N3M/F one-CLF_{TREE} tree
 'I see a tree.'
- b. yukigi pa'i-ki ni-i
 tree big-CLF COP-V.P.N3M/F
 'The tree is big.'

NP use

- a. *[[yĩ'i ññã-ã] pa'i-ki yukigi ni-i
 1SG see-V.P.N3M/F big-CLF tree COP-V.P.N3M/F
 'The one I see is a big tree.'
- b. *[[yĩ'i ññã(-ã)-ri] pa'i-ki yukigi ni-i
 1SG see-V.P.N3M/F-NZR big-CLF tree COP-V.P.N3M/F
 'The one I see is a big tree.'

⁵² We did not observe the laryngealization in the past forms, but we follow the transcriptions by Ramirez (1997) and Welch and West (2000).

- c. *[[yɪ'i ɪ̃nã(-ã)-ri-ki] pa'i-ki yukigi ni-i
 1SG see-V.P.N3M/F-NZR-CLF big-CLF tree COP-V.P.N3M/F
 'The one I see is a big tree.'

- d. [[yɪ'i ɪ̃nã(-ã)-ki] pa'i-ki yukigi ni-i
 1SG see(-V.P.N3M/F)-CLF big-CLF tree COP-V.P.N3M/F
 'The one I see is a big tree.'

Modification use

- e. *[[yɪ'i ɪ̃nã(-ã)] yukigi] pa'i-khi
 1SG see(-V.P.N3M/F) tree big-CLF
 ni-i
 COP-V.P.N3M/F
 'The tree I see is big.'

- f. *[[yɪ'i ɪ̃nã(-ã)-ri] yukigi] pa'i-ki ni-i
 1SG see(-V.P.N3M/F)-NZR tree big-CLF COP-V.P.N3M/F
 'The tree I see is a big tree.'

- g. *[[yɪ'i ɪ̃nã(-ã)-ri-ki] yukigi] pa'i-ki ni-i
 1SG see(-V.P.N3M/F))-NZR-CLF tree big-CLF COP-V.P.N3M/F
 'The tree I see is a big.'

- h. [[yɪ'i ɪ̃nã(-ã)-ki] yukigi] pa'i-khi ni-i
 1SG see(-V.P.N3M/F)-CLF tree big-CLF COP-V.P.N3M/F
 'The tree I see is a big.'

Past Baseline structures

- a. yɪ'i ɪ̃nã-api nika-gi yukigi
 1SG see-V.PA.N3M/F one-CLF tree
 'I saw a tree.'

- b. yukigi pa'i-khi ni-api
 tree big-CLF COP-V.PA.N3M/F
 'The tree was big.'

NP use

- a. *[[yɪ'i ɪ̃nã-api] pa'i-khi yukigi ni-api
 1SG see-V.PA.N3M/F big-CLF tree COP-V.PA.N3M/F
 'The one I saw was a big tree.'

- b. [yĩ'i ññã-'ki] pa'i-khi yukigi ni-api
 1SG see-PST.CLF big-CLF tree COP-V.PS.N3M/F
 'The one I saw was a big tree.'

Modification use

- c. *[[yĩ'i ññã-api] yukigi] pa'i-ki ni-api
 1SG see- V.PS.N3M/F tree big-CLF COP-V.PS.N3M/F
 'The tree I saw was big.'

- d. [[yĩ'i ññã-'ki] yukigi] pa'i-ki ni-api
 1SG see-PST.CLF tree big-CLF COP-V.PS.N3M/F
 'The tree I saw was big.'

The Tukano marking patterns of V-based nominalizations are summarized as in Tables 16 and 17.

Table 16. NP-use of V-based nominalizations in Tukano

Animate	*-Ø, *-ri, *-ri-M/F, -M/F
Inanimate	*-Ø, *-ri, *-ri-CLF, -CLF

Table 17. Prenominal modification use of V-based nominalizations in Tukano⁵³

Animate	*-Ø, *-ri, *-ri-M/F, -M/F
Inanimate	*-Ø, *-ri, *-ri-CLF, -CLF

A striking feature of the Tukano summaries above is the perfect convergence in the marking patterns of animate and inanimate classifiers, reinforcing the view that these are not distinct grammatical phenomena. Equally notable is the consistent absence of the nominalizer *-ri* in V-based nominalizations, whether functioning as NP heads or modifiers. This contrasts with Tuyuka, where inanimate nominalizations may retain *-ri* in both present and past tense forms, as illustrated in the comparison in (50) below.

(50) Tuyuka-Tukano comparison on *-ri* marking

- Tuyuka a. [ñãdia-ri] yukigi yaa-gi ni- ã
 fall-NZR tree 1SG.GEN-CLF COP-V.P.O
 'The tree falling is mine.'

⁵³ Postnominal modification patterns alike and the relevant data and summaries are omitted.

Tukano	a'. %[biridhia- ri] fall-NZR	yukigi tree	ya-gi 1SG.GEN-CLF	ni-i ⁵⁴ COP-V.P.N3M/F
	'The tree falling is mine.'			
Tuyuka	b. [ñãdia- ri -gi] fall-NZR-CLF _{TREE}	yukigi tree	yaa-gi 1SG.GEN-CLF	ni-ã COP-V.P.O
	'The tree falling is mine.'			
Tukano	b'. *[biridhia- ri -ki] fall-NZR-CLF	yukigi tree	ya-gi 1SG.GEN-CLF	ni-i COP-V.P.N3M/F
	'The tree falling is mine.'			
Tuyuka	c. [ñãdia-gi] fall-CLF _{TREE}	yukigi tree	yaa-gi 1SG.GEN-CLF	ni-ã COP-V.P.O
	'The tree falling is mine.'			
Tukano	c'. [biridhia-ki] fall-CLF _{TREE}	yukigi tree	ya-gi 1SG.GEN-CLF	ni-i COP-V.P.N3M/F
	'The tree falling is mine.'			

As it so turns out, Tukano also retains some *-ri* nominalization constructions. These are identified by Ramirez (1997: 207) as “inanimado (forma básica)” (inanimate (basic form)), and by Welch and West (2000: 429) as “sustantivos (nominalizados) de la clase *-ri*” ((nominalized) nouns of the *-ri* class) (see Appendices F and G). Our data include forms like those illustrated in (51) below, where the first in each pair represents the “regular” pattern and the second that of the *-ri*-class form.

(51) Tukano V-based nominalizations marked by *-ri*

Present tense

a.	%[[biridhia- ri] fall-NZR	pigi hammock	ya-gi 1SG.GEN-CLF	ni-i COP-V.P.N3M/F
	'The hammock falling down is mine.'			
a'.	[biridhia- ri] fall-NZR	wi'i house	ya-wi 1SG.GEN-CLF	ni-i COP-V.P.N3M/F
	'The house falling down is mine.'			

⁵⁴ Our Tuyuka-Tukano-Portuguese trilingual consultant (born 1985), who grew up and lived in Puerto Colombia, Colombia over 20 years before moving to Villa Tuyuka, San Gabriel da Cachoeira, Amazonas, Brazil, often accepts this type of modifier with *-ri*, but occasionally disallows it. A younger Tukano-Portuguese bilingual speaker (born 1994), who grew up in San Jose II, Medio Tikie, Amazonas, Brazil, rejects this type of modifier with *-ri*. Perhaps the former speaks a more conservative variety of Tukano or her Tukano is influenced by Tuyuka.

b. *[biridhia-**ri-ki**] pigi ya-gi ni-i
 fall-NZR-CLF hammock 1SG.GEN-CLF COP-V.P.N3M/F
 ‘The hammock falling down is mine.’

b'. [biridhia-**ri-wi**] wi'i ya-wi ni-i
 fall-NZR-CLF house 1SG.GEN-CLF COP-V.P.N3M/F
 ‘The house falling down is mine.’

c. [biridhia-**ki**] pigi ya-gi ni-i
 fall-CLF hammock 1SG.GEN-CLF COP-V.P.N3M/F
 ‘The hammock falling down is mine.’

c'. *[biridhia-**wi**] wi'i ya-wi ni-i
 fall-CLF house 1SG.GEN-CLF COP-V.P.N3M/F
 ‘The house falling down is mine.’

Past tense

d. [biridhia-'**ki**] pigi ya-gi ni-api
 fall-PST.CLF hammock 1SG.GEN-CLF COP-V.PS.N3M/F
 ‘The hammock that was falling down was mine.’

d'. *[biridhia-'**wi**] wi'i ya-wi ni-api
 fall-PST.CLF house 1SG.GEN-CLF COP-V.PS.N3M/F
 ‘The house that was falling down was mine.’

e. *[biridhia-**ka-ki**] pigi ya-gi ni-api
 fall-PST-CLF hammock 1SG.GEN-CLF COP-V.PS.N3M/F
 ‘The hammock that was falling down was mine.’

e'. [biridhia-**ka-wi**] wi'i ya-wi ni-api
 fall-PST-CLF house 1SG.GEN-CLF COP-V.PS.N3M/F
 ‘The house that was falling down was mine.’

In the past tense, the *-wi* classifier does not take the laryngealized past-tense form (51d') and instead requires the analytic past tense construction *V-ka-CLF*. In contrast, regular forms use the laryngealized past tense (51d) and shun the analytic construction (51e).⁵⁵

The size of the *-ri* class forms remains unclear. However, given that West and Welch(2000) set them out as a special class of “nominalized nouns” (see Appendix G) and that they are absent from West’s (1990) table of suffixes for V-based nominalizations, it may be reasonable to assume that they are residual features from an earlier stage in the evolution of Tukano classifier marking.

⁵⁵ It is unclear whether the *-ri* class forms are defined by the initial sound of classifiers or by lexically, e.g., “house”, etc.

7. DYNAMIC FUNCTIONAL ACCOUNT OF THE KOTIRIA-TUYUKA-TUKANO V-BASED NOMINALIZATIONS

Table 18 presents the crosslinguistic patterns of classifier marking in V-based nominalizations used for prenominal modification in Kotiria, Tuyuka, and Tukano. Prenominal modification is selected for analysis because it exhibits the greatest diversity in marking patterns.

Table 18. Crosslinguistic marking patterns in prenominal modification of V-based nominalizations

Kotiria⁵⁶

Animate	-Ø, -ri, -ri-M/F, -M/F
Inanimate	-Ø, -ri, -ri-CLF, -CLF

Tuyuka

Present

Animate	*-Ø, *-ri, *-ri-MF, -M/F
Inanimate	*-Ø, -ri, -ri-CLF, *-CLF

Past

Animate	*-Ø, *-ri, -ri-M/F, *-M/F
Inanimate	*-Ø, -ri, -ri-CLF, *-CLF

Tukano

Animate	*-Ø, *-ri, *-ri-M/F, -M/F
Inanimate	*-Ø, *-ri, *-ri-CLF, -CLF

Given the crosslinguistic typology of classifier marking-patterns presented in Table 18, Shibatani's dynamic model of functional typology calls for a diachronic account—or at least a hypothesis—of the evolutionary paths that different languages have followed to arrive at their current synchronic marking patterns. In this penultimate section, we take up this task.

Our account hypothesizes that the Kotiria pattern represents the beginning of a new cycle in the development of classifier marking, characterized by the greatest variety of classifier-

⁵⁶ Both Kotiria and Tukano show the same marking patterns in both present and past tense; and, therefore, only one table is given for them, while two separate tables are presented for Tuyuka, which shows distinct marking patterns depending on the tense.

marking strategies. These include: (i) no overt marking of V-based nominalizations (-Ø), and the use of finite verb forms themselves as nominalizations; (ii) the use of the nominalizer *-ri* alone; (iii) combinations of *-ri* and classifiers (*-ri-M/F*, *-ri-CLF*); and (iv) the use of classifiers alone (*-M/F*, *-CLF*), without any overt nominalizer. Notable also is that no distinctions are made between animate and inanimate classifiers in these marking options. To the best of our knowledge, Kotiria—along with possibly its closest sister language, Wa'ikhana—is the only documented Tukanoan language to exhibit all these possibilities.

The key distinction in marking patterns across the three languages lies in the behavior of the nominalizer suffix *-ri*, which is widely used in Kotiria but largely absent in Tukano. Tuyuka represents an intermediate stage in the evolution of classifier marking: the nominalizer *-ri* has been lost in present-tense contexts with animate classifiers, resulting in a classifier-only marking pattern—an option not available for inanimate classifiers. This innovation marks the initial divergence between animate and inanimate classifier marking. In past-tense contexts, *-ri* is generally retained with both classifier types, although animate classifiers show a stronger tendency toward *-ri* loss across both tenses. This trend leads to growing disparities between animate and inanimate marking patterns.

Alongside Tuyuka, many Eastern Tukanoan languages exhibit variant intermediate patterns of classifier marking. As the comparison between Tuyuka and Yuruti in Appendices B and C shows, Yuruti displays a system highly comparable to Tuyuka in its pattern of *-ri* loss and retention.

In contrast, Siriano and Desano represent a more advanced stage in the loss of *-ri*, as shown in Appendices D and E, where *-ri* is not seen for all animate tense categories. It is likely that the past-tense markers *-di*, *-dii*, and *-dee* in these languages have evolved from the nominalizer *-ri*, possibly due to its association with past-tense contexts for animate classifiers, as seen in Tuyuka and Yuruti. This shift likely created gaps in *-ri* marking within past-tense contexts. As in Appendix E, Desano (Miller 1999: 140) retains *-ri* for inanimate in all tenses, unlike Siriano, which has lost *-ri* (or gave rise to the tense interpretation) in the past tense.

Finally, as previously noted, Tukano has lost the nominalizing suffix *-ri* in both animate and inanimate V-based nominalizations—except for forms that belong to the *-ri* class, as shown in Ramirez's "inanimado/inanimate" row in Appendix F and Welch and West's (2000) last row for the *ri*-class forms in Appendix G. This change restores a harmonious marking pattern between animate and inanimate classifiers.

Naturally, we are interested in whether any other Tukanoan languages have reached a similar final stage of development. One possible parallel is Tatuyo, studied extensively by Gomez-Imbert (1982, 2007), who claims that in this language, "[i]n most respects, agreement ... fits canonical agreement, which is syntactically simple and **obligatory** in two domains: at the sentence level and in noun phrases" (2007: 403; emphasis added); "noun phrase agreement is syntactically simple because it has **no exceptions**: it targets all modifiers of the noun controller in a noun phrase [for both animate and inanimate classifiers]" (ibid. p. 405; emphasis added)⁵⁷ Languages that exhibit obligatory inanimate-classifier marking across all structural categories (e.g., numerals, demonstratives, etc.) in both NP- and modification-use contexts

⁵⁷ NB: In these quotations, Gomez-Imbert is referring to both animate and inanimate classifiers, indicating that both exhibit "obligatory" agreement when classifier-marked structures are used as noun modifiers.

represent the most advanced stage in the development of classifier systems. Tukano may be approaching this stage as well, although some variation is evident among speakers. The youngest Tukano speaker (born in 1994) in our study, who speaks no other indigenous languages, consistently uses animate and inanimate classifiers across all modifying structures, except demonstratives, whose marking remains optional⁵⁸.

A final task in a dynamic, functionally-oriented typological study is to explore the functional motivations underlying the hypothesized paths of diachronic change, as observed in Table 18 above. Four major developments can be identified in the evolution of V-based nominalization and associated marking strategies in the languages under discussion⁵⁹:

1. **Obligatory nominalization marking**

The introduction of a dedicated nominalizer *-ri* on V-based nominalizations:

Kotiria $\text{-}\emptyset > \text{Kotiria (Tuyuka, Tukano) } \text{-ri}$.

2. **Classifier marking**

The addition of a classifier onto *-ri*-marked nominalizations:

Kotiria (Tuyuka, Tukano) $\text{-ri} > \text{-ri-M/F, -ri-CLF}$

3. **Nominalizer loss before classifiers, 1**

The dropping of the *-ri* nominalizer preceding an animate classifier in the present (and other) tense(s):

Kotiria $\text{-ri-M/F(PRS)} > \text{Tuyuka -M/F (PRS)}$

4. **Nominalizer loss before classifiers, 2**

The subsequent erosion of the *-ri* nominalizer when it directly precedes a classifier

Kotiria $\text{-ri-M/F, -ri-CLF} > \text{-M/F, -CLF}$

Tuyuka $\text{-ri-M/F (PST)} > \text{Tukano -M/F}$

Tuyuka $\text{-ri-CLF} > \text{Tukano -CLF}$

In the exploration of the diachrony of nominalization markers, including ordinary (non-classifying) markers in Japanese and Ryukyuan dialects and gender and classifier marking in South Asian languages (Indo-Aryan and Dravidian), Shibatani (2023) and Shibatani and Nitta (2025)—inspired by André Martinet’s theory of two opposing forces in language change (see the articles cited for details)—invoke the dual principles of **speaker’s economy**, which favors simplification and uniformity, and **hearer’s economy**, which promotes structural diversity to aid interpretation. These principles also map neatly onto the Tukanoan developmental stages outlined above. In Step 1, the introduction of the overt nominalizer *-ri* formalizes the distinction between finite verb forms (used predicatively) and V-based nominalizations in favor of the hearer. In Step 2, extending classifier marking to these *-ri*-nominalizations consolidates the system: a single classifier paradigm now uniformly applies from numerals through demonstratives all the way to V-based nominalizations. This uniform marking not only streamlines speakers’ production but also eases hearers’ parsing by providing a single, consistent cue to nominalization across all underlying structural categories involved. Moreover, converting ordinary nominalizations into classifying nominalizations provides additional

⁵⁸ This gap requires some explanation.

⁵⁹ All the posited processes remain optional for Kotiria for now.

information about the nature of the referents, thereby aiding the hearer in narrowing the search domain.

Step 3 is a crucial stage, where animate-classifier and inanimate-classifier marking begins to diverge, first by dropping the nominalizer *-ri* in a specific tense context, like the present tense contexts in Tuyuka, and subsequently expanding the dropping contexts, as in Siriano and Desano. Also, this stage is where we observe that animate classifiers outpace inanimate classifiers in turning animate classifiers to sole nominalizers (in the absence of the nominalizer *-ri*), which can be considered a case of grammaticalization. This change creates a divergent pattern of marking for animate (-M/F) and inanimate (*-ri*-CLF) in favor the hearer's economy.

Finally, the reduction of the *-ri* nominalizer before a classifier in Step 4 eliminates redundant nominalization marking and promotes formal uniformity across the nominalization domain (e.g., NUM-CLF, DEM-CLF ... V-NZR-CLF → NUM-CLF, DEM-CLF...V-CLF), aligning with principles of speaker economy. A similar pattern of nominalizer reduction occurs in several East Asian classifier languages—such as Zauzou (Tibeto-Burman), Hmong (Miao-Yao), and Southern Min and Yue varieties of Chinese—where nominalizers are routinely omitted before classifiers (Shibatani, Pardeshi, & Jiang 2023). As noted earlier, comparable functional motivations have been proposed for the reduction of non-classifying nominalizers in Japanese and Ryukyuan dialects as well (Shibatani & Nitta 2025), underscoring a broader typological link between ordinary nominalizers and classifiers (including so-called gender markers).

8. CONCLUSION

In this paper, we have argued that Tukanoan languages challenge the conventional division between animacy-based classifiers and those based on physical or functional attributes—a dichotomy typically framed in terms of *gender* versus *classifier* systems (cf. the global typological surveys in *World Atlas of Linguistic Structures* 2005/WALS Online 2013). Instead, our analysis supports a unified approach to these systems. In line with the morphosemantic continuum proposed in earlier studies, the syntactic marking patterns examined here reveal substantial distributional overlap and functional convergence between animate and inanimate classifiers.

Whether animacy-based distinctions constitute a nominal classification system separate from those based on physical and functional attributes remains an empirical question. The mere presence of special grammatical treatment of animacy-based contrasts (e.g., obligatory marking or agreement), however, does not, in itself, justify a categorical division; rather, it may simply reflect a more advanced stage of grammaticalization—much like some fourth graders progress more quickly than their peers.

The broader and more productive task, then, is to analyze and compare the grammatical behavior of the full range of features underlying nominal classification systems—such as animate, human, masculine, round, flat, long, flexible, vertical, vehicle, or utensil—and to determine whether any of these features receive privileged grammatical treatment over others. For example, many classifier languages of Asia, such as Thai and Zhuang (both Kra-Dai languages), include animacy-based classifiers (e.g., animal, human, masculine, feminine), yet

these do not seem to exhibit grammatical behavior that clearly distinguishes them from inanimate classifiers. In contrast, many Indo-Aryan and Dravidian languages—like the Tukanooan languages studied here—grant varying degrees of grammatical privilege to animate classifiers over inanimate ones. Still other languages, such as those in the Indo-European family, typically characterized as gender languages, focus narrowly on animacy-based distinctions.

These cross-linguistic patterns suggest that the world's languages do not conform neatly to a binary opposition between gender and classifier systems, but instead exhibit gradual patterns of grammatical asymmetry between animacy-based features and inanimate features. The goal of typology is to identify such asymmetries and to provide explanations—whether cognitive, functional, or otherwise—for why certain features tend to become integrated into grammar (i.e., grammaticalize) more readily than others (see the abstract of this paper for our proposed reasoning).

ABBREVIATIONS

1SG	first person singular
2SG	second person singular
3SG	third person singular
1F.SG	first person feminine singular
2F.SG	second person feminine singular
3F.SG	third person feminine singular
1PL	first person plural
3PL	third person plural
ADJ	adjective
ANI	animate
ART	article
ASSERT	assertive
CLS	classifier
PL.CLS	plural classifier
CLASS	classifier
CLF	classifier
COLL	collective
COP	copula
DEIC	deictic
DEM	demonstrative
DEP	dependent
DISTAL	distal
EMPH	emphatic
EVI.PST	evidential past tense
F	feminine
IMP(ERF)	imperfective
INA	inanimate
INDEF	indefinite
M	masculine

Mod	modification
^M OPT	marginal optional marking
N	neutral or non-human
NON3M/F	non-3rd person masculine/feminine
NP	noun phrase
NUM	numeral
NZR	nominalizer
OBL	obligatory marking
OBL/NO	obligatory or no marking
OPT	optional marking
PERF	perfective
PL	plural
PoN	postnominal
PrN	prenominal
PROX	proximal
PRS	present tense
PST	past tense
PST.DISTANT	distant past
SG	singular
V.IMP.1	visual, imperfective, first person
V.IMP.N1	visual, imperfective, non-first person
V.P.N3M/F	visual, present, non-third person masculine/feminine
V.P.F.SG	visual, present, third person, feminine
V.PA.N3M/F	visual, past, non-third person, masculine/feminine
V.PA.N3M/F	visual, past, non-third masculine/feminine
V.P.O	visual, present tense, other person than 3M.SG, 3F.SG and 3PL
V.PA.O	visual, past tense, other person than 3M.SG, 3F.SG and 3PL
V.PST	visual past tense

ACKNOWLEDGMENTS

The present study has been supported by the project “Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective” (PI: Prashant Pardeshi) of National Institute for Japanese Language and Linguistics in Tokyo.

REFERENCES

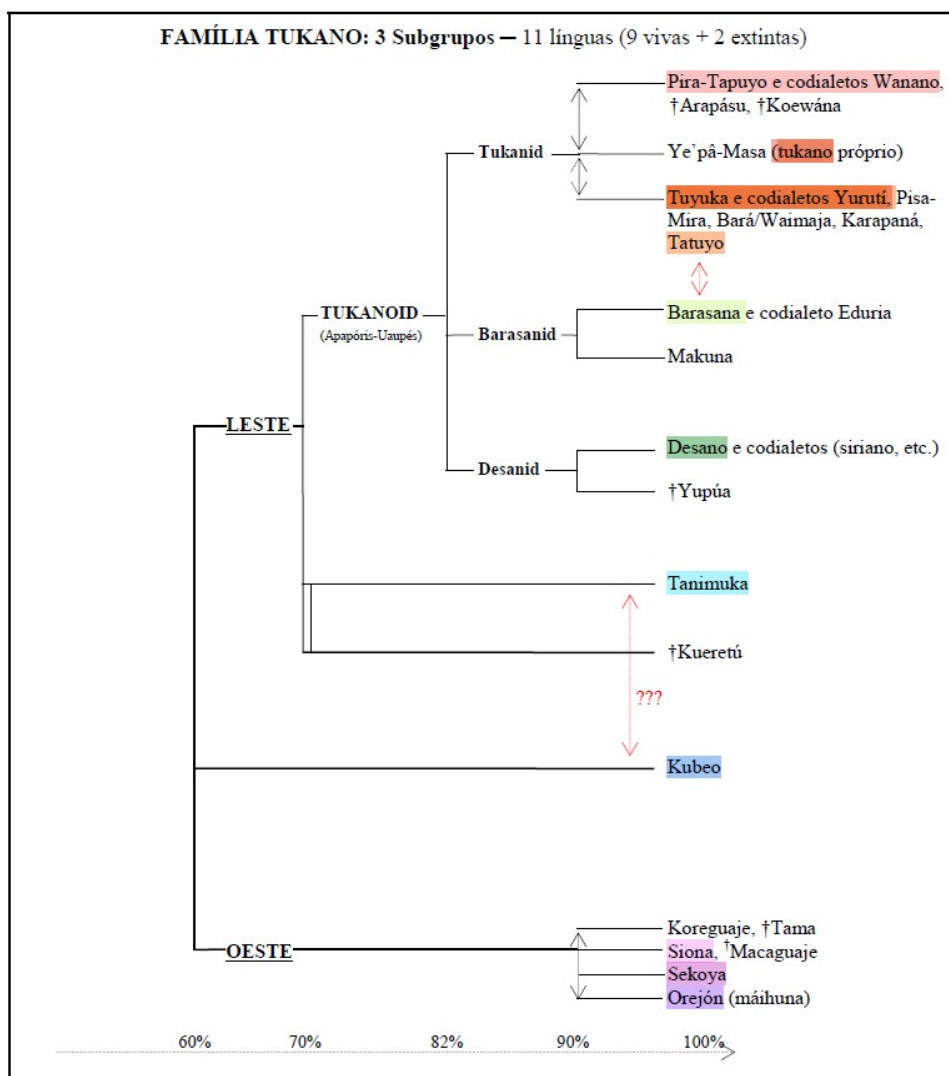
- Aikhenvald, Alexandra Y. 2007. Classifiers in multiple environments: Baniwa of Içana/Kurripako—A North Arawak perspective. *International Journal of American Linguistics* 73(4), 475–500.
- Aikhenvald, Alexandra Y. 2019. Noun categorization devices: A cross-linguistic perspective, In Alexandra Y Aikhenvald and Elena I. Mihás (eds.) *Genders and classifiers: A cross-linguistic typology*, 1–29. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2000. *Classifiers. A typology of noun categorization devices*. Oxford: Oxford University Press.

- Allan, Keith. 1977. Classifiers. *Language* 53, 285–311.
- Barnes, Janet. 1984. Evidentials in the Tuyuka Verb. *International Journal of American Linguistics* 50(3), 255–271.
- Barnes, Janet. 1990. Classifiers in Tuyuka. In Doris L. Payne (ed.), *Amazonian linguistics: Studies in lowland south American language*, 273–292. Austin: University of Texas Press.
- Barnes, Janet. 1999. Tucano. In R. M. W. Dixon and Alexandra Y. Aikhenvald (eds.) *The Amazonian languages*, 207–225. Cambridge: Cambridge University Press.
- Barnes, Janet, and Sheryl Takagi. 1976. Fonología del Tuyuka. *Sistemas fonológicos de idiomas colombianos* 3, 123–137. Bogotá: Ministerio de Gobierno.
- Barnes, Janet, and Terrel Malone. 2000. El tuyuka. In María Stella González de Pérez and María Luisa Rodríguez de Montes (eds.) *Lenguas Indígenas de Colombia: una visión descriptiva*, 37–452. Santa Fé de Bogotá: Instituto Caro y Cuervo.
- Bruil, Martine. 2014. *Clause-typing and evidentiality in Ecuadorian Siona*. LOT Dissertation Series 350. Leiden: Leiden University.
- Chacon, Thiago Costa. 2012. *The phonology and morphology of Kubeo: The documentation, theory, and description of an Amazonian language*. Doctoral dissertation. University of Hawai'i at Mānoa.
- Chacon, Thiago Costa. 2022. Gender and noun class markers in Kubeo (Tukanoan). *Faits de Langues* 52(2), 707–805.
- Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.
- Criswell, Linda, and Beverly Brandrup. 2000. Un bosquejo fonológico y gramática del Siriano. In María Stella González de Pérez and María Luisa Rodríguez de Montes (eds.) *Lenguas Indígenas de Colombia: una visión descriptiva*, 395–418. Santa Fé de Bogotá: Instituto Caro y Cuervo.
- Crystal, David. 1980 *A dictionary of linguistics and phonetics*. Oxford: Wiley-Blackwell.
- Dryer, Matthew S. and Martin Haspelmath (eds.) *World atlas of linguistic structure* (2005)/WALS Online (2013WALS Online (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591> (Available online at <https://wals.info>, Accessed on 2025-06-23.)
- Eraso, Keller, and Natalia María. 2015. *Gramática tanimuka, una lengua de la Amazonía colombiana*. Doctoral dissertation. Université Lumière Lyon 2.
- Farmer, Stephanie. 2015. Establishing reference in Máihiki. Doctoral dissertation. University of California, Berkeley.
- Frank, Michael C., Daniel L. Everett, Evelina Fedorenko, and Edward Gibson 2008. Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108(3), 819–824 DOI:10.1016/j.cognition.2008.04.007
- Gomez-Imbert, Elsa. 1982. *De la forme et du sens dans la classification nominale en tatuyo (langue tukano orientale d'Amazonie colombienne)*. Doctoral dissertation. Université de Paris-Sorbonne IV.
- Gomez-Imbert, Elsa. 2007. Tukanoan nominal classification: The Tatuyo system. In Leo Wetzels (ed.) *Language endangerment and endangered languages: Linguistic and anthropological studies with special emphasis on the languages and cultures of the Andean-Amazonian border area*. 401–428. Leiden: CNWS Publications.

- González de Pérez, María Stella, and Rodríguez de Montes, María Luisa (eds). 2000. *Lenguas Indígenas de Colombia: una visión descriptiva*. Santa Fé de Bogotá: Instituto Caro y Cuervo.
- Grinevald, Colette. 2000. A morphoynctactic typology of classifiers. In Gunter Senft (ed.) *Systems of nominal classification*, 50–92. Cambridge: Cambridge University Press.
- Jones, Wendell, and Paula Jones. 1991. *Barasano Syntax*. Studies in the Languages of Colombia 2. Dallas: The Summer Institute of Linguistics and The University of Texas at Arlington.
- Kinch, Rodney, and Pamela Kinch. 2000. El yuruti. In María Stella González de Pérez and María Luisa Rodríguez de Montes (eds.) *Lenguas Indígenas de Colombia: una visión descriptiva*, 469–488. Santa Fé de Bogotá: Instituto Caro y Cuervo.
- Miller, Marion. 1999. *Desano grammar (Studies in the indigenous languages of Colombia 6)*. Summer Institute of Linguistics and the University of Texas at Arlington.
- Panther, Klaus-Uwe, and Linda L. Thornburg (eds.) 2003. *Metonymy and pragmatic inferencing*. Amsterdam: John Benjamins.
- Payne, Doris, L. 1985. *Aspects of the grammar of Yagua: A typological perspective*. Doctoral dissertation. University of California, Los Angeles.
- Payne, Thomas. 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.
- Ramirez, Henri. 1997. *A Fala Tukano dos Ye'pa-Masa. Tomo I: Gramática*. Manaus: CEDEM.
- Rose, Françoise. 2024. Mojeño Trinitario classifiers: a multi-loci and multi-functional system. Book manuscript.
- Shibatani, Masayoshi. 2019. What is nominalization? Towards theoretical foundations of nominalization. In Roderto Zariquiey, Masayoshi Shibatani, and David W. Fleck (eds.) *Nominalization in languages of the Americas*, 15–168. Amsterdam: John Benjamins.
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni [Grammar of adnominal modification: Centering on classifiers and gender]. In Sung-Yeo Chung and Masayoshi Shibatani (eds.) *Taigenka riron to gengobunseki* [Nominalization theory and linguistic analysis], 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.
https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/pdf/-char/en
- Shibatani, Masayoshi, and Tetsuo Nitta. 2025. Japonic nominalizations in cross-dialectal perspective. In Nobuko Kibe, Tetsuo Nitta, and Kan Sasaki (eds.) *Handbook of Japanese dialects*, 397–450. Berlin: Mouton de Gruyter.
- Shibatani, Masayoshi, and Sung Yeo Chung. 2018. Nominal-based nominalization. *Japanese/Korean Linguistics* 25, 63–88.
- Shibatani, Masayoshi, Prashant Pardeshi, and Haowen Jiang. 2023. Nominalization-classifier continua and overlaps in languages of China. Paper presented at the International Workshop “Classifiers in Asian Languages,” Zhejiang University, Hangzhou, China, May 20, 2023.
- Silva, Wilson de Lima. 2012. *A descriptive grammar of Desano*. Doctoral dissertation, University of Utah.

- Stenzel, Kristine Sue. 2004. *A reference grammar of Wanano*. Doctoral dissertation. University of Colorado. (Published as: Stenzel, Kristine. 2013. *A reference grammar of Kotiria (Wanano)*. Lincoln: University of Nebraska Press.)
- Steele, Susan. 1978. Word order variation: A typological study. In Joseph Greenberg, Charles A. Furguson, and Edith A. Moravcsik (eds.) *Universals of human language, Vol. 4: Syntax*, 85–623. Stanford: Stanford University Press.
- Strom, Clay. 1992. *Retuarã syntax (Studies in the languages of Colombia 3)*. Dallas: The Summer Institute of Linguistics & The University of Texas at Arlington.
- Vallejos, Rosa. 2021. Nominal classification without grammatical agreement: Evidence from Sekoya. *International Journal of American Linguistics* 87(3), 423–455.
- Welch, Betty, and Birdie West. 2000. El tucano. In María Stella González de Pérez and María Luisa Rodríguez de Montes (eds.) *Lenguas Indígenas de Colombia: una visión descriptiva*, 419–436. Santa Fé de Bogotá: Instituto Caro y Cuervo.
- West, Birdie. 1980. *Gramatica Popular del Tucano*. Bogotá: Instituto Lingüístico de Verano, Ministerio de Gobierno-República de Colombia.

APPENDIX A: Tukanoan family (Ramirez 1997: 6)



APPENDIX B: Tuyuka (Barnes and Malone 2000: 445)

	ANIMADO			INANIMADO			
	Singular		Plural	Contable		No contable	De tiempo y lugar
	masculino	femenino		singular	plural		
PRESENTE	-gi	-go	-ra	-ri+ CLS	-re+ CLS	-re	-ro
PASADO	-rigi	-rigo	-rira			-rige	-riro
PASADO RECIENTE	-arigi	-arigo	-arira			-arige	-ariro
FUTURO	-idi	-odo	-adara			-adare	-adaro

APPENDIX C: Yuruti (Kinch and Kinch 2000: 476)

	ANIMADO			INANIMADO			
	Singular		Plural	Contable		No contable	De tiempo y lugar
	Masc.	Fem.		Singular	Plural		
PRESENTE	-gi	-go	-ra	-ri + CLS -ria + CLS	-re + PL.CLS	-re -ri	-ro
PASADO	-rigi	-rigo	-riara	-ria + CLS	-ria + PL.CLS -rige + PL.CLS	-rige	-riaro
PASADO RECIENTE	-arigi	-arigo	-ariara	-aria + CLS	-aria + PL.CLS -arige + PL.CLS	-arige	-ariaro
FUTURO	-akī	-ako	-adā	-ati + CLS	-ati + PL.CLS	-ate	-ato

APPENDIX D: Siriano (Criswell and Brandrup 2000: 408)

	ANIMADO		INANIMADO	
	objeto/idea	tiempo y lugar		
PRESENTE	masc. -gi fem. -go pl. -rā	-ri	-ro	
PASADO	masc. -dii-gi fem. -dee-go pl. -dēē-rā	-dee-a	-dee-ro	
FUTURO	masc. -bu-gi fem. -bu-go pl. -bū-rā	-bu-ri	-bu-ro	

APPENDIX E: Desano (Miller 1999: 140)

		animate	inanimate	
			concrete/ abstract noun	time/place/ concept
present	ms	-gi		
	fs	-go	-ri	-ro
	p	-rã		
past	ms	-di-gi		
	fs	-di-go	-ra (<ri-a)	-di-ro
	p	-di-rã		
future	ms	-bu-gi		
	fs	-bu-go	-bu-ri	-bu-ro
	p	-b-õ-rã		

APPENDIX F: Tukano (Ramirez 1997: 207)

	Simultâneo (presente, habitual)	Anterior (perfectivo)	Posterior (futuro relativo)
Animado/-fsg	-gí	-'kí	-akíhi
Animado/+fsg	-gó	-'ko	-akoho
Animado/pl	-rã /-dã/	-'karã /-ka-dã/	-arãha /-a-dãha/
Inanimado (forma básica)	-ri+ /-di+ /	-'ka+	-atihi+
Inan.pl & n-cont.	-sehé / -sé	-'ke	-atehe
Inan.lugar (locativo)	-ró /-dó/	-'karo /-'ka-do/	-atoho
Inan. forma roliça	-kahá (sg) -sehépaga (pl)	-'kaga (sg) -'kepaga (pl)	-akaha (sg) -atehepaga (pl)
Inan. forma de panela	-tihi -sehépari /-sehé-pa-di/	-'kari /-'ka-di/ -'kepari /-'ke-pa-di/	-atihi -atehepari /-a-tehe-pa-di/
Inan. forma retilínea	-kíhi -sehépagi	-'kagi -'kepagi	-akihi -atehepagi
Inan. forma tubular e oca	-pihi -sehépawi	-'kawi -'kepawi	-apihi -atehepawi
Inan. forma de abóbada	-pahá -sehépawa	-'kawa -'kepawa	-apaha -atehepawa
Inan. forma de lago	-tahá -sehépara /-sehé-pa-da/	-'kara /-'ka-da/ -'kepara /-'ke-pa-da/	-ataha -atehepara /-a-tehe-pa-da/

APPENDIX G: Tukano (Welch and West 2000: 429)

	CONTINUATIVO	PERFECTIVO	POTENCIAL
ANIMADO			
masculino	-gi	-ʔki	-á-k ^h i
femenino	-go	-ʔko	-á-k ^h o
plural	-rã	-ʔkã-rã	-á-hã
INANIMADO			
general	-se	-ʔke	-á-t ^h e ⁴
neutro, ubicación	-ro	-ʔka-ro	-át ^h o
SUSTANTIVOS DE LA CLASE -pa			
forma esférica	ga/-ká/-k ^h á	-ʔka-ga	-á-k ^h a
olla	-ri/-ti/-t ^h i	-ʔka-ri	-á-t ^h i
forma larga y recta	-gi/-ki/-k ^h i	-ʔki	-á-k ^h i
vehículo o instrumento	-wi/-pi/-p ^h i	-ʔka-wi	-á-p ^h i
raudal o algo que cierre	-wa/-pá/-p ^h á	-ʔka-wa	-á-p ^h a
SUSTANTIVOS DE LA CLASE -ri	-ri	-ka	-át ^h i

Nominalization in Vietnamese: Roles of classifiers and the genitive marker *của*

Masaaki Shimizu¹ and Trang Phan²

ABSTRACT

This study examines the use of classifiers and the genitive marker *của* as devices for grammatical nominalization in Vietnamese, focusing on their distribution in both NP-use and modification-use contexts. The data on classifier marking largely supports Shibatani's (2021, 2023) proposed functional hierarchy, with one minor exception. In contrast, the genitive marker *của* shows partial similarity to Mandarin *de*, functioning as a nominalizer in both NP and modification uses. However, its use is semantically and contextually restricted. These findings contribute to our understanding of Vietnamese nominalization patterns and offer further insight into cross-linguistic typologies of nominalization.

1. INTRODUCTION

This study investigates two primary strategies of grammatical nominalization in Vietnamese: one involving the use of classifiers, and the other employing the genitive marker *của* as a nominalizing device. These elements are analyzed in terms of their distribution across NP-use and modification-use contexts. Examples (1) and (2) illustrate the roles of classifiers and *của* in nominal constructions:

- (1) Cái đang treo trong tủ là áo của tôi.
CLF PROG hang in closet COP clothes GEN 1SG
'What is hanging in the closet are my clothes.'
- (2) Quyển sách này là của thầy Nam soạn.
CLF book this COP GEN teacher Nam edit
'This book is the one Mr. Nam edited.'

Drawing on the typological framework proposed by Shibatani (2021: 524; 2023: 35), this study evaluates the extent to which Vietnamese adheres to the hierarchical pattern of classifier/gender marking. Vietnamese allows for classifier marking in all NP-use contexts outlined in the hierarchy—Numerals > Demonstratives > Genitives > Adjectives > Verb-based nominalizations—but restricts classifier marking in modification-use to numerals only. In contrast, the genitive marker *của* functions as a nominalizer in both NP and modification contexts for genitives and verb-based nominalizations, showing partial parallels with Mandarin

¹ Graduate School of Humanities, The University of Osaka, 3-5-10, Senba-Higashi, Mino, Osaka, 562-8678, Japan. email: shmz.hmt@osaka-u.ac.jp

² Department of Asian and North African Studies, Ca' Foscari University of Venice, Dorsoduro 3246, 30123 Venice, Italy. email: thihuyentrang.phan@unive.it

de (Shibatani 2018: 390-391). However, its use is more limited in both semantic range and contextual flexibility. This study explores these patterns in detail, contributing to the cross-linguistic understanding of nominalization mechanisms.

As outlined by Nguyen (2004), Vietnamese noun phrases consist of both pre-nominal and post-nominal modifiers. The pre-nominal elements include classifiers (CLF), measure phrases (MP), articles, the focus particle *cái*, numerals (Num), and quantifiers (Q). In contrast, post-nominal modifiers comprise noun adjuncts, adjective phrases, prepositional phrases, relative clauses, demonstratives, and possessive constructions. While the sequence of pre-nominal elements is structurally fixed, as illustrated in (3), the arrangement of post-nominal elements is more variable and depends on the context (see Phan & Duffield 2022 for discussion).

(3) Q – Num/Article – Particle *cái* – CLF/MP – N

The following section provides a detailed examination of each component of the noun phrase, with the exception of articles and the particle *cái*, both of which are uniquely Vietnamese and warrant separate preliminary discussion due to their distinctive properties.

Although the existence of lexical articles in Vietnamese remains a topic of debate, Nguyen (1975) and Nguyen (2004) argue that elements such as the plural markers *những* and *các*, as well as the numeral *một* ('one'), functionally encode definiteness and number. For a more nuanced account of these elements, see Phan & Lam (2021) and Phan & Chierchia (2022). Illustrative examples are provided in (4).

- (4) a. **những** con trâu
 PL CLF buffalo
 'the) buffaloes'
- b. **các** con trâu
 PL CLF buffalo
 'the buffalos'
- c. **một** con trâu
 one CLF buffalo
 'a buffalo'
- d. [zero] con trâu
 CLF buffalo
 'a/the buffalo' (NT³, p.104)

³ The abbreviations used to indicate data sources are as follows: NT = Nguyen. T. (2004)

Vietnamese features a function word homonymous with the classifier *cái*, which typically marks inanimate nouns, but distinct in its grammatical role. This particle *cái* appears between a numeral and a classifier and serves as a focus marker. An example is provided in (5):

- (5) Tôi muốn mượn hai **cái** cuốn sách kia kia.
 1SG want borrow two cái CLF book that there
 ‘I’d like to borrow those VERY two books OVER THERE’ (NT, p.48)

Importantly, the focus particle *cái* cannot precede the classifier *cái* itself, but it is obligatory when focus is applied to a noun that requires a classifier. The particle is always phonologically stressed, whereas classifiers are not. As Nguyen (2004: 52) notes, although focus in Vietnamese can be independently marked through prosody, *cái* functions as a formal syntactic device to indicate focus within the noun phrase. See also Simpson and Ngo (2018) for a more recent and detailed discussion of the particle *cái*.

In addition to the typical noun phrase structure described above, Vietnamese employs several nominalization strategies that derive noun phrases from other parts of speech. The most common among these are the nominalizers *sự*, *việc*, and *chuyện*, which originate from the Chinese nouns *shì* 事 ‘matter,’ *yì* 役 ‘work,’ and *zhuàn* 傳 ‘story,’ respectively. Unlike classifiers, these elements do not perform a classifying function. While *sự* functions strictly as a grammatical nominalizer and cannot serve as the head of a noun phrase, *việc* and *chuyện* retain more of their lexical meanings and may function as independent heads in NP constructions.

- (6) Sự phát triển của ngành công nghệ thông tin được đánh giá cao.
 NMLZR develop GEN sector IT PASS evaluate high
 ‘Development of the IT sector was highly praised.’
- (7) Việc học ngoại ngữ rất thú vị.
 NMLZR study foreign languages very interesting
 ‘To study foreign languages is very interesting.’
- (8) Chuyện nó không đến làm tôi vô cùng buồn.
 NMLZR he NEG come make 1SG extremely sad
 ‘That he did not come made me extremely sad.’

Another strategy involves zero-marked nominalization, where no explicit nominalizer is used. This commonly occurs in cleft constructions, which do not require nominalization markers (such as classifiers) when the referent is sufficiently identifiable from context. Illustrative examples are provided in (9)–(10).

- (9) Đang treo trong tủ đằng kia là áo của tôi.
 PROG hang in closet over there COP clothes GEN 1SG
 ‘What is hanging in the closet over there are my clothes.’

- (10) Tôi thích nhất là bún chả.
 1SG like most COP bun cha (vermicelli with grilled pork)
 ‘What I like most is bun cha.’

Another instance of zero-marked nominalization is found in complement clauses. Vietnamese complementizers such as *là* (derived from the copula) and *rằng* (originally a verb meaning ‘to say’) are frequently omitted, particularly in colloquial speech.

- (11) Tôi nghĩ (là/rằng) phương pháp này khả thi.
 1SG think COMP method this feasible
 ‘I think this method is feasible.’

In this study, data drawn from published sources are marked with standard abbreviations. Examples without explicit source attribution are original constructions based on the second author’s native-speaker intuition.

2. OVERVIEW OF CLASSIFIER MARKING AND GRAMMATICAL NOMINALIZATIONS IN VIETNAMESE

2.1 Classifier marking in Vietnamese

Nguyen D. H. (1957) compiled a list of approximately 200 classifiers used in Vietnamese. In a later dictionary (Nguyen D. H. 1971), we identified 1,005 lexical items categorized by these classifiers. However, as noted by Emeneau (1951), Nguyen D. H. (1957), Simpson & Ngo (2018), Phan (2019), Phan & Chierchia (2025), and others, not all Vietnamese nouns require or even permit classifier use when combined with numerals. Cao (1992), for instance, identified a set of 129 monosyllabic nouns that do not co-occur with classifiers. Additionally, Nguyen D. H. (1957: 132) observed that many compound nouns of Sino-Vietnamese origin (e.g., *kết quả* ‘result’, *ảnh hưởng* ‘influence’) and collective noun constructions denoting semantically related or synonymous pairs (e.g., *nhà cửa* ‘dwelling, house’; *giày dép* ‘footwear’; *chăn chiếu* ‘bedclothes’; *sách vở* ‘books and supplies’) also tend to appear without classifiers.

Classifier categories in Vietnamese are diverse; however, the most fundamental distinction is based on animacy and humanness. Nguyen (2004: 11) identifies a three-way animacy-based classification that serves as the core of the system.

- (12)
-
- ```

 graph LR
 ANIMACY --- Animate
 ANIMACY --- Inanimate["Inanimate: cái (CLF for things)"]
 Animate --- Human["Human: người ‘person’ (CLF for people)"]
 Animate --- NonHuman["Non-human: con (CLF for animals)"]

```

The fundamental distinctions among the core classifiers, particularly *con* and *cái*, become most apparent in contexts where the referent is unknown. When speakers lack specific knowledge about an object, they begin by categorizing it based on animacy, distinguishing between living and non-living entities. In such cases, *con* is typically used for animate (non-human) beings, while *cái* is used for inanimate objects.

- (13) a.    Con                này                là                con                gì?  
              CLF<sub>animate</sub>        this                COP                CLF<sub>animate</sub>        what  
              ‘What (animal) is this?’
- b.    Cái                này                là                cái                gì?  
              CLF<sub>inanimate</sub>        this                COP                CLF<sub>inanimate</sub>        what  
              ‘What (thing) is this?’

In contrast, the use of *người*, the classifier for humans, is more restricted and often optional.

- (14) a.    hai    con    chó                                hai    con    cá  
              two CLF dog                                two CLF fish  
              ‘two dogs’                                    ‘two fishes’
- b.    hai    cái    bàn                                hai    cái    kéo  
              two CLF table                                two CFL scissors  
              ‘two tables’                                    ‘two scissors’
- c.    hai    (người)    bạn                                hai    (người)    chị  
              two CLF        friend                                two CLF        sister  
              ‘two friends’                                    ‘two sisters’

Once the speaker is able to identify specific properties of the referent, classifiers are chosen more precisely. For instance, *con* continues to mark animals, while inanimate objects may be classified using more semantically specialized classifiers, such as *quả/trái* for round objects, *quyển/cuốn* for books or notebooks, and *tấm* for flat, two-dimensional items, etc.

## 2.2 Grammatical nominalizations in Vietnamese

Among the various strategies for grammatical nominalization in Vietnamese, one of the most prominent involves the use of classifiers. Classifiers in Vietnamese serve functions in both lexical and grammatical nominalization. In the case of lexical nominalization, the most typical are the so-called event classifiers, which constitute a closed set (Nguyen 2004: 186–193). These classifiers are used to nominalize verbs or verbal expressions by encoding events or actions as discrete entities.



- (15) a. một cuộc họp  
one CLF meet  
'a meeting'
- b. một kì thi  
one CLF exam  
'an exam'
- c. một giấc ngủ  
one CLF sleep  
'a sleep'
- d. niềm tự hào  
CLF proud  
'pride'
- e. một vụ ám sát  
one CLF assassinate  
'an assassination'

In contrast, the classifiers used for grammatical nominalization are drawn from a broader and more open set, extending beyond the restricted group of event-classifiers. In the sections that follow, we provide examples illustrating grammatical nominalization across different syntactic contexts.

## [1] Numerals

Vietnamese distinguishes between cardinal and ordinal numerals based on their syntactic position within the noun phrase. Cardinal numerals function as pre-nominal modifiers, appearing before the sequence {CLF + head noun}, whereas ordinal numerals are positioned post-nominally, following the same sequence.

### [1a] Cardinal numerals

- (16) a. Tôi muốn mua bốn con chó?  
1SG want buy four CLF dog  
'I want to buy four dogs.'
- b. Anh muốn mấy con chó?  
2SG want how.many CLF dog  
'How many dogs do you want?'

- c. Bốn con.  
four CLF  
'Four.'

### [1b] Ordinal numerals

- (17) a. Anh thích bài thơ thứ mấy?  
2SG like CLF poem which order  
'Which poem do you like?'
- b. Tôi thích bài thứ hai.  
1SG like CLF second  
'I like the second one.'
- c. Bài thơ thứ hai.  
CLF poem second  
'The second poem.'

### [2] Demonstratives

Demonstratives function as post-nominal modifiers in Vietnamese noun phrases. They can be categorized into three types based on their syntactic behavior: free forms, bound forms, and forms that can function as both (see Table 1). The demonstratives illustrated in the following examples are either bound or free/bound forms.

- (18) a. Con chó này to nhỉ.  
CLF dog this big SFP  
'This dog is big, isn't it?'
- b. Con này to. Con kia nhỏ.  
CLF this big CLF that small  
'This one is big.' 'That one is small.'
- c. Chó này to.  
dog this big  
'This (kind of) dog(s) is(/are) big.'

Table 1. Demonstratives in Vietnamese

|                     | Indefinite            | Proximal             | Medial               | Distal               |
|---------------------|-----------------------|----------------------|----------------------|----------------------|
| Free morpheme       | <i>đâu</i><br>‘where’ | <i>đây</i><br>‘this’ | <i>đấy</i><br>‘that’ |                      |
| Free/Bound morpheme |                       |                      | <i>đó</i><br>‘that’  | <i>kia</i><br>‘that’ |
| Bound morpheme      | <i>nào</i><br>‘which’ | <i>này</i><br>‘this’ | <i>ấy</i><br>‘that’  | <i>nọ</i><br>‘that’  |

### [3] Genitive noun phrases

The genitive marker *của* functions as a post-nominal modifier and is historically derived from a noun meaning ‘property’. In contrast to other syntactic contexts, the use of classifiers in genitive noun phrases is generally optional.

- (19) a. (Cái) váy của em là cái nào?  
CLF skirt GEN 2SG COP CLF which  
‘Which is your skirt?’
- b. Cái này là váy của em.  
CLF this COP skirt GEN 1SG  
‘This is my skirt.’
- c. (Cái) của chị đâu?  
CLF GEN 2SG where  
‘Where is yours?’

### [4] Adjectives

Adjectives of native Vietnamese origin typically appear as post-nominal modifiers, whereas those of Sino-Vietnamese origin frequently occur in pre-nominal position, as in *nam sinh viên* ‘male student’.

- (20) a. Tôi muốn mua con chó bé.  
1SG want buy CLF dog small  
‘I want to buy a small dog.’
- Anh muốn con chó nào?  
2SG want CLF dog which  
‘Which dog do you want?’
- b. Tôi muốn con bé.  
1SG want CLF small  
‘I want small one.’

## [5] Verb-based nominalizations

### [5a] Relative clauses

Vietnamese forms relative clauses using two primary strategies: the gapping strategy, as illustrated in examples (21b–c) and (22a–b), and pronominal retention, as shown in (23). These constructions may appear with or without the relativizer *mà*.

- (21) a. Anh có thể mang cho tôi cái váy mới không?  
 2SG can bring for 1SG CLF skirt new NEG  
 ‘Can you bring me the new skirt?’
- b. Cái đang treo trong tủ áo?  
 CLF PROG hang in closet SFP  
 ‘The one hanging in the closet?’
- c. Cái váy đang treo trong tủ áo.  
 CLF skirt PROG hang in closet SFP  
 ‘The skirt hanging in the closet.’
- (22) a. Cho tôi xem quyển (mà) anh mua hôm qua.  
 let 1SG see CLF LNK 2SG buy yesterday  
 ‘Let me see the one (book-like thing) you bought yesterday.’
- b. Quyển tiểu thuyết (mà) tôi mua hôm qua áo?  
 CLF novel LNK 1SG buy yesterday SFP  
 ‘The novel I bought yesterday?’

In example (23), the head noun phrase *cái đứa* ‘the guy’ is retained within the relative clause as the pronoun *nó* ‘(s)he/3rd person singular’, illustrating the pronominal retention strategy.

- (23) Cái đứa mà tôi cao hơn nó là em trai của tôi.  
 cái CLF LNK 1SG high more 3SG COP younger brother GEN 1SG  
 ‘The guy whom I am taller than is my brother.’ (Natchanan & Amara 2009)

### [5b] Complement clauses

Complement clauses in Vietnamese may be formed with or without an overt complementizer. As mentioned above, common complementizers include *là*, which historically derives from the copula, and *rằng*, which originally meant ‘to say’ in earlier stages of the language.

- (24) a. Nó gửi một bài thơ là/rằng Gió mưa là bệnh của trời...  
 3SG send one CLF poem COMP wind rain COP sick GEN sky  
 ‘She sent a poem saying Wind and rain is the sick of the sky...’
- b. Anh ấy gửi lại một bài là/rằng Anh về anh ốm tương tư...  
 3SG send back one CLF COMP 1SG return 1SG sick lovesick  
 ‘He sent back one saying I went home and got lovesick...’

### 3. DATA

#### 3.1 Classifiers

This section presents empirical data illustrating the distribution of classifier marking and grammatical nominalization in Vietnamese across a range of syntactic environments. The focus is on the distinction between NP-use and modification-use in contexts involving numerals, demonstratives, genitive phrases, adjectives, and verb-based nominalizations.

#### [1] Numerals

##### [1a] Cardinal numerals

NP use

- (25) Tôi muốn [bốn con].  
 1SG want 4 CLF  
 ‘I want four.’

Mod use

- (26) Tôi muốn [bốn con] chó.  
 1SG want 4 CLF dog  
 ‘I want four dogs.’

##### [1b] Ordinal numerals

NP use

- (27) Tôi thích [bài thứ hai].  
 1SG like CLF second  
 ‘I like the second one.’

Mod use

- (28) Tôi thích bài thơ [(\*)bài] thứ hai.  
 1SG like CLF poem CLF second  
 ‘I like the second poem.’

## [2] Demonstratives

NP use

- (29) [Con này] to.  
CLF this big  
'This one is big.'

Mod use

- (30) Con chó [(con) này] to.  
CLF dog CLF this big.  
'This dog is big.'

## [3] Genitive noun phrases

NP use

- (31) [(Cái) của chị] đâu?  
CLF GEN 2SG where  
'Where is yours?'

Mod use

- (32) (Cái) váy [(cái) của chị] đâu?  
CLF skirt CLF GEN 2SG where  
'Where is your skirt?'

## [4] Adjectives

NP use

- (33) Tôi muốn [con bé].  
1SG want CLF small  
'I want small one.'

Mod use

- (34) Tôi muốn con chó [(con) bé].  
1SG want CLF dog CLF small  
'I want a small dog.'

## [5] Verb-based nominalizations

### [5a] Relative clauses

NP use

- (35) [Cái           đang       treo       trong       tủ]       á?  
          CLF       PROG     hang     in       closet     SFP  
          ‘The one hanging in the closet?’

Mod use

- (36) Cái       váy       [(\*cái)     đang       treo trong     tủ]       á?  
          CLF     skirt       CLF     PROG     hang in       closet     SFP  
          ‘The skirt hanging in the closet?’

### [5b] Complement clauses

NP use

- (37) Nó     gửi     một     bài     là/rằng     Gió     mưa     là     bệnh     của     trời...  
        3SG   send   one     CLF   COMP     wind   rain   COP   sick   GEN   sky  
        ‘He sent one (poem) saying Wind and rain is the sick of the sky...’

Mod use

- (38) Nó     gửi     một     bài     thơ       (\*bài)     là/rằng     Gió     mưa     là     bệnh     của  
        3SG   send   one     CLF   poem     CLF       COMP     wind   rain   COP   sick   GEN  
        trời     ...  
        sky  
        ‘He sent a poem saying Wind and rain is the sick of the sky...’

The patterns observed in the data are summarized in Figure 1, which outlines the distribution of classifier marking across NP-use and modification-use contexts. The figure captures both obligatory and restricted marking patterns, highlighting where classifier use is permitted, optional, or blocked.

Figure 1. Classifier marking pattern in Vietnamese

NP use:     CARD.NUM >   DEM >     <sup>OPT</sup>GEN >   ADJ >     V-based NMLZ  
               ORD.NUM

Mod use:   CARD.NUM >   \*DEM >     \*GEN >     \*ADJ >     \*V-based NMLZ  
               \*ORD.NUM

(\* = no marking or impossible; OPT = optional marking; \*OPT = marking applies to a specific referent; unmarked = obligatory marking)

### 3.2 Genitive marker *của*

In contrast to classifiers, the genitive marker *của* plays a more limited role in grammatical nominalization. It functions as a nominalizer primarily in genitive noun phrases and relative clauses, and, to a lesser extent, as a marker of NP use in adjectival constructions, though the latter is largely confined to literary contexts. The nominalizing function of *của* is also semantically constrained. It typically requires a clear possessive or affiliative relationship between the head noun and the modifying element. In cases of verb-based nominalization, the head noun is generally interpreted as being controlled by or created by the subject expressed in the relative clause.

#### [3] Genitive noun phrases

NP use

- (39) Cái váy này là [của em].  
 CLF skirt this COP GEN 1SG  
 ‘This skirt is mine.’

Mod use

- (40) Cái này là [váy [của em]].  
 CLF this COP skirt GEN 1SG  
 ‘This is my skirt.’

#### [4] Adjectives

NP use

- (41) Của rẻ là của ôi.  
 GEN cheap COP GEN spoiled  
 ‘Cheap things are worthless.’

#### [5] Verb-based nominalizations

NP use

- (42) [Của tôi đào tạo] là thằng Nam, còn ...  
 GEN 1SG train COP CLF Nam, and  
 ‘The one I trained is Nam, and ...’

Mod use

- (43) [Thằng [của tôi đào tạo]] là thằng Nam, còn ...  
 CLF GEN 1SG train COP CLF Nam and  
 ‘The one I trained is Nam, and ...’



The distributional patterns of *của* as a nominalizer across different syntactic contexts are summarized in Figure 2.

Figure 2. *Của* marking pattern in Vietnamese

|          |        |        |       |            |              |
|----------|--------|--------|-------|------------|--------------|
| NP use:  | *NUM > | *DEM > | GEN > | *OPT ADJ > | V-based NMLZ |
| Mod use: | *NUM > | *DEM > | GEN > | *ADJ >     | V-based NMLZ |

#### 4. ANALYSIS AND DISCUSSION

The Vietnamese data on classifier marking align closely with the predictions of Shibatani’s (2023) functional-typological hierarchy, which posits that classifier/gender marking typically originates in NP-use and only later extends to modification-use. The pattern observed in Vietnamese largely supports this claim: classifier marking is consistently present in NP-use across structural categories (e.g., numerals, demonstratives, genitives, adjectives, and verb-based nominalizations), but its distribution in modification-use is strictly limited to cardinal numerals. A notable exception involves genitive noun phrases, where classifier marking is optional in NP-use. This deviates slightly from the hierarchy, which predicts either obligatory or no marking. This optionality may reflect a transitional stage or context-specific pragmatic influences within the Vietnamese system.

By contrast, the behavior of the genitive marker *của* provides an interesting complementary perspective. When functioning as a nominalizer, *của* appears in both NP-use and modification-use for genitive constructions and V-based nominalizations, mirroring the function of Mandarin Chinese *de* (Shibatani 2018: 390–391). This supports Shibatani’s broader claim that certain genitive or linker elements can evolve into non-classifying nominalizers. However, the use of *của* as a nominalizer is semantically constrained, typically requiring a possessive, affiliative, or control-based relationship between the head and modifier. Its productivity is also contextually limited, particularly in adjectival and numeral contexts, where its nominalizing function is not available.

Overall, the Vietnamese data contribute valuable cross-linguistic evidence in support of Shibatani’s theory, while also demonstrating some language-specifics and constraints that shape the distribution of nominalization strategies. The contrast between classifier-based and *của*-based nominalization in Vietnamese exemplifies the diversity of nominalization mechanisms and offers insight into the functional expansion and grammaticalization processes that operate across languages.

#### 5. CONCLUSION

The Vietnamese classifier marking pattern, when viewed through the lens of Shibatani’s (2023) functional-typological framework, provides important support for the proposed hierarchy of nominalization, particularly the precedence of NP-use over modification-use in the development of classifier systems. However, a comparison with typologically similar languages, such as Northern Thai, reveals key syntactic divergences, most notably in word order. Whereas Northern Thai typically exhibits the [NUM–CLF–N] order, Vietnamese shows

a strong preference for the [NUM–CLF–N] structure in modern usage but may have historically followed a [N–NUM–CLF] pattern. Alves (2023) suggests that this shift reflects a broader syntactic change from early Vietic to modern Vietnamese. Evidence from 15th-century poetic texts, such as the following line from ‘Ba Tiêu’ in *Quốc Âm Thi Tập*, illustrates this earlier pattern:

- (44) tình thư    một            bức            phong            còn            kín  
 love letter   one            CLF            seal            still            tight  
 ‘A love letter still sealed tight.’ (From the poem “Ba Tiêu” in “Quốc Âm Thi Tập,” Alves 2023: 11 of 23)

Although such examples are primarily found in poetic contexts, they offer diachronic insight into how classifier constructions in Vietnamese may have evolved structurally.

In parallel, the development of the genitive marker *của* also warrants historical investigation. While *của* functions today as a semantically restricted nominalizer in certain syntactic environments (e.g., genitive phrases and V-based nominalizations), its origin as a lexical noun meaning ‘property’ suggests a grammaticalization pathway worth tracing. Examining early texts may shed light on the semantic shifts underlying *của*’s evolution from denoting ownership or concrete possessions (e.g., ‘things’, ‘clothes’, ‘food’) to serving as a grammatical marker of nominal relations.

Taken together, the Vietnamese data appear broadly consistent with key aspects of Shibatani’s typological model and suggest the potential value of historical and comparative perspectives in exploring how nominalization systems develop across languages. Further diachronic and cross-linguistic investigation, particularly into classifier word order and the evolution of nominalizing markers, may offer deeper insights into the interaction between form, function, and structure in nominal syntax.

## ABBREVIATIONS

|          |                   |
|----------|-------------------|
| 1        | first person      |
| 2        | second person     |
| 3        | third person      |
| SG       | singular          |
| ADJ      | adjective         |
| CARD.NUM | cardinal numeral  |
| CLF      | classifier        |
| COMP     | complementizer    |
| COMP.CL  | complement clause |
| DEM      | demonstrative     |
| GEN      | genitive marker   |
| NMLZ     | nominalization    |
| NMLZR    | nominalizer       |
| NUM      | numeral           |
| ORD.NUM  | ordinal numeral   |

## ACKNOWLEDGMENTS

The present study has been supported by grants from the Japan Society for the Promotion of Science (KAKENHI Project IDs: 22H00659 and 23K21931, ‘Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers’, PI: Kazuhiro Kawachi) and from the National Institute for Japanese Language and Linguistics (‘Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective’, PI: Prashant Pardeshi). For academic purpose, Shimizu Masaaki is responsible for Sections 2, 3.2, and 5, while Trang Phan is responsible for Sections 1, 3.1, and 4.

## REFERENCES

- Alves, Mark. 2023. An updated overview of the Austroasiatic components of Vietnamese. *Languages* 9 (12), 377.  
<https://doi.org/10.3390/languages9120377>
- Cao, Xuân Hạo. 1992. Về cấu trúc của danh ngữ trong tiếng Việt (On the structure of noun phrases in Vietnamese). In Viện Khoa học Xã hội tại Thành phố Hồ Chí Minh. (eds.) *Tiếng Việt và các ngôn ngữ phía Nam (Vietnamese and the languages in the South)*, 7–31. Hà Nội: Nhà xuất bản Khoa học Xã hội.
- Emeneau, Murray B. 1951. *Studies in Vietnamese (Annamese) grammar*. Berkeley, CA: University of California Publications in Linguistics 8.
- Natchanan, Yaowapat, and Amara Prasithrathsint. 2009. A typology relative clauses in mainland Southeast Asian languages. *Mon-Khmer Studies* 38, 1–23.  
<http://sealang.net/sala/archives/pdf8/natchanan2008typology.pdf>
- Nguyễn, Tài Cẩn. 1975. *Từ loại Danh từ trong tiếng Việt hiện đại (Nouns in modern Vietnamese)*. Hanoi: Nhà xuất bản Khoa học Xã hội.
- Nguyen, Dinh Hoa. 1957. Classifiers in Vietnamese. *Word* 13, 124–52.
- Nguyen, Dinh Hoa. 1971. *Vietnamese-English student dictionary*. London and Amsterdam: Southern Illinois University Press.
- Nguyen, Tuong. 2004. *The structure of the Vietnamese noun phrase*. Boston: Boston University dissertation.
- Phan, Trang. 2019. The absence of classifiers in numeral classifier constructions in Vietnamese. In WebFestschrift, Metin Bağrıaçık, Anne Breitbarth and Karen de Clercq. (eds.) *Mapping linguistic data. Essays in honour of Liliane Haegeman*, 208–213. Ghent: Ghent University.
- Phan, Trang, and Gennaro Chierchia. 2022. Identifying (in)definiteness in Vietnamese. *Journal of the Southeast Asian Linguistics Society* 15 (2), 27–49.  
<http://hdl.handle.net/10524/52502>
- Phan, Trang, and Gennaro Chierchia. in press. Putting three pieces of a puzzle into place: Classifiers- plurals-articles in Vietnamese. In Andrew Simpson. (ed.) *New insights into theoretical syntax from Asian languages*, Amsterdam and Philadelphia: John Benjamins.
- Phan, Trang, and Duffield Nigel. 2022. A Roadmap to Vietnamese Phrase Structure. In Chris Shei and Saihong Li. (eds.) *The Routledge handbook of Asian linguistics*, 165–185. London: Taylor & Francis.

- Phan, Trang, and Lam Quang Dong. 2021. Decomposing Definiteness in Vietnamese. *Journal of the Southeast Asian Linguistics Society* 14 (1), 1–18.  
<http://hdl.handle.net/10524/52476>
- Shibatani, Masayoshi. 2018. Nominalization in crosslinguistic perspective. In Prashant Pardeshi and Taro Kageyama. (eds.) *Handbook of Japanese contrastive linguistics*, 345–410. Berlin: Walter de Gruyter.
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)
- Simpson, Andrew, and Binh Ngo. 2018. Classifier syntax in Vietnamese. *Journal of East Asian Linguistics* 29, 211–246.  
<https://doi.org/10.1075/sl.35.1.10sim>

# Classifying nominalizers in Northern nDrapa

Satoko Shirai

## ABSTRACT

This study describes grammatical nominalization markers that classify notions evoked by nominalizations in Northern nDrapa. nDrapa features classifying grammatical nominalizers, also known as classifiers. It mainly examines the application of Shibatani's (2021) implicational hierarchy to classifiers in nDrapa.

Classifiers must follow basic numerals when they express quantification. However, in noun modification, loan numerals representing large numbers are frequently used without classifiers. Classifiers may also directly follow NPs, but not demonstratives or genitive phrases. Verbs are obligatorily marked by verb-based nominalizers when used as NPs or in adnominal clauses. V-based nominalizations may be followed by *ɾA* 'GEN' but not by classifiers.

A potential exception to Shibatani's hypothesis is that nDrapa classifiers can be attached to adjectives. However, we analyze this phenomenon as a functional extension of bare classifiers (i.e., those used without numerals) to indicate the discourse property of NPs. Classifiers can be attached to reduplicated adjectives (which can be used as adnominal modifiers or NPs by themselves) but not to adjective stems. By examining the discourse functions of bare classifiers in contrast to the NP-use marker *ɾA*, we conclude that they function as discourse markers: classifiers mark indefinite NPs, while *ɾA* marks definite NPs.

In summary, nDrapa classifiers only nominalize numerals. Moreover, loan numerals may modify nouns without classifiers, whereas those used as NPs in the field data have a classifier. These facts support Shibatani's hypothesis.

## 1. INTRODUCTION

The study examines the application of Shibatani's (2021, 2023) implicational hierarchy to classifying grammatical nominalizers in the Northern dialect of the nDrapa language. nDrapa (ISO 639-3 zhb) is a Qiangic language of the Sino-Tibetan language family and is spoken in Sichuan Province, China. Northern nDrapa features two types of classifying grammatical nominalizers: numeral-based and verb (V)-based nominalizers. The study focuses on numeral-based nominalizers, also known as numeral classifiers. In nDrapa, numeral classifiers are enclitics that are obligatorily used when quantifying numerals, which function as noun phrases (NPs) or adnominal modifiers. Moreover, numeral classifiers can be attached to NPs. Conversely, classifying V-based nominalizers are verbal suffixes that form adnominal clauses and NPs from verb phrases or clauses. In addition, Northern nDrapa uses the non-classifying N-based nominalizer or the genitive marker, *ɾA* 'GEN'.

## 1.1 Theoretical background

This chapter examines whether nDrapa classifiers are used to conform to Shibatani's (2021) hypothesis. Shibatani (2021: 498) argued that the primary functions of classifiers are to nominalize numerals and classify an entity denoted by numeral-based nominalizations instead of classifying the head nouns they modify. Moreover, Shibatani (2021) proposed two hypotheses on the following hierarchy about the uses and syntactic functions of classifying nominalization markers, including classifiers: [1] they primarily have NP use more than modification use and [2] their syntactic functions develop according to the hierarchy of their base: NUM > DEM > GEN > ADJ > V-based NMLZ. They initially nominalize numerals, which, in turn, develop to nominalize demonstratives, genitive phrases, adjectives, and V-based nominalizations.

Shibatani (2019) also defined and discussed N-based nominalization. For example, in nDrapa, the genitive phrase *ŋa=rá* {1SG=GEN} 'my/mine' denotes an entity evoked regarding *ŋa* '1SG' after contexts such as 'my cup' or 'my family.' This function parallels V-based nominalization. An example is *ki-ttsí-ma(=ra)* {INW-eat-NMLZ(=GEN)}, a V-based nominalization that consists of *ki-ttsi* 'to have eaten' and a nominalizer (similar to Japanese *tabe-ta(=no)* {eat-PST(=FORMAL.NOUN)}), which denotes something evoked in relation to the completed act of eating, such as an apple (food) or a bowl (instrument). Therefore, both are termed nominalizations.

## 1.2 Basic information of nDrapa

The nDrapa language belongs to the Qiangic group of the Sino-Tibetan family. It is spoken along the Xianshuihe River in the Western Sichuan region of China. No written tradition exists. The language is spoken by an estimated number of 10,000 speakers in three dialect groups, namely, Southern, Central, and Northern (Huang 2024). This study examines the northernmost variety of the Northern dialect group, which is spoken in the Mätro (Mazhong/麻中) village of Zhongni (仲尼) township. However, the study also refers to elicited examples from the Tratho (Zhatuo/扎拖) variety of the Northern dialect group with the notation "Tratho nDrapa." Unless noted otherwise, all examples in this chapter were collected from the author's fieldwork. Notes enclosed in < > after Mätro examples indicate a source category, whether from elicitation or folktale narratives (title in italics).

Phonemes in Mätro nDrapa include the following: consonants /ph [p<sup>h</sup>], th [t<sup>h</sup>], tʰ [t<sup>h</sup>], ch [ç<sup>h</sup>], kh [k<sup>h</sup>]; p, t, t̚, c, k; b, d, d̚, ʝ, g; tsh [ts<sup>h</sup>], tɕh, [tɕ<sup>h</sup>]; ts, tɕ; dz, dz̚; m, n, ŋ, ɲ; m̥ [m̥m], ɲ̥ [ɲ̥n], ɲ̥̥ [ɲ̥̥n], ɲ̥̥̥ [ɲ̥̥̥n]; fh [f<sup>h</sup>], sh [s<sup>h</sup>], ɕh [ɕ<sup>h</sup>]; f, s, ɕ, x, h; v, z, ʒ, ʃ, fi; w, j; l, r [ɾ~z]; ʎ [ʎ̥], ʎ̥̥ [ʎ̥̥~ʂ]; vowels /i, i̥, u, u̥, e [ɪ], ɐ, o, ɛ, ʌ, a; ei, ʌu/. In addition, Mätro nDrapa has word tones in which the first two syllables of a phonological word have one of the following pitch patterns: (i) high-level (ó/óó), (ii) high-falling (ǒ/óó [ǒǒ]), (iii) low-rising (ǒ/óó [ǒó]), and (iv) low-rising-falling (ǒó [ǒǒ]) (the low-rising-falling tone is not attested in monosyllabic phonological words). The third and further syllables are unstressed and have a low pitch. Low-level pitch is left unmarked in phonemic transcription. However, tones may vary according to post-lexical prosody, which reflects specific contexts (Shirai 2019).

nDrapa native morphemes tend to be monosyllabic, and their morphology is mainly agglutinative and uses prefixes and suffixes. The basic case system is nominative-accusative.

The nominative presents no overt marker; other case markers may be omitted if grammatical relations are clear based on the context (Shirai 2010).

The basic constituent order is subject–object–verb. Demonstratives, nouns, and nominalizations can modify nouns. In this context, nominalizations include nominal (N)-based (genitive phrases), V-based (adnominal clauses), and adjective (A)-based (adjectives in reduplicated form) nominalizations, and quantifiers, including classifier phrases. Moreover, particles may follow the NP; for example, the genitive particle *ra* functions as an NP-use marker (Shirai 2020). Noun modifiers may precede or follow a head noun depending on the type (Figure 1).

Figure 1. Basic order of noun modification.

| <i>Pre-head modifiers</i>                   |                                | <b>Head noun</b> | <i>Post-head modifiers</i> |            | <i>Post-NP</i> |
|---------------------------------------------|--------------------------------|------------------|----------------------------|------------|----------------|
| Demonstrative/<br>N-based<br>nominalization | Noun/V-based<br>nominalization |                  | A-based<br>nominalization  | Quantifier | Particle       |

## 2. OVERVIEW OF GRAMMATICAL NOMINALIZATIONS IN NDRAPA

### 2.1 Grammatical nominalizations of verbal, adjectival, and nominal bases

Shirai (2020, 2022) described the basic facts of nDrapa grammatical nominalizations from the perspective of nominalization theory. It concluded that adnominal modifiers are nominals or nominalizations, because every type of adnominal modifier can also be used as an NP (Shirai 2020). This section provides overviews of the V-based, A-based, and N-based nominalizations. Section 3 provides data on numeral-based nominalizations.

#### 2.1.1 V-based nominalizers

nDrapa uses the following nominalizing suffixes to nominalize verbal bases: verbs, VPs, or verbal clauses (Shirai 2020). Their distinction is primarily based on syntax, although they partially demonstrate classification-like functions; for example, the use of *-pi* is nearly limited to nominalizations that evoke humans.

- ma*: the default nominalizer, which makes nominalizations that evoke various objects except for those which *-pi* or *-hti* below, is typically used.
- pi*: typically makes nominalizations that evoke the human subject (animals can also be accepted in certain cases).
- hti*: typically makes nominalizations that evoke place (recipients can also be accepted).
- ze*: only makes event nominalizations; no examples of argument nominalization.

V-based nominalizations are used as NPs or noun modifiers, which precede the head noun. The subsequent examples illustrate NP-use ((1)–(4)) and modification use ((5)–(7)). The event-specific nominalizer *-ze* does not have modification use. Section 2.2 discusses the non-typical function of *ra* ‘GEN’ in (1), (2), and (5), which appear after NPs.

- (1) [[hkeʒó ɲá-ndé]-mΛ]=rΛ shá =sa=ré.  
PSN OUT-beat-NMLZ=GEN who =ADM=FAC  
'Whom did Kezo beat?' <elicitation>
- (2) ɲóró [[[tehî ɲi]-pí](=rΛ)] ndú mé=ʈó a-təhó-a.  
that condition be.ill-NMLZ=GEN PLN medical=place DWN-go-FAC.PFV  
'That sick person went to the hospital in Ndu.' <elicitation>
- (3) [[zipáŋɣΛ tó-təú-a]-hti] zΛntəhí vó=æu, ...  
son NTL-exist-PFT-NMLZ daughter come=CONC  
'The daughter came to the place where the son had been, but (they couldn't meet).' <*Bowl and Bracelet*>
- (4) [[somúɲi xǒ nchará jí]-zé] nΛ=sá=rɛ.  
next.day more fun/picnic go/come-NMLZ exist=ADM=FAC  
'It happened that they went for a picnic again the next day.' <*A Frog*>
- (5) ɲoró=la=ne [[[mě tΛ-khé]-mΛ]=rΛ nteholó]=ji po=ʈé.  
3SG=LOC=TOP mother NTL-give-NMLZ=GEN bowl=CLF exist=IPFV.FAC  
'He has a bowl given by his mother.' <*Bowl and Bracelet*>
- (6) [[[sheppú ɲgú=wu təú]-pí] shwí]=rΛ záhʈí a-ʈhá=ne...  
tree top=DAT exist-NMLZ man=GEN shoestring DWN-untie=then  
'The man on the top of the tree untied his shoestrings, then ...' <*A Shape-shifter*>
- (7) [[[ɲóró a-jí]-hti] satsá] ndú rě.  
3SG DWN-go/come-NMLZ place PLN COP<sub>4</sub>  
'The place he went was Ndu.' <elicitation>

### 2.1.2 A-based nominalization

Adjectives are nominalized through the reduplication of the last syllable. A-based nominalizations follow the head NP when used as a modifier. The following texts are examples of NP use (8) and modification use (9):

- (8) ɲjé=rá ɲi~ɲí-ɲgo=rΛ nóné kí-ttsí hce-á mǒ.  
1PL=GEN delicious~NMLZ-DIM=GEN 2DU INW-eat PST-FAC.PFV CFM  
'You two ate our delicious ones, right?' <*Two Goblins*>
- (9) htěi ʈhé~ʈhé pálá~lΛ=ji kΛ-té=rΛ,  
cloud white~NMLZ round~NMLZ=CLF INW-come=COND  
lΛʈhé hsó ó-mo.  
white.incense burn UPW-make.IMPR  
'If a white and round cloud comes, burn white incense.' <*Two Goblins*>



In (8), the nominalized adjective  $\eta i \sim \eta i$  ‘delicious one’ is modified by the genitive phrase  $\eta j \acute{e} = r \acute{a}$  ‘our’ and diminutive  $- \eta go$  ‘DIM’ in the quantifier position. In (9), two coordinated nominalized adjectives, namely,  $\acute{t} h \acute{e} \sim \acute{t} h \acute{e}$  ‘white one’ and  $p \acute{a} l \acute{a} \sim l \acute{a}$  ‘round one’, modify the head noun  $h t \acute{e} i$  ‘cloud’.

The abovementioned descriptions are positive degree adjectives. Comparative and superlative expressions are not reduplicated but use the same strategy as verbs. For example, in (10), the superlative adjective phrase  $h t \acute{i} z \acute{i}$  ‘most miserable’ is nominalized using a V-based nominalizer  $- pi$ .

- (10)  $z \Delta n t \acute{e} h \acute{i}$      $[[h t \acute{i} \quad z u] - p i] = r \Delta = k \Delta$                        $m \acute{o} = h c \acute{e} - a = r \acute{e}$ .  
 daughter    most    miserable-NMLZ=GEN=inside    live=PST-PFV=FAC  
 ‘The daughter stayed among the ones (servants) of the lowest rank.’ <*Bowl and Bracelet*>

### 2.1.3 N-based nominalization

The so-called genitive particle nominalizes nominals. N-based nominalizations in NP use require the genitive particle, while those in modification use may stand without it. For example, in (11a),  $\eta a = r \acute{a}$  {1SG=GEN} ‘mine’ denotes an object that belongs to the speaker, which denotes cups based on the given context. The genitive particle cannot be omitted in this case, as illustrated in (11b). In (12),  $h t \acute{e} a l \acute{a}$  (a household name) has the genitive particle and modifies the following NP,  $\eta \acute{a} \eta k h a z \acute{i}$  ‘gold spoon’. Conversely,  $\eta \acute{a}$  ‘gold’ in the same example modifies the head noun  $\eta k h a z \acute{i}$  ‘spoon’ without the genitive particle.

- (11) a.     $\eta \acute{o} r \acute{o}$      $\acute{t} \acute{o} n \acute{d} \acute{o}$      $sei = t \acute{e} u = r \Delta$                        $\eta a = r \acute{a}$                        $w \acute{a}$ .  
           that    cup    three=CLF=GEN    1SG=GEN                      COP<sub>1</sub>  
           ‘Those three cups are mine.’

- b.     $* \eta \acute{o} r \acute{o}$      $\acute{t} \acute{o} n \acute{d} \acute{o}$      $sei = t \acute{e} u = r \Delta$                        $\eta \check{a}$                        $w \acute{a}$ .  
           that    cup    three=CLF=GEN    1SG                      COP<sub>1</sub>  
           (Lit. ‘those three cups are me.’)

- (12)  $ther \acute{o} - n \acute{e}$      $h t \acute{e} a l \acute{a} = r \Delta$      $\eta \acute{a}$      $\eta k h a z \acute{i} = t \acute{e} u = r \Delta$                        $\eta \acute{a} - z \acute{e}$                        $r \acute{e} - a$ .  
 goblin-DU    PSN=GEN    gold    spoon=CLF=GEN    steal-NMLZ    get-FAC.PFV  
 ‘The two goblins got to steal Chala family’s gold spoon.’ <*Two Goblins*>

### 2.2 Functional extension of nominalizers: Bare classifier and NP-use markers

The genitive marker in Northern nDrapa also functions as NP-use markers, which supports the status of complex nominalizations as NPs. However, they are frequently considered optional in elicitation, as in (13).

- (13)  $[[w \acute{o} t \acute{s} h \acute{i} \quad s \acute{e} i = t \acute{e} u \quad p \acute{o}] - p i] (= r \Delta)$                        $\acute{l} \acute{a} m \acute{o}$      $r \acute{e}$ .  
           hat                      three=CLF    exist-NMLZ=GEN    PSN                      COP<sub>4</sub>  
           ‘The one who has three hats is Lamo.’ <elicitation>

Analyzing narratives, we conclude that definite, specific, and complex NPs tend to have *rA* ‘GEN’ that functions as NP-use markers. For example, the sentence in (14) first introduces *hdæci htsí~htsí* ‘hot Tibetan bread’ into the narrative, which does lack *rA* ‘GEN’. Alternatively, in a later context, the center part of the same bread is mentioned with *rA* ‘GEN’ (15).<sup>1</sup>

- (14) *hdæci htsí~htsí=ji kó-mé t̥he=sá.*  
 Tibetan.bread hot~NMLZ=CLF INW-make HABIT=ADM  
 ‘(The family) made a hot Tibetan bread (for their herder).’ <Two Goblins>
- (15) *hdæci khóntshwí=rA kí-ttsí hce-á.*  
 Tibetan.bread marrow=GEN INW-eat PST-FAC.PFV  
 ‘(The goblins) ate the center part of the Tibetan bread (that the family prepared for the herder).’ <Two Goblins>

Northern nDrapa classifiers may directly follow NPs, including A-based nominalizations. This is parallel to a phenomenon called “bare classifiers” in many East and Southeast Asian languages (e.g., Bisang 1999: 150–159; Pichetpan and Post 2021; Li 2023). Bare classifiers in Northern nDrapa are used when the NP is single, indefinite, and specific, which is common when introducing new and important characters in narratives.<sup>2</sup> For example, (16) and (17) are the beginning of a narrative. In (17), the default classifier *ji* directly follows the nominalized adjective *ndzí~ndzí* ‘a great one’ on the surface; structurally, however, it is attached to the NP *lemé ndzí~ndzí* ‘(a) great monk’. The position of a bare classifier in (9), which is after the entire NP *ht̥ei th̥é~th̥é p̥ál̥á~lA* ‘white round cloud’ instead of after each adjective, also supports this analysis.

- (16) *t̥e=ccu=la [miwó]=ji t̥á-n̥á-a=r̥e.*  
 one=CLF=LOC old.woman=CLF NTL-exist<sub>2</sub>-PFV=FAC  
 ‘There was an old woman in a certain household.’ <A Frog>
- (17) *nda=ní waláhte=la [lemé ndzí~ndzí]=ji tó-pó-a=r̥e.*  
 past.time=TOP PSN=LOC monk great~NMLZ=CLF NTL-exist-PFV=FAC  
 ‘Long ago, there was a great monk in Walate’s house.’ <Two Goblins>

After their introduction, bare nouns are used to refer to the characters already introduced in the story. Example (18) is found in a context later than (17). The bare noun *lemé* ‘(the) monk’

<sup>1</sup> In most cases, the default V-based nominalizer *-mA* connects to *rA* ‘GEN’, as in examples (1) and (5), which may be becoming lexicalized. In Southern nDrapa, the particle *mbəra*, which originates from this combination, is synchronically inseparable.

<sup>2</sup> In the Sasho dialect of Southern nDrapa, parallel expressions use a particle *tu*, which is derived from the combination of the numeral *t̥e* ‘one’ and the default classifier *ji* (Huang 2024). This suggests that bare classifiers in Northern nDrapa also originate from the numeral ‘one’ plus a classifier, which dropped the numeral after the dialects split.

in (18) denotes the same character as *lemé ndzɪ~ndzɪ=ji* ‘a great monk’ in (17). Besides, indefinite and non-specific nouns also use the bare form, such as *paŋɟá* ‘child’ in (19).

- (18) **lemé**      kʌ-móki      hce-á=rɛ.  
monk      INW-ask      PST-PFV=FAC  
‘The monk asked.’ <Two Goblins>

- (19) jɛʈó      ɕohpá      hpèi      tɛí~tɛí=la      **paŋɟá**      mɐ-pó=sa.  
PLN      village      local.lord      big~NMLZ=LOC      child      NEG-exist<sub>1</sub>=ADM  
‘The great load of the Yeto village did not have a child.’ <Bowl and Bracelet>

A bare classifier and the NP-use marker *rʌ* ‘GEN’ may follow an NP together in this order. In these cases, the bare classifier indicates that the NP is single, while the NP-use marker indicates that the NP is definite. For example, *rérá kóló=pe=rʌ* ‘the one pot of tea’ in (21) refers to an identical pot of tea mentioned in the former context in (20).

- (20) rérá      kóló=pe      a-xú      lɛ=ʈhé=sa.  
tea      pot=CLF      DWN-insert      put=HABIT=ADM  
‘(The family) prepared one pot of tea (for their herder).’ <Two Goblins>

- (21) rérá      kóló=pe=rʌ      kʌ-théi      hce-á.  
tea      pot=CLF=GEN      INW-drink      PST-FAC.PFV  
‘(The goblins) drank the pot of tea (that the family prepared for the herder).’ <Two Goblins>

### 3. DATA ON CLASSIFYING GRAMMATICAL NOMINALIZATION

Northern nDrapa has relatively rich classifiers that function as classifying grammatical nominalizers. Shirai (2022) reported eight individual classifiers, seven collective classifiers, and many mensural classifiers, which are obligatorily used with basic numerals. The current study refers to a construction that involves a numeral and a classifier as a classifier phrase. Examples (22)–(24) are elicited from the Tratho variety of Northern nDrapa based on Pardeshi’s (2021) questionnaire.<sup>3</sup> Example (22) indicates that numerals require a classifier, while example (23) evidently indicates that a classifier follows reduplicated adjectives. Nevertheless, as discussed in Section 2.2, bare classifiers do not nominalize adjectives but are attached to NPs given discourse conditions. Other candidates, including demonstratives and genitive phrases, may not have a classifier. For example, classifiers cannot be directly attached to demonstratives (24a). If one needs to use a classifier with a demonstrative, then a numeral must be inserted between them, as in (24b).

<sup>3</sup> Fieldwork on the Mätro variety has not been possible since 2015 due to the political situation. Therefore, here we present data for the Tratho variety, which has almost the same grammatical structure as the Mätro variety.

(22) Tratho nDrapa

- a. nǒ khǎ tɛhú=tɛu ɛu-á?  
 2SG dog how.namy=CLF need-Q  
 ‘How many dogs do you want?’
- b. (ŋá) [dǎ=tɛu] (ɛǎ).  
 1SG four=CLF need  
 ‘(I want) four. CLF.’
- c. khǎ [dǎ=tɛu]  
 dog four=CLF  
 ‘four dogs’

(23) Tratho nDrapa

- a. nǒ khǎ tɛhí+ŋkhéi=tɛu ɛu-á?  
 2SG dog what+like=CLF need-Q  
 ‘How many dogs do you want?’
- b. ŋá [lézi~zi]=tɛu ɛǎ.  
 1SG little~NMLZ=CLF need  
 ‘I want a little. CLF (one).’
- c. [khǎ [lézi~zi]] (=tɛu)  
 dog little~NMLZ=CLF  
 ‘(a) little dog’

(24) Tratho nDrapa

- a. ŋá kǎrǎ ɛǎ.  
 1SG this need  
 ‘I want this.’
- b. ŋá kǎrǎ tɛ=i=rǎ ɛǎ.  
 1SG this one=CLF=GEN need  
 ‘I want this one.’

Classifier phrases in Northern nDrapa may be used as NPs without the explicit appearance of a noun in the context. For example, in example (16), the classifier phrase *té=ccu*, which consists of the numeral ‘one’ and the classifier for households, refers to a particular household at the beginning of a narrative. A parallel phenomenon is observed in (25), *né=ccú* {two=CLF}. In addition, *ŋǎ=zjá* in (26), which consists of the numeral ‘five’ and the classifier for human, denotes five people without a head noun.

- (25) véŋá-nē      né=ccú      mē=hcj-á.  
 brother=DU   two=CLF      make=PST-FAC.PFV  
 ‘The two brothers were divided into two separate families. (Lit. made two households).’  
 <A Goblin in a Cave>

- (26) ndá      ŋjé      jé=ká      ŋé=zjá      tál-ná-a      =rě.  
 past.time   1PL   house=inside   five=CLF   NTL-exist-PFV   =FAC  
 ‘Formerly, there were five (family members) in our house.’ <elicitation>

Loan numerals that represent large numbers, such as *fidzi* ‘hundred’, are typically used without classifiers in noun modification. For example, in (27), the numeral *fidzi* ‘hundred’ modifies the head noun *thí* ‘dollar, (Chinese) yuan’ without a classifier. In this context, we expect the use of the default classifier *ji* if the numeral is basic (e.g., *ně* ‘two’ in (28)).

- (27) ŋa=rá      páopáo=ká      thí      fidzi      tēǎ.  
 1SG=GEN   bag=inside   dollar   hundred   exists<sub>3</sub>  
 ‘There are one hundred yuans (元) in my bag.’ <elicitation>

- (28) tajá      thí      ně=ji  
 money   yuan   two=CLF  
 ‘two yuans (元) of money’ <elicitation>

The field notes present a few examples in which large-number numerals form NPs. These cases involve a classifier (29). This fact implies that NP use is more primary than modification use for Northern nDrapa classifiers, but additional data are required to confirm.

- (29) somúŋi      sheifidzi=zja=htei      ŋá-jí      ɛu-é.  
 tomorrow   three.hundred=CLF=about   INW-go/come   need-FAC.IPFV  
 ‘Tomorrow, about 300 (servants) need to go.’ <Bowl and Bracelet>

#### 4. CONCLUSION

This study described grammatical nominalization markers in Northern nDrapa and focused on classifying nominalizers or numeral classifiers. It also examined the application of Shibatani’s (2021) implicational hierarchy to numeral classifiers in Northern nDrapa.

Classifiers in Northern nDrapa obligatorily follow the basic numerals if a phrase serves a quantification function in NP use and modification use. However, loan numerals that represent large numbers are typically used without classifiers in noun modification, while those used as NPs in the field data have a classifier.

Numerals may also directly follow NPs, including A-based nominalizations in NP use, but they cannot directly follow demonstratives and genitive phrases (bare classifiers). Verbs are obligatorily marked by V-based nominalizers when used as NPs or adnominal clauses. V-based nominalizations may be followed by *rá* ‘GEN’ but not by classifiers.

A potential exception to Shibatani's hypothesis is that nDrapa classifiers can be attached to adjectives when they function as NPs or adnominal modifiers. However, the study analyzed the phenomenon as a functional extension of classifiers to indicate the discourse property of NPs. Classifiers can be attached to reduplicated adjectives, which can be used as adnominal modifiers or NPs by themselves, but not as adjective stems. By examining the functions of bare classifiers that contrast *ra* 'GEN', the study infers that they function as discourse markers: the former marks indefinite NPs, while the latter marks definite NPs.

Thus, the study concludes that nDrapa classifiers only nominalize numerals. Moreover, by observing data on loan numerals that represent large numbers, NP use may be more primary than modification use. These facts support Shibatani's hypothesis.

## ABBREVIATIONS

|       |                    |
|-------|--------------------|
| 1     | first person       |
| 2     | second person      |
| 3     | third person       |
| ADM   | admirative         |
| CFM   | confirmative       |
| CLF   | classifier         |
| CONC  | concessive         |
| COND  | conditional        |
| COP   | copular            |
| DAT   | dative             |
| DIM   | diminutive         |
| DWN   | downward directive |
| DU    | dual               |
| FAC   | factual            |
| GEN   | genitive           |
| HABIT | past habitual      |
| IPFV  | imperfective       |
| INW   | inward directive   |
| LOC   | locative           |
| NEG   | negative           |
| NMLZ  | nominalizer        |
| NTL   | neutral directive  |
| OUT   | outward directive  |
| PFV   | perfective         |
| PL    | plural             |
| PLN   | proper place name  |
| PSN   | proper person name |
| PST   | past               |
| SG    | singular           |
| TOP   | topic              |
| UPW   | upward directive   |

## ACKNOWLEDGMENTS

The study forms part of the results of the National Institute for Japanese Language and Linguistics (“Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective,” PI: Prashant Pardeshi). In addition, it has received support from JSPS KAKENHI Grants (Project IDs: JP22H00659 and JP23K00476).

## REFERENCES

- Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages: Counting and beyond. In Jadranka Gvozdanović. (ed.) *Numeral types and changes worldwide*, 113–186. Berlin: de Gruyter.
- Huang, Yang. 2024. *Sichuan Yajiang Zhabahua* [The nDrapa language in Yajiang, Sichuan.] Beijing: The Commercial Press.
- Li, Yu. 2023. Referentiality and thematic importance: The discourse functions of Zauzou classifiers. *Lingua* 285, 103477.
- Pardeshi, Prashant. 2021. Data-collection template. ms.
- Pichetpan, Nitipong and Mark W. Post. 2021. Bare classifier phrases in Thai and other mainland Asian languages: implications for classifier theory and typology. *Linguistic Typology* 25(3), 461–506.
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shirai, Satoko. 2010. Dapago no “kaku” o arawasu keesiki [Case markers in nDrapa]. In Hideo Sawada. (ed.) *Tibetto=Birumakeegengo no bunpoogensyoo 1: Kaku to sono syuuhon [Grammatical phenomena in Tibeto-Burman languages 1: Case and related phenomena]*, 287–310. Fuchu: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Shirai, Satoko. 2020. Dapago no meishishuushokuhyoogen to meishikuhyooshiki (Noun modification and NP marker in nDrapa). In Prashant Pardeshi and Kaoru Horie. (eds.) *Nihongo to sekai no gengo no meishishuushokuhyoogen (Noun modification in Japanese and world’s languages)*, 97–117. Tokyo: Hituzi Syobo.
- Shirai, Satoko. 2022. Classifiers in nDrapa: Definition and categorization. *Gengo Kenkyu* 162, 25–46.

# Nominalization in Guoli Zauzou

Yu Li<sup>1</sup> and Tetsuya Miyagishi<sup>2</sup>

## ABSTRACT

This study examines grammatical nominalization in Guoli Zauzou (Tibeto-Burman), analyzing two distinct strategies: classifying nominalizers (sortal classifiers) and non-classifying nominalizers (the genitive enclitic =ze<sup>55</sup> and linking particles). In alignment with Shibatani's (2021, 2023) structural hierarchy, classifiers exhibit contiguous extension from numerals to verb-based nominalizations (e.g., adjectival phrases, relative clauses), yet they are systematically excluded from genitive noun phrases—where the enclitic =ze<sup>55</sup> is employed instead. This constitutes a key exception to the proposed hierarchy. Non-classifying nominalizers, including linking particles, are restricted to modification contexts and lack independent nominalizing functions in noun phrase uses. While classifiers dominate in verb-based noun phrase uses, linking particles prevail in verb-based modification uses, with their co-occurrence being dispreferred. The findings generally support Shibatani's hypothesis but uncover a non-contiguous nominalization pattern in genitives, challenging the universality of a strict hierarchical progression.

## 1. INTRODUCTION

This paper examines two types of grammatical nominalization in the Guoli dialect of Zauzou (ISO 639-3: zal), a Tibeto-Burman language spoken in Yunnan, China. The first type involves sortal classifiers, which form a semi-open class of classifying nominalizers and primarily govern the nominalization of cardinal numerals, demonstratives, adjectives, and relative clauses. The second type comprises non-classifying nominalizers, including the genitive enclitic =ze<sup>55</sup> and the linking particles xo<sup>53</sup> and tɕ<sup>31</sup>, which are employed in genitive phrases, adjectival phrases, and verb-based nominalizations, respectively.

The findings demonstrate that classifier-based nominalizations span all levels of Shibatani's (2021: 524, 2023: 35) hierarchy, from cardinal numerals to relative clauses. In contrast, nominalizations formed by non-classifying nominalizers are restricted to genitive noun phrases, adjectives, and relative clauses.

Consistent with the predominant modifier-modified order in Sino-Tibetan languages (LaPolla 2007: 46), Guoli Zauzou displays a rigid [MODIFIER-MODIFIED] order within noun phrases. Notably, in addition to nouns, sortal classifiers can function as nominal heads—modified by various adnominal modifiers when the head noun is implicit. The basic NP structure of Zauzou is illustrated in Figure 1. The noun (N) is optional in a noun phrase, whereas a classifier (CLF) is obligatory. Demonstratives and relative clauses consistently

---

<sup>1</sup> School of Chinese Language and Literature, Wuhan University, A218 Zhenhua Building, Wuhan University, Wuhan, Hubei, China. email: yli57@buffalo.edu

<sup>2</sup> Department of Japanese, Faculty of Letters, Yasuda Women's University, 6-13-1 Yasuhigashi, Asaminami-ku, Hiroshima 731-0153, Japan. email: miyagisi@yasuda-u.ac.jp



precede N or CLF, while adjectives exhibit greater positional flexibility. Possessors and noun modifiers are prenominal and restricted to N. Numerals must first form a constituent with a CLF. When N and CLF co-occur, the structure [N MOD CLF] is employed (cf. Li 2020; Miyagishi 2022).

Figure 1. Basic structure of Zauzou noun phrases (Li 2020: 21)

(POSS/N/REL)-(ADJ)-(DEM)-(N)-(REL/ADJ)-(DEM)-(NUM)-CLF-(ADJ)

Classifiers in Zauzou serve to indicate referential properties of noun phrases. Sortal classifiers exhibit two distinct syntactic functions: (1) they may follow a bare noun (i.e., classified noun) or combine with demonstratives to express definiteness, or (2) they co-occur with numerals to designate specific referents (Li 2023). Unclassified nouns, by contrast, typically function as kind-referring expressions. While pluralizers (see Miyagishi and Li, this issue) share functional similarities with sortal classifiers, they are excluded from the present study due to insufficient data availability.

Guoli Zauzou utilizes the same set of non-classifying nominalizers as those documented in the Jiangmo dialect (Miyagishi & Li, this issue). The most productive non-classifying nominalizers are the linking particles *xo*<sup>53</sup> (used in adjectival phrases) and *tə*<sup>31</sup> (used in relative constructions). Other non-classifying nominalizers with more limited distributions are thus excluded from the current analysis. Those nominalizers include: (1) the prefix *tə*<sup>55</sup>-, which nominalizes only adjectives and intransitive verbs, (2) the enclitic =*ku*<sup>55</sup>, which is associated with undergoers and adjuncts, and (3) the enclitic =*ci*<sup>55</sup>, which nominalizes either verb arguments or actions (Li 2020: 267-272). Moreover, in genitive noun phrases, Guoli Zauzou preferentially employs the genitive enclitic =*ze*<sup>55</sup> rather than classifiers.

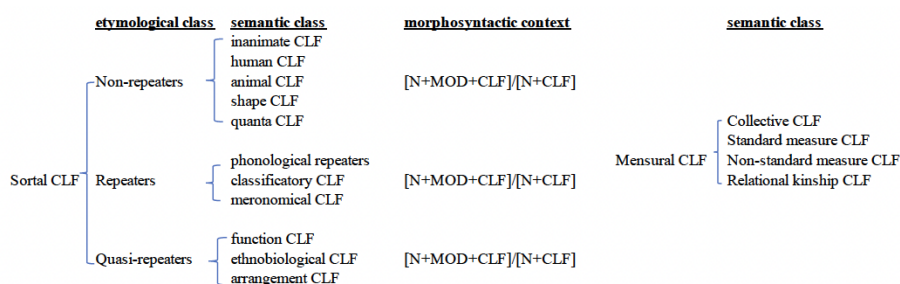
This overview of Guoli Zauzou synthesizes data from existing reference grammars of Zauzou (Li 2020; Miyagishi 2022) and Sun et al.'s (2002) grammatical sketch of the Guoli dialect. Fieldwork conducted in August 2024 in Lijiang and Dali, Yunnan Province, provided additional primary data on nominalization patterns. Two native speakers of Guoli Zauzou participated in this research, and the data analyzed in Section 4 originate from these elicited sessions.

## 2. OVERVIEW OF CLASSIFIER MARKING AND GRAMMATICAL NOMINALIZATIONS IN GUOLI ZAUZOU

### 2.1 Classifier marking in Guoli Zauzou

Guoli Zauzou employs both sortal and mensural classifiers, as displayed in Figure 2. The classifier system in the Guoli dialect exhibits minimal morphosyntactic and semantic variation relative to the Jiangmo dialect, though phonological differences occur in individual lexical items. Consequently, the present analysis follows the general description of Zauzou classifiers, with specific dialectal differences explicitly indicated where relevant.

Figure 2. Numeral classifier system in (Guoli/Jiangmo) Zauzou



This paper focuses specifically on sortal classifiers, as mensural classifiers—which categorize nouns based on quantity—fall beyond the scope of nominalization analysis. The sortal classifier system in Zauzou consists of three etymologically distinct categories: repeaters, non-repeaters, and quasi-repeaters. Example (1) illustrates these three subtypes in numeral phrases. Repeaters refer to ‘a noun that functions as its own classifier’ (Senft 1993: 104). The repeater *k<sup>h</sup>õ<sup>31</sup>* in (1a) repeats the noun stem *k<sup>h</sup>õ<sup>31</sup>* ‘river’. Non-repeaters, by contrast, form an independent class of sortal classifiers that do not appear in the noun stem, as shown by the human classifier *ja<sup>53</sup>* in (1b). Quasi-repeaters exhibit hybrid properties: while obligatory with isolated noun stems, they are frequently omitted in full noun phrases, as illustrated by *tõ<sup>13</sup>* in (1c).

(1)[N+NUM+CLF]: repeater, non-repeater and quasi-repeater

a. repeater

*k<sup>h</sup>õ<sup>31</sup> tã<sup>31</sup> k<sup>h</sup>õ<sup>31</sup> ku<sup>55</sup> nẽ<sup>33</sup> tẽ<sup>55</sup> tẽ<sup>h</sup>a<sup>31</sup> sẽ<sup>55</sup> k<sup>h</sup>uẽ<sup>33</sup> lõ<sup>13</sup>*  
 river one CLF:river cross TOP COP 3.6\_Yuan  
 ‘Crossing one river is (charged) 3.6 Yuan.’ (Zauzou marriage)

b. non-repeater

*jẽ<sup>33</sup> lã<sup>31</sup> ka<sup>33</sup> ʔõ<sup>55</sup> zu<sup>31</sup> tã<sup>31</sup> ja<sup>53</sup> ni<sup>55</sup> to<sup>53</sup>*  
 room CLF:inani LOC man one CLF:human sit DUR  
 ‘A man is sitting in the room.’

c. quasi-repeater

*na<sup>53</sup>-(tõ<sup>13</sup>) tã<sup>31</sup> tõ<sup>13</sup>*  
 pea-CLF:pod one CLF:pod  
 ‘one pea pod’

These three etymological classes are further subcategorized into eleven semantic classes as presented in Figure 2. In Zauzou, non-repeaters employ a widely attested set of universal semantic parameters for nominal classification, involving animacy, shape, rigidity, size, and quanta (Li 2024). Repeaters, by contrast, often stem from a phonological process (i.e., repeat the final syllable of the noun) rather than semantic correlation. They may nevertheless exhibit semantic relationships with their head nouns. Parameters of classification and meronymy are

particularly prominent in repeater-based systems. Quasi-repeaters, meanwhile, classify nouns according to distinct semantic parameters such as function, ethnobiological classification, and arrangement. Examples of these three classes of sortal classifiers, organized according to their semantic parameters, are presented in Table 1.

Notable lexical differences emerge between the Guoli and Jiangmo dialects in classifier-noun pairings. For example, while Jiangmo uses separate classifiers for animals of different sizes (Miyagishi & Li, this issue), Guoli employs a single, general animal classifier *ko*<sup>33</sup> ‘CLF:animal’ that categories both small and big animals.

All three etymological classes of sortal classifiers regularly occur in two primary morphosyntactic contexts: the bare classifier phrase [N+CLF] and the adnominal modification phrase [(N)+MOD+CLF]. In the latter context, the MOD position may be occupied by numerals, demonstratives, or relative clauses (including adjectives). While the CLF in the [(N)+MOD+CLF] is obligatory, the head noun is often elided in discourse. Examples in (2)–(5) illustrate the use of sortal classifiers across four distinct morphosyntactic contexts, with reference to specific individuals.

Table 1. Guoli Zauzou sortal classifiers and the semantic parameters in classification

| class of sortal classifier | semantic parameter     | classifier                                                           |                                                                                                    |
|----------------------------|------------------------|----------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| non-repeater               | animacy                | <i>ja</i> <sup>53</sup> ‘CLF:human’                                  |                                                                                                    |
|                            |                        | <i>ko</i> <sup>33</sup> ‘CLF:big animal’                             |                                                                                                    |
|                            |                        | <i>læ</i> <sup>31</sup> ‘CLF:inanimate’                              |                                                                                                    |
|                            | shape                  | 1-dimensional                                                        | <i>teo</i> <sup>55</sup> ‘CLF:string’<br><i>ʔa</i> <sup>55</sup> ‘CLF:stick’                       |
|                            |                        | 2-dimensional                                                        | <i>k<sup>h</sup>a</i> <sup>53</sup> ‘CLF:plank’<br><i>p<sup>h</sup>a</i> <sup>53</sup> ‘CLF:piece’ |
|                            |                        | 3-dimensional                                                        |                                                                                                    |
|                            | quanta                 | single in a pair                                                     | <i>p<sup>h</sup>o</i> <sup>53</sup>                                                                |
|                            |                        | paired                                                               | <i>tsẽ</i> <sup>13</sup>                                                                           |
| repeater                   | phonological repeating | <i>k<sup>h</sup>õ</i> <sup>31</sup> ‘CLF:river’                      |                                                                                                    |
|                            | classificatory         | <i>ma</i> <sup>31</sup> ‘CLF:hair’                                   |                                                                                                    |
|                            | meronymy               | <i>væ</i> <sup>53</sup> ‘CLF:flower’                                 |                                                                                                    |
| quasi-repeater             | ethnobiology           | <i>tsẽ</i> <sup>55</sup> ‘CLF:tree’                                  |                                                                                                    |
|                            | function               | <i>ka</i> <sup>33</sup> ‘CLF:locality’                               |                                                                                                    |
|                            | arrangement            | <i>tæ</i> <sup>33</sup> ‘CLF:cob’, <i>tõ</i> <sup>31</sup> ‘CLF:pod’ |                                                                                                    |

(2) [N+CLF]: non-repeater and quasi-repeater

- a. **ʔo**<sup>53</sup>**vur**<sup>31</sup> **ja**<sup>55</sup>                      tsu<sup>13</sup>te<sup>h</sup>o<sup>53</sup>    tɔ<sup>31</sup>    læ<sup>31</sup>                      wo<sup>13</sup>    wu<sup>13</sup>  
man                      CLF:human                      football                      one    CLF:inani    play    PROG  
‘The man is playing a football.’

- b.    ɲu<sup>55</sup>    tsa<sup>55</sup>            tsu<sup>31</sup> wu<sup>13</sup>            tə<sup>31</sup>            na<sup>53</sup>-tə<sup>31</sup>    tə<sup>31</sup>            ne<sup>33</sup>    sɛ̃<sup>31</sup>  
       1       pick\_off    eat    COMPL    LINK       pea-pod    CLF:pod    TOP    three  
       tə<sup>31</sup>  
       CLF:pod  
       ‘The pea pod I picked off are three pea pods.’

- (3) [N+DEM+CLF]: non-repeater  
       tsə<sup>33</sup>k<sup>h</sup>ə<sup>35</sup>    tə<sup>55</sup>    **ja**<sup>33</sup>                      wũ<sup>35</sup>    lə<sup>31</sup>            ɲə<sup>55</sup>  
       elderly       DIST CLF:human       sell    FP:assr       say  
       ‘That old man said that he will sell (the grains).’

- (4) [N+ADJ+CLF]  
       təo<sup>55</sup>    ʔo<sup>53</sup>    ts<sup>h</sup>u<sup>55</sup>    **ko**<sup>33</sup>  
       DIST pig    fat       CLF:animal  
       ‘that fat pig’

- (5) [REL+CLF]  
       [tɛ<sup>h</sup>a<sup>31</sup>pa<sup>55</sup>    xɛ̃<sup>31</sup>]<sub>REL</sub>    **ja**<sup>33</sup>                      ʔɛ<sup>31</sup>    tə<sup>55</sup>sə̃<sup>31</sup>            tə<sup>33</sup>    lə<sup>33</sup>  
       money       collect       CLF:human       also    DIST       one    CLF:inani  
       ‘The person who receives money also (uses) this one (this kind of bag).’

## 2.2 Grammatical nominalizations in Guoli Zauzou

### [1] Numerals

#### [1a] Cardinal numerals

Guoli Zauzou employs a base-ten numeral system composed of ten monosyllabic cardinal numerals denoting values from ‘one’ to ‘ten’. Notably, the language lacks a lexical form for the concept of ‘zero’. The full inventory of these basic numerals is provided in Table 2. Comparative analysis reveals that the numeral systems of the Guoli and Jiangmo dialects are principally distinguished by phonological variation.

Table 2. Guoli Zauzou simple cardinal numbers (Sun et al. 2002: 60)

| Simple Cardinal number          | Gloss |
|---------------------------------|-------|
| tuu <sup>31</sup>               | 1     |
| nɛ̃ <sup>53</sup>               | 2     |
| sɛ̃ <sup>33</sup>               | 3     |
| yi <sup>31</sup>                | 4     |
| ŋo <sup>33</sup>                | 5     |
| k <sup>h</sup> a <sup>33</sup>  | 6     |
| nɛ̃ <sup>55</sup>               | 7     |
| ia <sup>33</sup>                | 8     |
| ku <sup>33</sup>                | 9     |
| ts <sup>h</sup> e <sup>55</sup> | 10    |

### [1b] Ordinal numerals

Sun et al. (2002: 62-63) noted that ordinal numerals in Guoli Zauzou are formed through a specific construction: a classified cardinal numeral followed by the classifier phrase *tuu<sup>31</sup> ia<sup>55</sup>* ‘one X’. For example:

- (6) a.    nɛ̃<sup>53</sup>    ia<sup>55</sup> <sup>3</sup>                      tuu<sup>31</sup> <sup>4</sup> ia<sup>55</sup>  
          two   CLF.human       one   CLF.human  
          ‘the second person’ (Sun et al. 2002: 62)
- b.    yi<sup>31</sup>    ia<sup>55</sup>                      tuu<sup>31</sup>    ia<sup>55</sup>  
          four   CLF.human       one   CLF.human  
          ‘the fourth person’ (Sun et al. 2002: 62)
- c.    khi<sup>31</sup>    nɛ̃<sup>53</sup>    kō<sup>55</sup>                      tuu<sup>31</sup>    kō<sup>55</sup>  
          dog       two   CLF.animal       one   CLF.animal  
          ‘the second dog’ (Sun et al. 2002: 63)

### [2] Demonstratives

Demonstratives in Guoli Zauzou display a six-way semantic distinction that encodes both deictic and vertical dimensions. Functionally, these demonstratives are bound morphemes that require co-occurrence with numeral classifiers to form either pronominal or adverbial constructions. Table 3 provides examples of demonstrative pronouns and demonstrative adverbs in Guoli Zauzou.

<sup>3</sup> The classifier for human in Guoli Zauzou is transcribed as *ia<sup>55</sup>* by Sun et al. (2002), which is a phonological variant of *ja<sup>53</sup>/ja<sup>33</sup>* ‘CLF:human’.

<sup>4</sup> *tuu<sup>31</sup>* is the transcription of ‘one’ in Sun et al. (2002), which is a phonological variant of *tə<sup>31</sup>* ‘one’ in this paper.

Table 3. Demonstratives in Guoli Zauzou (Sun et al. 2002: 73-75)

| Deictic distinction   | Distance to the speaker            | demonstrative morpheme  | demonstrative pronoun                                             | demonstrative adverb                                       |
|-----------------------|------------------------------------|-------------------------|-------------------------------------------------------------------|------------------------------------------------------------|
| PROX (proximal)       | close to the speaker               | <i>ʔa<sup>55</sup></i>  | <i>ʔa<sup>55</sup> ia<sup>55</sup></i> ‘this person’              | <i>ʔa<sup>55</sup> ka<sup>55</sup></i> ‘here’              |
| DIST (distal)         | far from the speaker               | <i>tɛo<sup>35</sup></i> | <i>tɛo<sup>55</sup> ia<sup>55</sup></i> ‘that person’             | <i>tɛo<sup>35</sup> ka<sup>55</sup></i> ‘there (specific)’ |
| DIST (distal/general) | any place not close to the speaker | <i>tu<sup>55</sup></i>  | <i>tu<sup>55</sup> ia<sup>55</sup></i> ‘that person’              | <i>tu<sup>55</sup> ka<sup>55</sup></i> ‘there (general)’   |
| HIGH/UP               | higher than the speaker            | <i>ʔo<sup>55</sup></i>  | <i>ʔo<sup>55</sup> ia<sup>53</sup></i> ‘the upper person’         | <i>ʔo<sup>55</sup> ka<sup>55</sup></i> ‘up’                |
| LOW                   | lower than the speaker             | <i>mo<sup>55</sup></i>  | <i>mo<sup>55</sup> ia<sup>53</sup></i> ‘the lower person’         | <i>mo<sup>55</sup> ka<sup>55</sup></i> ‘down’              |
| DOWN                  | bottom                             | <i>yo<sup>33</sup></i>  | <i>yo<sup>33</sup> ia<sup>53</sup></i> ‘the person in the bottom’ | <i>yo<sup>33</sup> ka<sup>55</sup></i> ‘bottom’            |

### [3] Genitive noun phrases

In Guoli Zauzou, pronominal genitive noun phrases are formed through combining possessive pronouns with possessed nouns, as shown in Table 4. According to Sun et al. (2002), possessive pronouns in Guoli Zauzou exhibit two notable characteristics: (1) they exist exclusively in singular forms, lacking plural counterparts, and (2) they represent phonologically reduced variants of their corresponding personal pronouns (Sun et al. 2002: 57).

Table 4. Pronouns in Guoli Zauzou (Sun et al. 2002: 57)

| PERSON | personal pronouns      | possessive pronouns    | genitive noun phrase                                   |
|--------|------------------------|------------------------|--------------------------------------------------------|
| 1SG    | <i>ŋu<sup>55</sup></i> | <i>ŋu<sup>5</sup></i>  | <i>ŋu<sup>5</sup> pɔ<sup>31</sup></i> ‘my father’      |
| 2SG    | <i>no<sup>33</sup></i> | <i>nu<sup>31</sup></i> | <i>nu<sup>31</sup> pɔ<sup>31</sup></i> ‘your father’   |
| 3SG    | <i>tu<sup>55</sup></i> | <i>tu<sup>5</sup></i>  | <i>tu<sup>5</sup> pɔ<sup>31</sup></i> ‘his/her father’ |

Nominal genitive noun phrases in Guoli Zauzou comprise a nominal possessor and a possessed noun. Crucially, the genitive enclitic =*ze<sup>55</sup>* can be used after the possessor to indicate a possessive relationship between the two nominals in both pronominal and nominal genitive constructions. However, =*ze<sup>55</sup>* is optional when the possessed noun is overtly expressed (Sun et al. 2002: 119). See (7) and (8) for examples. Note that the use of =*ze<sup>55</sup>* is prohibited in the context of inalienable possession, as in (7a).

#### (7) Pronominal genitive noun phrases

- a. *ŋu<sup>55</sup>* (= \**ze<sup>33</sup>*)    *su<sup>31</sup> ŋã<sup>55</sup>*    *ja<sup>55</sup>*                      *ɛi<sup>55</sup>*    *zɔ<sup>31</sup>*  
 1SG=GEN    child            CLF:human    die    PFV  
 ‘My child died.’

- b.     $\eta u^{55}$     $ts^{h}a^{55}po^{33}$     $ts^{h}ue^{55}$     $tu^{55}$     $iu^{55}ei^{55}$     $la^{35}$     $z\phi^{31}$   
      1SG   book       CLF:book   3SG   take       go     PFV  
      ‘He took away my book.’ (Sun et al. 2002: 73)

- (8)   Nominal genitive noun phrases  
       $a^{55}ma^{55}$     $mi^{31}$         $ku^{33}$     $tu^{55}$         $ne^{33}$   
      mother   clothes   CLF   prefix       red  
      ‘Mother’s clothes is red.’ (Sun et al. 2002: 58)

The enclitic  $=ze^{33}$  is required in genitive noun phrases when the possessed noun is implicit, as demonstrated in (9).

- (9)    $x\epsilon^{35}z\phi^{33}$     $io^{35}$     $\text{?}a^{55}$         $k^{h}a^{33}$     $tea^{53}tu^{33}=ze^{33}$ ,    $\text{?}a^{31}-pe^{55}=ze^{33}$     $\text{?}a^{31}$     $\eta\epsilon^{53}$   
      maize       field   PROX       CLF   other=GEN       1-PL =GEN       NEG COP  
      ‘This maize field is others’, not ours.’ (Sun et al. 2002: 80)

#### [4] Adjectives

Adjectives in Guoli Zauzou can be nominalized through the nominalizing prefix  $tu^{55}$ , which creates nouns referring to the kind of objects possessing the quality denoted by the adjective, as illustrated in (10).

- (10) a.     $tu^{55}-i^{33}$  ‘the big ones’  
      b.     $tu^{55}-mia^{55}$  ‘the tall ones’  
      c.     $tu^{55}-ue^{55}$  ‘the yellow ones’ (Sun et al. 2002: 106)

Nominalized adjectives can modify a noun as a prenominal modifier, as demonstrated in (11), or alternatively, may appear postnominally, as in (12) (Sun et al. 2002: 106).

- (11)  $tu^{55}-ne^{33}$     $me^{33}$   
      NMLZ-red clothes  
      ‘red clothes’

- (12)  $me^{33}$         $ne^{33}$   
      clothes       red  
      ‘red clothes’

Sun et al. (2002: 141) pointed out that post-nominal attributive adjectives constitute the dominant word order in adjectival phrases of Guoli Zauzou.

## [5] Verb-based nominalizations

### [5a] Relative clauses

Based on the presence or absence of linking particles (i.e.,  $t\alpha^{13}$  and  $xo^{53}$ ), Li (2020) distinguishes two types of relative constructions in Zauzou: (1) zero-marked relative clauses and 2) marked relative clauses. Zero-marked relative clauses attach directly to the head NP in the matrix clause without any linking particles. In Guoli Zauzou, relative clauses often appear as prenominal modifiers, as illustrated in (13), where the relative clause appears unmarked. By contrast, marked relative clauses, exemplified in (14), employ a linking particle to connect the relative clause to the head NP.

(13) Zero-marked relative clauses (Sun et al. 2002: 142)

- a.  $\eta o^{33}$   $tsa^{33}$   $ts^{h}a^{55}k^{h}o^{13}$   
 fish fish old\_man  
 ‘the old man who fishes’
- b.  $tur^{55}$ - $pe^{55}$   $ku\epsilon^{55}$   $s\epsilon^{53}$   
 3-PL chop wood  
 ‘the wood they chopped’

(14) Marked relative clauses (Sun et al. 2002: 116)

|                                     |                  |                                        |                      |                                       |               |
|-------------------------------------|------------------|----------------------------------------|----------------------|---------------------------------------|---------------|
| $\eta a^{55}t\epsilon\epsilon^{55}$ | $nu^{35}xo^{33}$ | $\eta ue^{35}$ $t_2^{53}$ <sup>5</sup> | $tur^{55}$ $mo^{31}$ | $tur^{55}$ - $t\epsilon\epsilon^{33}$ | $\eta a^{31}$ |
| old_brother                         | beef             | buy LINK                               | DIST PL:inani        | NMLZ-fresh                            | NEG           |
| $\eta\epsilon^{53}$                 |                  |                                        |                      |                                       |               |
| COP                                 |                  |                                        |                      |                                       |               |

‘The beef the old brother bought is not fresh.’

Relative clauses can be classified according to the head noun’s position relative to the restricting clause (Keenan & Comrie 1977; Keenan 1985). In external-headed relative clauses, the head noun appears in its full form in the matrix clause but is absent in the relative clause. The examples in (13) demonstrate typical external-headed relative clauses in Guoli Zauzou. The external-headed relative clauses in (13) also represent the gapping strategy of relativization (Keenan 1985). Alternatively, the head noun may be fully expressed in both the relative clause and the matrix clause, yielding a correlative relative clause. (14) is an example of correlative relative clause in Guoli Zauzou. The head noun  $nu^{35}xo^{33}$  ‘beef’ is overtly expressed in the relative clause and is coreferential with the demonstrative phrase in the matrix clause. In terms of Keenan’s (1985) relativizing strategy, the correlative relative clause in (14) represents the non-reduction strategy, as it maintains a full noun within the relative clause.

<sup>5</sup>  $t_2^{53}$  can possibly be a particle indicating aspect. There are examples of aspectual  $t_2^{53}$  in Sun et al. (2002: 82, 98, 100, 282). In the current context, there is no way to tell the distinction between the two functions of  $t_2^{53}$ .



Guoli Zauzou lacks nominalizers in complement clause constructions. Complement clauses attach directly to matrix verbs without intervening nominalizing morphology, as shown in example (15).

- ### 3. DATA

The NP use of bare classifiers within genitive phrases is ungrammatical unless accompanied by the genitive enclitic =*ze*<sup>55</sup>, though the grammaticality of classifier =*ze*<sup>55</sup> co-occurrence remains partially indeterminate. Crucially, =*ze*<sup>55</sup> becomes obligatory when the possessed noun is unexpressed. Similarly, the modification use of bare classifiers in genitive phrases is also ungrammatical. While overt possessed nouns may permit =*ze*<sup>55</sup> alongside classifiers (with continued inter-speaker variation), =*ze*<sup>55</sup> itself exhibits optional modification use constrained by alienability distinctions.

Adjectival phrases employ both classifiers and the non-classifying nominalizer *xo*<sup>53</sup>. Classifiers show obligatory NP use with adjectives but optional modification use. The nominalizer *xo*<sup>53</sup> appears to function primarily as a linking particle, restricted to [ADJ N] constructions. It demonstrates neither independent adjectival nominalization (lacking NP use), nor [N ADJ] occurrence (lacking Mod use). Crucially, the co-occurrence of classifiers and non-classifying nominalizers is systematically prohibited in both NP and modification uses.

Relative clauses utilize both classifiers and non-classifying nominalizers ( $xo^{53}/ta^{51}$ ), with classifiers being optional in both NP and Mod uses. Bare nominalizers cannot independently mark relative clauses in NP use and must be accompanied by an additional classifier. In modification uses, non-classifying nominalizers are optional but rarely co-occur with a classifier.

Figure 3 through 5 illustrate the marking patterns of classifiers, genitive marker, and linking particles in Guoli Zauzou.

Figure 3. Classifier marking pattern in Guoli Zauzou

|          |       |        |         |          |                 |
|----------|-------|--------|---------|----------|-----------------|
| NP use:  | NUM > | DEM >  | ?*GEN > | ADJ >    | OPTV-based NMLZ |
| Mod use: | NUM > | ?DEM > | ?*GEN > | OPTADJ > | OPTV-based NMLZ |

Figure 4. Non-classifying linking particle pattern in Guoli Zauzou

NP use:           \*NUM >   \*DEM >   \*GEN >   \*ADJ >   \*V-based NMLZ

Mod use:        \*NUM >   \*DEM >   \*GEN >   \*<sup>OPT</sup>ADJ > <sup>OPT</sup>V-based NMLZ

Figure 5. Non-classifying genitive marking pattern in Guoli Zauzou

NP use:           \*NUM >   \*DEM >   GEN >   \*ADJ >   \*V-based NMLZ

Mod use:        \*NUM >   \*DEM >   <sup>OPT</sup>GEN >   \*ADJ >   \*V-based NMLZ

(\* = no marking or impossible; OPT = optional marking; \*OPT = marking applies to a restricted context; unmarked = obligatory marking; ? = no agreement)

### 3.1 Classifiers

#### [1] Numerals

##### [1a] Cardinal numerals

In Guoli Zauzou, classifiers mandatorily co-occur with cardinal numerals in both noun-phrase and modification uses. It should be noted that in modification uses, the [NUM CLF] sequence demonstrates flexible word order relative to the head noun, as exemplified in (17).

NP use

- (16)   ŋu<sup>55</sup>   yi<sup>31</sup>   ko<sup>33</sup>               na<sup>33</sup>  
          1SG   four   CLF:animal       like  
          ‘I want four (cats).’

Mod use

- (17) a.   ŋu<sup>55</sup>   nia<sup>53</sup>ni<sup>33</sup>       yi<sup>31</sup>   ko<sup>33</sup>               na<sup>33</sup>  
          1SG   cat               four   CLF:animal       like  
          ‘I want four cats.’
- b.   yi<sup>31</sup>   ko<sup>33</sup>               nia<sup>53</sup>ni<sup>33</sup>  
          four   CLF:animal       cat  
          ‘four cats’

#### [2] Demonstratives

Classifiers are obligatory with demonstratives in noun-phrase use, as shown in (18). In modification uses, its usage shows cross-speaker variation. While classifiers remain generally obligatory with demonstratives in modification uses—with (19a) representing the canonical

demonstrative phrase structure with an overt head noun—some Guoli dialect speakers occasionally permit classifier omission, as attested in (19b). The reduction of classifier in modification use is a marginal phenomenon in demonstrative phrases.

NP use

- (18) ʔa<sup>33</sup> ko<sup>33</sup> ji<sup>31</sup>  
 PROX CLF:animal big  
 ‘This is big (cat).’

Mod use

- (19) a. nia<sup>53</sup>ni<sup>33</sup> ʔa<sup>33</sup> ko<sup>33</sup> ji<sup>31</sup>  
 cat PROX CLF:animal big  
 ‘This cat is big.’
- b. ʔa<sup>33</sup> nia<sup>53</sup>ni<sup>33</sup> ji<sup>31</sup>  
 PROX cat big  
 ‘This cat is big.’

### [3] Genitive noun phrases

Guoli Zauzou employs the possessive enclitic =ze<sup>55</sup> in genitive noun phrases (§3.2). This enclitic is required when the possessed noun is unexpressed. Crucially, in NP uses, bare classifiers cannot generate grammatical genitive noun phrases with a possessor in the absence of =ze<sup>55</sup>, as demonstrated by the ungrammatical examples in (20a-b). However, acceptability judgments vary among speakers concerning classifier occurrence in =ze<sup>55</sup>-marked genitive phrases. See (20c) and (20d) for example.

Parallel to NP uses, the modification use of classifiers is prohibited when the possessive enclitic =ze<sup>55</sup> is absent. A bare classifier cannot grammatically occur between possessor and possessed noun without =ze<sup>55</sup>, as shown by the ill-formed (21a). Even with the presence of =ze<sup>55</sup>, inter-speaker variation persists regarding classifier acceptability in modification uses, as reflected in the divergent judgments for (21b)–(21c).

NP use

- (20) a. \*ʔa<sup>33</sup> kə<sup>31</sup> ŋu<sup>55</sup> kə<sup>31</sup>  
 PROX CLF:clothes 1SG CLF:clothes  
 Intended: ‘This is mine (skirt).’
- b. \*ŋə<sup>31</sup> kə<sup>31</sup> xa<sup>33</sup>ka<sup>33</sup> ?  
 2SG CLF:clothes where  
 Intended: ‘Where is yours (skirt)?’
- c. ʔa<sup>33</sup> kə<sup>31</sup> ŋu<sup>55</sup>=ze<sup>55</sup> kə<sup>31</sup>  
 PROX CLF:clothes 1SG=GEN CLF:clothes  
 Intended: ‘This is mine (skirt).’

- d.     $\eta\mathfrak{z}^{31}=\text{ze}^{55}$      $k\mathfrak{a}^{31}$                        $x\mathfrak{a}^{33}k\mathfrak{a}^{33}$     ?  
       2SG=GEN CLF:clothes        where  
       ‘Where is yours (skirt)?’

Mod use

- (21) a.     $*\eta\mathfrak{z}^{31}$   $k\mathfrak{a}^{31}$                        $t\mathfrak{e}^h\mathfrak{y}^{53}t\mathfrak{s}^{31}$   
       2SG CLF:clothes        skirt  
       Intended: ‘your skirt’
- b.     $*\eta\mathfrak{z}^{31}=\text{ze}^{55}$      $k\mathfrak{a}^{31}$                        $m\mathfrak{e}^{31}$   
       2SG=GEN CLF:clothes        shirt  
       Intended: ‘your shirt’ [speaker MYX]
- c.     $\eta\mathfrak{z}^{31}=\text{ze}^{55}$      $k\mathfrak{a}^{31}$                        $t\mathfrak{e}^h\mathfrak{y}^{53}t\mathfrak{s}^{31}$   
       2SG=GEN CLF:clothes        skirt  
       ‘your skirt’ [speaker ZJM]

#### [4] Adjectives

Adjectives in Guoli Zauzou may take an optional non-classifying nominalizer (e.g.,  $xo^{53}$ ) (§3.3) when modifying a noun or a classifier. In the absence of this nominalizer, classifiers become obligatory for adjectives in noun-phrase uses to derive either kind-referring or specific referents, as exemplified in (22a) and (22b), respectively. However, when the nominalizer is present, as in (22c), the NP use of classifiers with adjectives is less acceptable, though additional data are needed to confirm this preliminary observation.

In modification uses where the head noun is overtly expressed, classifiers occur optionally with adjectives. The occurrence of classifiers also interacts with the non-classifying nominalizer  $xo^{53}$ . When  $xo^{53}$  is absent, a classifier is optional, as can be seen in (23a)–(23e). Both orders [N ADJ] and [ADJ N] are attested in Guoli Zauzou. With the [N ADJ] order, a classifier forming a constituent with ADJ is optional. The classified and unclassified adjective differ in referentiality. (23a) illustrates that bare [N ADJ] renders a kind interpretation, while (23b) is ambiguous between kind and specific reference. The indefinite ‘one’ in (23c) can cancel this ambiguity.

With the [ADJ N] order, the classifier is also optional. (23d) demonstrates that the juxtaposition of an ADJ and a classified N<sup>6</sup> without an intervening classifier renders a specific reference in Guoli Zauzou. (23e) exemplifies the modification use of classifier in adjectival phrases, permitting either specific or generic interpretation.

<sup>6</sup> Note this post-nominal classifier forms a constituent with the noun rather than the adjective.

NP use

- (22) a.  $j\dot{i}^{55}\eta\tilde{a}^{55}$   $ko^{33}$   
 small CLF:animal  
 ‘small ones (cat as a kind)’
- b.  $j\dot{i}^{55}\eta\tilde{a}^{55}$   $t\emptyset^{31}$   $ko^{33}$   
 small one CLF:animal  
 ‘a small one (cat as an individual)’
- c.  $j\dot{i}^{31}/ *j\dot{i}^{31}\eta\tilde{a}^{55}$   $xo^{53}$   $ko^{33}$   
 big small LINK CLF:animal  
 ‘the big/small one (cat as an individual)’

Mod use: no NMLZ

- (23) a.  $\eta a^{55}ni^{33}$   $j\dot{i}^{55}\eta\tilde{a}^{55}$   
 cat small  
 ‘small cat (kind)’
- b.  $\eta a^{55}ni^{33}$   $j\dot{i}^{55}\eta\tilde{a}^{55}$   $ko^{33}$   
 cat small CLF:animal  
 ‘a small cat/small cat (kind)’
- c.  $\eta a^{55}ni^{33}$   $j\dot{i}^{55}\eta\tilde{a}^{55}$   $t\emptyset^{31}$   $ko^{33}$   
 cat small one CLF:animal  
 ‘a small cat’
- d.  $j\dot{i}^{55}\eta\tilde{a}^{55}$   $\eta a^{55}ni^{33}$   $ko^{33}$   
 small cat CLF:animal  
 ‘a small cat’
- e.  $j\dot{i}^{55}\eta\tilde{a}^{55}$   $ko^{33}$   $\eta a^{55}ni^{33}$   
 small CLF:animal cat  
 ‘a small cat/small cat (kind)’

When the nominalizer  $xo^{53}$  is overtly expressed, classifiers appear to be prohibited from co-occurring with it in modification uses. Both [N ADJ] and [ADJ N] orders demonstrate a strong preference for the absence of classifiers when  $xo^{53}$  is present, as evidenced by (24). Notably, we found no examples of [ADJ  $xo^{53}$  CLF N] in our data.

Mod use: overt NMLZ  $xo^{53}$

- (24) a.  $*\eta a^{55}ni^{33}$   $j\dot{i}^{31}$   $xo^{53}$   $ko^{33}$   
 cat big NMLZ CLF:animal  
 Intended: ‘big cat’

- b.    ʔkã<sup>13</sup>       xo<sup>53</sup>       tsẽ<sup>55</sup>       se<sup>53</sup>tsẽ<sup>55</sup>  
       tall       NMLZ       CLF:tree    tree  
       Intended: ‘tall tree’

## [5] Verb-based nominalizations

### [5a] Relative clauses

Classifiers are optional in relative clauses in their noun-phrase uses. Guoli Zauzou displays both unmarked and marked relative clauses. The former lacks a nominalizer (*tə<sup>31</sup>* or *xo<sup>53</sup>*), while the latter is overtly marked by *tə<sup>31</sup>/xo<sup>53</sup>*. As shown in (25a), the classifier may be absent when no nominalizer is present, whereas (25b) demonstrates its optional presence in such contexts. In NP uses, the non-classifying nominalizer *tə<sup>31</sup>* may co-occur with a classifier, as seen in (25c).

NP use:

- (25) a.   ʔa<sup>33</sup>ni<sup>33</sup>   ei<sup>55</sup>   zo<sup>31</sup>   ni<sup>33</sup>   ŋu<sup>55</sup>   ŋa<sup>55</sup>ni<sup>33</sup>   ko<sup>33</sup>  
           yesterday die   PFV TOP 1SG cat           CLF:animal  
           ‘The one that died yesterday is my cat.’
- b.    ʔa<sup>33</sup>ni<sup>33</sup>   ei<sup>55</sup>   zo<sup>31</sup>   ko<sup>33</sup>           ŋu<sup>55</sup>=ze<sup>55</sup>  
           yesterday die   PFV CLF:animal   1SG=GEN  
           ‘The one that died yesterday is mine.’
- c.    ʔa<sup>33</sup>ni<sup>33</sup>   ei<sup>55</sup>   tə<sup>31</sup><sup>7</sup>   ko<sup>33</sup>           ŋu<sup>55</sup>=ze<sup>55</sup>   ŋa<sup>55</sup>ni<sup>33</sup>  
           yesterday die   NMLZ   CLF:animal   1SG=GEN cat  
           ‘The one that died yesterday is my cat.’

Classifiers remain optional in relative clauses in their modification uses. For unmarked relative clauses, the classifier appears optionally, as demonstrated in (26a) and (26b). However, it is important to note that classifiers are not allowed when the relative clause is non-finite (i.e., specifically when the perfective marker *zo<sup>31</sup>* is absent), as shown in (26c). This restriction indicates that the modification use of classifiers with relative clauses is grammatically constrained. In marked relative clauses, inter-speaker variation exists regarding the co-occurrence of nominalizers and classifiers. While some speakers accept constructions where the nominalizer (*tə<sup>31</sup>* or *xo<sup>53</sup>*) appears with a classifier, as in (26d), others reject such combinations, judging sentences like (26e) unacceptable.

<sup>7</sup> *tə<sup>31</sup>* is ambiguous between a linking particle and the numeral word ‘one’. We need to test the co-occurrence condition of LINK and ‘one’ in a single NP.

Mod use

- (26) a.    ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    zo<sup>31</sup>    ŋa<sup>55</sup>ni<sup>33</sup>    ko<sup>33</sup>    ni<sup>33</sup>    ŋu<sup>55</sup>=ze<sup>55</sup>  
           yesterday die    PFV cat    CLF:animal    TOP 1SG=GEN  
           ‘The cat that died yesterday is mine.’
- b.    ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    zo<sup>31</sup>    ko<sup>33</sup>    ŋa<sup>55</sup>ni<sup>33</sup>    ko<sup>33</sup>    ŋu<sup>55</sup>=ze<sup>55</sup>  
           yesterday die    PFV CLF:animal    cat    CLF:animal    1SG=GEN  
           ‘The cat that died yesterday is mine.’
- c.    \*ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    ko<sup>33</sup>    ŋa<sup>55</sup>ni<sup>33</sup>    ko<sup>33</sup>    ŋu<sup>55</sup>=ze<sup>55</sup>  
           yesterday die    CLF:animal    cat    CLF:animal    1SG=GEN  
           Intended: ‘The cat that died yesterday is mine.’
- d.    ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    tə<sup>31</sup>    ko<sup>33</sup>    ŋa<sup>55</sup>ni<sup>33</sup>    ŋu<sup>55</sup>=ze<sup>55</sup>  
           yesterday die    NMLZ    CLF:animal    cat    1SG=GEN  
           ‘The cat that died yesterday is mine.’
- e.    \*ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    xo<sup>53</sup>    ko<sup>33</sup>    ŋa<sup>55</sup>ni<sup>33</sup>    ko<sup>33</sup>    ŋu<sup>55</sup>=ze<sup>55</sup>  
           yesterday die    NMLZ    CLF:animal cat    CLF:animal    1SG=GEN  
           Intended: ‘The cat that died yesterday is mine.’

### 3.2 Non-classifying nominalizers: the possessive marker =ze<sup>55</sup>

The possessive enclitic =ze<sup>55</sup> in Guoli Zauzou occurs exclusively in genitive noun phrases. It is obligatory when the possessed noun is unexpressed. Examples (27a) and (27b) demonstrate the mandatory presence of =ze<sup>55</sup> in the NP use, both with and without co-occurring classifiers. As previously established in (20a)–(20b), bare classifiers in NP use without =ze<sup>55</sup> is unacceptable in genitive phrases,.

NP use

- (27) a.    ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    zo<sup>31</sup>    ŋa<sup>55</sup>ni<sup>33</sup>    ko<sup>33</sup>    ni<sup>33</sup>    ŋu<sup>55</sup>=ze<sup>55</sup>  
           yesterday die    PFV cat    CLF:animal    TOP 1SG=GEN  
           ‘The cat that died yesterday is mine.’
- b.    ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    zo<sup>31</sup>    ŋa<sup>55</sup>ni<sup>33</sup>    ko<sup>33</sup>    ni<sup>33</sup>    ŋu<sup>55</sup>=ze<sup>55</sup>  
           yesterday die    PFV cat    CLF:animal    TOP 1SG=GEN  
           ko<sup>33</sup>  
           CLF:animal  
           ‘The cat that died yesterday is mine.’

In genitive noun phrases containing an explicit possessed noun, =ze<sup>55</sup> is obligatory in inalienable possession contexts, as in (28a), while remaining optional in alienable possession constructions, as in (28b). The grammaticality of =ze<sup>55</sup> co-occurring with classifiers in

modification use warrants further investigation, given the variable acceptability patterns observed in examples (21b)–(21c).

Mod use

- (28) a.     $\eta\mathfrak{o}^{31}$ - $\text{pe}^{33}$ = $\text{ze}^{55}$      $\text{me}^{33}$      $\text{m}\mathfrak{o}^{53}$   
          2-PL=GEN    clothes    PL:inani  
          ‘your clothes’
- b.     $\eta\mathfrak{o}^{31}$ (= $\text{ze}^{55}$ )  $\text{su}^{31}\eta\mathfrak{a}^{55}$   
          2=GEN    child  
          ‘your child’

### 3.3 Non-classifying nominalizers: the linking particle $x\mathfrak{o}^{53}/t\mathfrak{o}^{31}$

The two non-classifying nominalizers  $x\mathfrak{o}^{53}$  and  $t\mathfrak{o}^{31}$  in Guoli Zauzou are restricted to adjectival phrases and relative clauses. The particle  $x\mathfrak{o}^{53}$  is used to nominalize adjectives, and both forms are found with relative clauses.

#### [4] Adjectives

A nominalizer is not obligatory in adjectival phrases, regardless of whether the modified noun is present or absent (§3.1). In NP use, the linking particle  $x\mathfrak{o}^{53}$  cannot independently nominalize an adjective without a classifier, as shown in (29). However, the presence of an additional classifier may improve the acceptability of an  $x\mathfrak{o}^{53}$ -marked adjectival phrase, as in (22c). This suggests that  $x\mathfrak{o}^{53}$  functions primarily as a linking particle that requires a following nominal head. In modification use,  $x\mathfrak{o}^{53}$  can appear alone in [ADJ N] to connect ADJ and N, as in (30a). However, a bare  $x\mathfrak{o}^{53}$  is rarely used in the reversed [N ADJ] order, as seen in (30b). Additionally, when  $x\mathfrak{o}^{53}$  is overtly expressed, classifiers generally do not co-occur with it in its modification use, as previously illustrated in (24).

NP use

- (29)  $*\text{j}\mathfrak{i}^{55}\eta\mathfrak{a}^{55}$      $x\mathfrak{o}^{53}$   
          small    NMLZ  
          Intended: ‘the small one’

Mod use

- (30) a.     $\text{k}\mathfrak{a}^{33}$   $x\mathfrak{o}^{53}$      $\text{se}^{53}\text{ts}\mathfrak{e}^{55}$   
          tall NMLZ    tree  
          ‘tall trees’
- b.     $?\eta\mathfrak{a}^{55}\text{ni}^{33}$      $\text{j}\mathfrak{i}^{31}$      $x\mathfrak{o}^{53}$   
          cat    big NMLZ  
          Intended: ‘big cats’



Integrating the findings from §3.1 and §3.3 on adjectival phrases reveals that Guoli Zauzou systematically favors classifiers over nominalizers for adjective nominalization in both NP and Mod uses. The non-classifying nominalizer  $xo^{53}$  is strictly limited to [ADJ N], functioning solely as a linking particle. Crucially, when a classifier appears in an adjectival phrase, co-occurrence with  $xo^{53}$  typically results in ungrammaticality in both NP and Mod uses.

### [5a] Relative clauses

Both  $xo^{53}$  and  $tə^{31}$  serve as linking particles in marked relative constructions in Guoli Zauzou. They follow the relative clause to signal its nominalized status.

First of all, a bare nominalizer cannot occur on relative clauses in NP use. A relative clause can be joined to the matrix without any marking, as shown in (25a). The nominalizer preferably co-occurs with a classifier in NP use, as illustrated by (31b). A bare NMLZ  $tə^{31}$  in (31a) is ungrammatical.

NP use

- (31) a. \*ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    tə<sup>31</sup>            ŋu<sup>55</sup>=ze<sup>55</sup>    ŋa<sup>55</sup>ni<sup>33</sup>  
 yesterday    die    NMLZ    1SG=GEN cat  
 Intended: ‘The one that died yesterday is my cat.’
- b.    ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    tə<sup>31</sup>            ko<sup>33</sup>            ŋu<sup>55</sup>=ze<sup>55</sup>    ŋa<sup>55</sup>ni<sup>33</sup>  
 yesterday    die    NMLZ    CLF:animal    1SG=GEN cat  
 ‘The one that died yesterday is my cat.’

In modification use, the nominalizers  $xo^{53}/tə^{31}$  occur optionally. This optionality is exemplified by the unmarked relative construction in (26a). A bare nominalizer can appear between the relative clause and head noun, as illustrated in (32a). The co-occurrence of nominalizer and classifier in modification uses is dispreferred, as demonstrated by (32b). However, inter-speaker variation exists, with some speakers accepting such combinations in modification uses, as in (32c).

Mod use

- (32) a.    ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    tə<sup>31</sup>/ $xo^{53}$             ŋa<sup>55</sup>ni<sup>33</sup>    (ko<sup>33</sup>)            ŋu<sup>55</sup>=ze<sup>55</sup>  
 yesterday    die    NMLZ/NMLZ    cat            CLF:animal    1SG=GEN  
 ‘The cat that died yesterday is mine.’
- b.    ʔa<sup>33</sup>ni<sup>33</sup>    ei<sup>55</sup>    tə<sup>31</sup>            ko<sup>33</sup>            ŋa<sup>55</sup>ni<sup>33</sup>    ŋu<sup>55</sup>=ze<sup>55</sup>  
 yesterday    die    NMLZ    CLF:animal    cat            1SG=GEN  
 ‘The cat that died yesterday is mine.’

|    |                                                  |                  |                  |                  |                                   |                  |
|----|--------------------------------------------------|------------------|------------------|------------------|-----------------------------------|------------------|
| c. | *ʔa <sup>33</sup> ni <sup>33</sup>               | ei <sup>55</sup> | xo <sup>53</sup> | ko <sup>33</sup> | ŋa <sup>55</sup> ni <sup>33</sup> | ko <sup>33</sup> |
|    | yesterday                                        | die              | NMLZ             | CLF:animal       | cat                               | CLF:animal       |
|    | ŋu <sup>55</sup> =ze <sup>55</sup>               |                  |                  |                  |                                   |                  |
|    | 1SG=GEN                                          |                  |                  |                  |                                   |                  |
|    | Intended: ‘The cat that died yesterday is mine.’ |                  |                  |                  |                                   |                  |

Results from §3.1 and §3.3 point to that classifiers and non-classifying nominalizers exhibit complementary distribution in NP and Mod uses. Relative clauses in Guoli Zauzou display a preference for non-classifying nominalizers over classifiers in modification uses. Conversely, classifiers are preferred over non-classifying nominalizers in NP uses. Critically, non-classifying nominalizers alone do not seem to fulfill the nominalization function in relative constructions unless paired with a classifier.

#### 4. ANALYSIS AND DISCUSSION

The Guoli dialect demonstrates nominalization patterns consistent with those observed in Jiangmo dialect (Miyagishi & Li, this issue), with both aligning closely with Shibatani’s (2023) nominalization hierarchy.

First, in both dialects, the genitive enclitic =ze<sup>55</sup> and classifiers exhibit complementary distribution across the five categories of Shibatani’s hierarchy. For Guoli Zauzou, classifiers serve nominalizing functions for four categories—cardinal numerals, demonstratives, adjectives, and relative clauses—but are excluded from genitive noun phrases. Classifier optionality increases progressively along the hierarchy rightward. Genitive noun phrases in both Guoli and Jiangmo dialects exclusively favor the genitive enclitic =ze<sup>55</sup>, which cannot extend to other categories on the nominalization hierarchy.

Second, non-classifying nominalizers in both dialects, specifically xo<sup>53</sup> and tə<sup>31</sup>, are confined to the context of adjectival phrases and relative clauses. In the case of Guoli Zauzou, these nominalizers do not function as genuine nominalizers. While true nominalizers exhibit NP use, these particles operate primarily as linking particles, optionally connecting adjectives or relative clauses to overt head nouns exclusively in modification uses.

Shibatani (2021, 2023) hypothesize that classifier/gender marking progresses following the hierarchy in both structural (i.e., NUM > DEM > GEN > ADJ > V-based nominalization) and functional (i.e., NP use > Mod use) dimensions. Classifier/gender marking starts out with numerals and then spreads to the constructions towards the right edge of the hierarchy in a contiguous manner (Shibatani 2023: 35).

Based on the observed patterns characterized in the above sections, classifiers emerge as the primary nominalizing device in Guoli Zauzou. Their obligatoriness and optionality across five categories of NPs align with Shibatani’s (2023) nominalization hierarchy, which effectively predicts the functional expansion of classifiers from N-based to V-based nominalizations, and from NP use to MOD use. The only unexpected pattern occurs in genitive noun phrases, which exhibits a preference for the genitive enclitic =ze<sup>55</sup> over classifiers, resulting in a non-contiguous nominalization hierarchy.

Several unresolved issues in the Guoli Zauzou data require further investigation. First, the interaction between classifying and non-classifying nominalizers within individual noun

phrases remains incompletely understood. This includes clarifying the co-occurrence conditions of classifiers with linking particles in adjectival and relative clauses, as well as classifiers with the genitive enclitic =*ze*<sup>55</sup> in genitive phrases. While the preceding analysis (§3.1 and §3.3) has identified several constraints on their co-occurrence within single noun phrases, additional data are necessary to confirm these patterns. Second, a comprehensive account of non-classifying nominalizers must incorporate the word order variation inherent to Zauzou syntax, such as the flexible order of adjectives and nouns (§3.3). The placement of the noun in either initial or final position within a noun phrase clearly affects the grammatical acceptability of the construction, suggesting that word order plays a critical role in licensing nominalizer usage. Finally, the exceptional status of genitive noun phrases in Zauzou is not an isolated phenomenon. Parallel patterns have been observed in a handful of Ngwi languages of the Tibeto-Burman family. A larger scale comparative study of genitive marking across these languages would yield valuable insights into the universality of Shibatani’s nominalization hierarchy and the typological status of possession within grammatical systems.

## 5. CONCLUSION

Grammatical nominalization in Guoli Zauzou encompasses two distinct strategies: classifying nominalizers (i.e., sortal classifiers) and non-classifying nominalizers (i.e., the genitive enclitic =*ze*<sup>55</sup> and two linking particles). Nominalization via classifiers follows Shibatani’s (2023) structural hierarchy, extending from numerals to verb-based nominalizations but excluding genitive noun phrases. The genitive enclitic =*ze*<sup>33</sup> is exclusively used in genitive noun phrases, forming an exceptional category to Shibatani’s structural hierarchy. Linking particles in Guoli Zauzou are restricted to modification uses within adjectival phrases and relative constructions, and are thus not genuine nominalizers. Classifiers dominate in V-based NP uses (i.e., adjectival phrases and relative clauses), whereas linking particles are favored in V-based Mod uses. Co-occurrence of classifiers and linking particles is generally dispreferred. The findings of this study generally support Shibatani’s (2023) hypothesis. However, several remaining issues, including the unexpected pattern in genitive noun phrases, present critical deviations from the predicted hierarchy and underscore the necessity for further analysis.

## ABBREVIATIONS

|       |                   |
|-------|-------------------|
| 1     | first person      |
| 3     | third person      |
| ADJ   | adjective         |
| NUM   | numeral           |
| CLF   | classifier        |
| COMPL | completive aspect |
| COP   | copula            |
| DEM   | demonstrative     |
| DIST  | distal            |
| DUR   | durative aspect   |
| FP    | final particle    |

|      |                    |
|------|--------------------|
| GEN  | genitive marker    |
| LINK | linking particle   |
| LOC  | locative           |
| N    | noun               |
| NEG  | negation           |
| NMLZ | nominalizer        |
| NUM  | numeral            |
| PFV  | perfective aspect  |
| PL   | plural             |
| POSS | possessor          |
| PROG | progressive aspect |
| REL  | relative clause    |
| SG   | singular           |
| TOP  | topic              |

## ACKNOWLEDGMENTS

We would like to express our deepest gratitude to our Guoli Zauzou consultant, Mu yongxin and Zhang jiangmei. The present study has been supported by grants from the Japan Society for the Promotion of Science (KAKENHI Project IDs: 22H00659 and 23K21931, ‘Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers’, PI: Kazuhiro Kawachi) and from the National Institute for Japanese Language and Linguistics (‘Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective’, PI: Prashant Pardeshi).

## REFERENCES

- Keenan, Edward L. 1985. Relative Clauses. In Timothy Shopen. (ed.) *Language typology and syntactic description (Vol. 2)*, 141–170. Cambridge: Cambridge University Press.
- Keenan, Edward L. & Comrie, Bernard. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8, 63–99.
- LaPolla, Randy J. 2007. Overview of Sino-Tibetan morphosyntax. In Graham Thurgood & Randy J. LaPolla. (eds.) *Sino-Tibetan languages* (2nd edition), 40–69. London & New York: Routledge.
- Li, Yu. 2020. *A grammar of Zauzou*. Ph.D dissertation. The State University of New York at Buffalo, Buffalo.
- Li, Yu. 2023. Referentiality and thematic importance: the discourse function of Zauzou classifiers. *Lingua* 285.  
<https://www.sciencedirect.com/science/article/abs/pii/S0024384123000013>
- Li, Yu. 2024. Sortal classifiers and noun categorization in Zauzou. *Language and Linguistics* 25(2), 270–317.
- Miyagishi, Tetsuya. 2022. *Zozogo (jakujuugo) bunpoo (A grammar of Zauzou)*. Hiroshima: Yasuda Women’s University.
- Miyagishi, Tetsuya, and Li, Yu. This issue. Grammatical nominalizations using classifiers, pluralizers, and non-classifying nominalizers in the Jiangmo dialect of Zauzou.

- Senft, Gunter. 1993. A grammaticalization hypothesis on the origin of Kilivila classificatory particles. *Sprachtypologie und Universalienforschung* 46, 100–112.
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)
- Sun, Hongkai, and Huang, Chenglong, and Zhou, Maocao. 2002. *Rouruoyu yanjiu (A study of Zauzou)*. Beijing: China Minzu University Press.

# Grammatical nominalization in the Jiangmo dialect of Zauzou

Tetsuya Miyagishi<sup>1</sup> and Yu Li<sup>2</sup>

## ABSTRACT

This paper presents an interim report on the use of classifiers, pluralizers, and non-classifying nominalizers as grammatical nominalizers in the Jiangmo dialect of Zauzou (hereafter referred to as Jiangmo Zauzou), a language of the Ngwi subgroup within the Tibeto-Burman family. The data from this study reveal that the three types of nominalizers in Jiangmo Zauzou exhibit distinct distributions and extensions due to varying degrees of classification. The findings also demonstrate that classifier-based nominalization attains the highest level in Shibatani's (2021: 524, 2023: 35) classifier hierarchy. However, genitive nominalization via Jiangmo Zauzou's nominalizers cannot be fully explained within the current framework of Shibatani's theory.

## 1. INTRODUCTION

This paper categorizes nominalizers in Jiangmo Zauzou into three types: classifiers, pluralizers, and non-classifying nominalizers. Classifiers categorize nouns into distinct semantic classes, while pluralizers encode plural number and classify nouns exclusively based on animacy. Non-classifying nominalizers, by contrast, operate without classification. This study examines the distribution and syntactic extension of grammatical nominalizations formed by each type.

The findings reveal that classifier-based nominalizations occur across all levels of Shibatani's (2021: 524, 2023: 35) hierarchy, spanning numerals to verbs. In comparison, pluralizer-derived nominalizations are attested from demonstratives to verbs but exclude numerals. Non-classifying nominalizers are restricted to the stages of adjectives and verbs. These results suggest a potential correlation between the degree of classification and the functional range of nominalizations.

Jiangmo Zauzou employs an array of nominalizers. In addition to the highly productive  $=xo^{53}$  and  $tə^{13}$ , several others display more constrained usage and are excluded from the present analysis. For example, the prefix  $tə^{55}$  nominalizes only adjectives and intransitive verbs, while the enclitic  $=ku^{55}$  specifically targets undergoers and adjuncts. The enclitic  $=ei^{55}$  nominalizes the participants or the action itself (Li 2020: 267–272).

The basic structure of Jiangmo Zauzou noun phrases, as shown in Figure 1, is a structure where modifiers can occur both before and after the head noun (N).

---

<sup>1</sup> Department of Japanese, Faculty of Letters, Yasuda Women's University, 6-13-1 Yasuhigashi, Asaminami-ku, Hiroshima 731-0153, Japan. email: miyagisi@yasuda-u.ac.jp

<sup>2</sup> School of Chinese Language and Literature, Wuhan University, A218 Zhenhua Building, Wuhan University, Wuhan, Hubei, China. email: yli57@buffalo.edu

Figure 1. Basic structure of Jiangmo Zauzou noun phrases (Li 2020: 21)

(Poss/NMod/RC)-(Adj)-(Dem)-(N)-(RC/Adj)-(Dem)-(Num)-CLF-(Adj)

Adjectives and nouns are two distinct categories in Jiangmo Zauzou, as they can be distinguished morphosyntactically by the type of negators, their compatibility with the degree word *lo<sup>53</sup>ei<sup>13</sup>*, and their ability to host the nominalizing prefix *tə<sup>55</sup>*- (Li 2020: 532). Sortal classifiers and pluralizers constitute two classes of classifying nominalizers. They are typically used with numerals, demonstratives, adjectives and relative clauses to designate specific referents, and may also directly follow bare nouns to express definiteness. As for non-classifying nominalizers, in addition to the highly productive *=xo<sup>53</sup>*, which will be discussed in detail later, there are also the prefix *tə<sup>55</sup>*- (e.g., *tə<sup>55</sup>-ni<sup>33</sup>* ‘blue one’), the enclitic *=ku<sup>55</sup>* (e.g., *te<sup>h</sup>ə<sup>55</sup>=ku<sup>55</sup>* ‘sweet things’), and the enclitic *=ei<sup>55</sup>* (e.g., *tsu<sup>33</sup>=ei<sup>33</sup>* ‘eating’), as mentioned previously.

The data come mainly from: (i) materials collected by Li Shaoen, a native speaker and language expert of Jiangmo Zauzou, including interviews with him; and (ii) materials from Li Yu's doctoral dissertation *A Grammar of Zauzou*, which contains spoken data and oral literature collected from native speakers of Jiangmo Zauzou for the dissertation.

## 2. OVERVIEW OF CLASSIFIER MARKING AND GRAMMATICAL NOMINALIZATIONS IN JIANGMO ZAUZOU

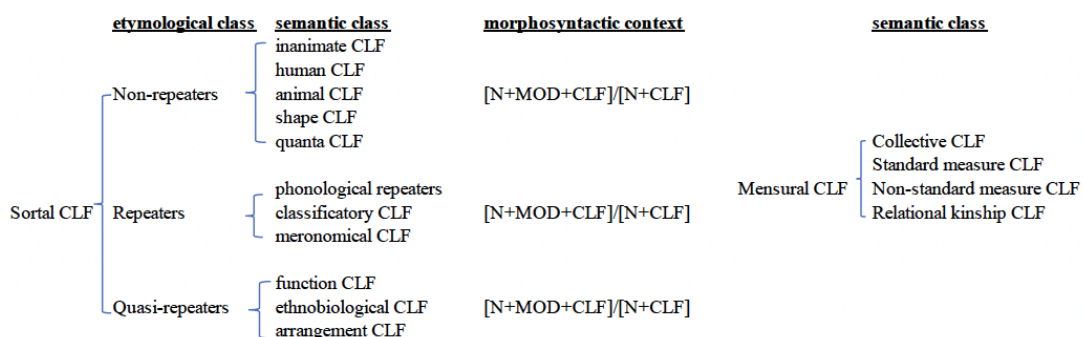
### 2.1 Classifier marking in Jiangmo Zauzou

The nominal classification systems of Jiangmo Zauzou comprise primarily a classifier system alongside a binary animate-inanimate classification system mediated by pluralizers. Both systems participate in grammatical nominalization processes. Furthermore, while masculine and feminine affixes may mark gender for animate entities, these occur exclusively in lexical nominalizations, as Jiangmo Zauzou lacks grammatical gender. This section will analyze Jiangmo Zauzou's classifiers and pluralizers.

#### 2.1.1 Sortal classifiers

Jiangmo Zauzou employs both sortal and mensural classifiers. Figure 2 characterizes the Jiangmo Zauzou numeral classifier system (Li 2024).

Figure 2. Numeral classifier system in Jiangmo Zauzou



This paper focuses exclusively on sortal classifiers, as mensural classifiers categorize nouns by quantity and are consequently irrelevant to nominalization. Jiangmo Zauzou sortal classifiers comprise three etymological subtypes: repeaters, non-repeaters, and quasi-repeaters. Example (1) illustrates these three classifier subtypes in numeral phrases. Repeaters are defined as nouns functioning as their own classifiers (Senft 1993: 104). The repeater *kʰɔ̃<sup>31</sup>* in (1a) repeats the noun stem *kʰɔ̃<sup>31</sup>* ‘river’. Non-repeaters, conversely, constitute an independent class of sortal classifiers that do not appear in the noun stems, exemplified by the human classifier *ja<sup>53</sup>* in (1b). Quasi-repeaters exhibit hybrid properties: they are obligatory in isolated noun stems but often omitted in noun phrases, as shown by *tɔ̃<sup>13</sup>* in (1c).

(1) [N NUM CLF]: repeater, non-repeater and quasi-repeater

a. repeater

**kʰɔ̃<sup>31</sup> tɔ̃<sup>31</sup> kʰɔ̃<sup>31</sup> ku<sup>55</sup> nɛ<sup>33</sup> tɔ̃<sup>55</sup> tɛ<sup>h</sup>a<sup>31</sup> sɛ̃<sup>55</sup> kʰuɛ<sup>33</sup> lɔ<sup>13</sup>**  
 river one CLF:river cross TOP COP 3.6\_Yuan  
 ‘Crossing one river is (charged) 3.6 Yuan.’ (Zauzou marriage)

b. non-repeater

**jɛ̃<sup>33</sup> lɛ̃<sup>31</sup> ka<sup>33</sup> ʔo<sup>55</sup> zu<sup>31</sup> tɔ̃<sup>31</sup> ja<sup>53</sup> ni<sup>55</sup> to<sup>53</sup>**  
 roomCLF:inani LOC man one CLF:human sit DUR  
 ‘A man is sitting in the room.’

c. quasi-repeater

**na<sup>53</sup>-(tɔ̃<sup>13</sup>) tɔ̃<sup>31</sup> tɔ̃<sup>13</sup>**  
 pea-CLF:pod one CLF:pod  
 ‘one pea pod’

Using semantic parameters, these three etymological classes are further subcategorized into eleven semantic classes as presented in Figure 1. Nominal classification through non-repeaters in Jiangmo Zauzou primarily follows universal semantic parameters repeatedly used in noun categorization, involving animacy, shape, rigidity, size, and quanta (Li 2024). Classification by repeaters does not necessarily correlate with the semantic category of the referent but rather stems from a phonological operation that repeats the final syllable of the noun. However, repeaters may also exhibit semantic relationships with their head nouns. In Jiangmo Zauzou, classification and meronymy emerge as the most prevalent parameters in repeater-based classification. Quasi-repeaters classify nouns according to semantic parameters such as function, ethnobiology, and arrangement. Table 1 presents examples of these three sortal classifier classes organized by their semantic parameters.

All three etymological classes of sortal classifiers regularly occur in two morphosyntactic contexts: the bare classifier phrase [N CLF] and the adnominal modification construction [(N) MOD CLF]. In the second context, the MOD position can be filled by a numeral, demonstrative, or relative clause (including adjectives). While the CLF in [(N) MOD CLF] is obligatory, the head noun frequently becomes optional in discourse. The use of numeral word in the MOD position has been illustrated in (1); examples (2)–(4) now illustrate classifier usage in other



morphosyntactic contexts. The bare classifier phrase [N CLF] in (1) renders a definite singular interpretation.

Table 1. Jiangmo Zauzou sortal classifiers and the semantic parameters in classification

| class of sortal classifier | semantic parameter     | Classifier                                              |                                                                                          |
|----------------------------|------------------------|---------------------------------------------------------|------------------------------------------------------------------------------------------|
| non-repeater               | animacy                | ja <sup>53</sup> ‘CLF:human’                            |                                                                                          |
|                            |                        | animal                                                  | ʔɔ <sup>55</sup> ‘CLF:small animal’,<br>kɔ <sup>55</sup> ‘CLF:big animal’                |
|                            |                        | lə <sup>31</sup> ‘CLF:inanimate’                        |                                                                                          |
|                            | shape                  | 1-dimensional                                           | tɕo <sup>55</sup> ‘CLF:string’<br>ʔa <sup>55</sup> ‘CLF:stick’                           |
|                            |                        | 2-dimensional                                           | k <sup>h</sup> a <sup>53</sup> ‘CLF:plank’<br>p <sup>h</sup> a <sup>53</sup> ‘CLF:piece’ |
|                            |                        | 3-dimensional                                           | to <sup>53</sup> ‘CLF:drop’                                                              |
|                            | quanta                 | single in a pair                                        | p <sup>h</sup> o <sup>53</sup>                                                           |
|                            |                        | paired                                                  | tsɛ̃ <sup>13</sup>                                                                       |
| Repeater                   | phonological repeating | k <sup>h</sup> ɔ̃ <sup>31</sup> ‘CLF:river’             |                                                                                          |
|                            | classificatory         | ma <sup>31</sup> ‘CLF:hair’                             |                                                                                          |
|                            | meronymy               | væ <sup>53</sup> ‘CLF:flower’                           |                                                                                          |
| quasi-repeater             | ethnobiology           | tsɛ̃ <sup>55</sup> ‘CLF:tree’                           |                                                                                          |
|                            | function               | ka <sup>33</sup> ‘CLF:locality’                         |                                                                                          |
|                            | arrangement            | tæ <sup>33</sup> ‘CLF:cob’, tɔ̃ <sup>31</sup> ‘CLF:pod’ |                                                                                          |

(1) [N CLF]: non-repeater and quasi-repeater

- a. ʔo<sup>53</sup>vuu<sup>31</sup> ja<sup>55</sup> tsu<sup>13</sup>te<sup>h</sup>o<sup>53</sup> tɔ̃<sup>31</sup> lə<sup>31</sup> wo<sup>13</sup> wu<sup>13</sup>  
man CLF:human football one CLF:inani play PROG  
‘The man is playing a football.’
- b. ŋu<sup>55</sup> tsa<sup>55</sup> tsu<sup>31</sup> wu<sup>13</sup> tɔ̃<sup>31</sup> na<sup>53</sup>-tɔ̃<sup>31</sup> tɔ̃<sup>31</sup> ne<sup>33</sup> sɛ̃<sup>31</sup>  
1 pick\_off eat COMPL LINK pea-pod CLF:pod TOP three  
tɔ̃<sup>31</sup>  
CLF:pod  
‘The pea pod I picked off are three pea pods.’

(2) [N DEM CLF]: non-repeater

- tsɔ̃<sup>33</sup>k<sup>h</sup>ɔ̃<sup>35</sup> tɔ̃<sup>55</sup> ja<sup>33</sup> wũ<sup>35</sup> lə<sup>31</sup> ŋə<sup>55</sup>  
elderly DIST CLF:human sell FP:assr say  
‘That old man said that he will sell (the grains).’

(3) [N ADJ CLF]

tɕo<sup>55</sup> ʔo<sup>53</sup> tsʰu<sup>55</sup> ʔõ<sup>55</sup>  
 DIST pig fat CLF:animal  
 ‘that fat pig’

(4) [REL CLF]

[tɕʰa<sup>31</sup>pa<sup>55</sup> xẽ<sup>31</sup>]<sub>REL</sub> ja<sup>33</sup> ʔɛ<sup>31</sup> tã<sup>55</sup>sã<sup>31</sup> tã<sup>33</sup> læ<sup>33</sup>  
 money collect CLF:human also DIST one CLF:inani  
 ‘The person who receives money also (uses) this one (this kind of bag).’

## 2.1.2 Pluralizers

Jiangmo Zauzou employs two plural quantifiers for common nouns: the animate *tã<sup>13</sup>* ‘PL:ani’ and the inanimate *mõ<sup>53</sup>* ‘PL:inani’. Li (2023) analyzes these as constituting an independent grammatical category in the language. The quantifiers stem from the pronominal plural suffixes *-tã<sup>13</sup>* and *-mõ<sup>53</sup>*, which mark plural number on non-1st person pronouns (Li 2020). Plural quantifiers can extend to mark plural number in Jiangmo Zauzou noun phrases. They primarily combine with bare nouns to specify plural quantity for animate and inanimate referents, as demonstrated in (5). Crucially, plural quantifiers differ from canonical classifiers in their semantic scope: while quantifiers only distinguish animacy, classifiers further categorize nouns into finer semantic classes. In (5), for example, *tã<sup>13</sup>* categorizes both human and animals, while *mõ<sup>53</sup>* categorizes count and mass nouns.

(5) [N PL]

- a. wu<sup>55</sup> kã<sup>55</sup> tea<sup>53</sup> tã<sup>13</sup> tɕʰyɪ<sup>31</sup> tã<sup>13</sup> ta<sup>55</sup> tsẽ<sup>53</sup> le<sup>13</sup>  
 village LOC other PL:ani dog PL:ani run meet come  
 ‘Other people’s dogs in the village ran to us.’ (Calling back souls)
- b. pɔ<sup>13</sup> jo<sup>55</sup> tã<sup>13</sup> mi<sup>55</sup> mõ<sup>53</sup> ʔɛ<sup>31</sup> tã<sup>31</sup> tẽã<sup>31</sup>  
 grandfather grandmother PL:ani name PL:inani also one VCL  
 ta<sup>33</sup>mi<sup>33</sup> xe<sup>31</sup>  
 ask FP  
 ‘I want to ask something about my grandparents’ names.’ (Names of ancestors)
- c. ʔɛ<sup>31</sup>ɲue<sup>53</sup> mõ<sup>53</sup> ka<sup>55</sup> jɛ<sup>55</sup> tẽã<sup>55</sup>  
 grind PL:inani LOC grind PURP  
 ‘Go to the grinds to grind.’ (Selling bean jelly)
- d. nã<sup>55</sup>-tã<sup>13</sup> ɛyɪ<sup>31</sup> mõ<sup>53</sup> nã<sup>31</sup> pɔ<sup>13</sup> fu<sup>53</sup> wu<sup>13</sup> zɔ<sup>31</sup>  
 2PL-PL:ani blood PL:inani 2 father drink COMPL PFV  
 ‘Your blood is drunk by your father.’ (A farmer’s family)

Moreover, plural quantifiers do not serve individuation functions. Unlike classifiers, which may combine with any cardinal number for enumeration, the plural quantifiers *tã<sup>13</sup>* and

*mõ*<sup>53</sup> cannot co-occur with cardinal number (i.e., *tə*<sup>13</sup> may co-occur with ‘one’, which has a special status as a morpheme signaling indefiniteness but not cardinality), as evidenced in (6) and (7). The incompatibility with cardinal numbers demonstrates that plural quantifiers are prohibited from counting contexts the primary domain of numeral classifiers.

(6) \*[N NUM PL]

- a. \*sɿ<sup>55</sup>mi<sup>13</sup>    næ<sup>53</sup> tə<sup>13</sup>  
       woman      two    PL:ani  
       Intended: ‘two women’
- b. \*jẽ<sup>33</sup>            næ<sup>53</sup> mõ<sup>53</sup>  
       house        two    PL:inani  
       Intended: ‘two houses’

(7) [N ‘one’ tə<sup>13</sup>/\*mõ<sup>53</sup>]

- a. tɛa<sup>55</sup>sɿ<sup>55</sup>    tə<sup>31</sup>    tə<sup>13</sup>  
       person     one    PL:ani  
       ‘a group of people’
- b. \*wu<sup>53</sup>vuu<sup>13</sup>    tə<sup>31</sup>    mõ<sup>53</sup>  
       egg            one    PL:inani  
       Intended: ‘a few eggs’

### 2.1.3 The versatile classifier *və*<sup>13</sup>

Jiangmo Zauzou employs a special classifier that exhibits properties of both soral classifiers and pluralizers the versatile classifier *və*<sup>13</sup>. Presumably, this classifier developed from a possessive marker, cognate with the *-vi* suffix in Nuosu Yi (Gerner 2013) and *ve* in Lahu (Matisoff 1973:142).

First, *və*<sup>13</sup> functions as a general classifier for all inanimate nouns (both count and mass nouns) (Li 2023). Categorization by *və*<sup>13</sup> is semantically underspecified (not distinguishing noun subclasses) and numerically underspecified (not defining counting units). It can substitute for any conventional Jiangmo Zauzou inanimate classifier in noun phrases requiring a numeral classifier, or it may be selected when the speaker cannot determine which classifier to use. It differs from prototypical classifiers in being unproductive in cardinal constructions but more frequent in other classifier phrases in spontaneous speech. In numeral phrases, though relatively uncommon, *və*<sup>13</sup> functions as a soratal numeral classifier for both count and mass nouns: in (8a) it classifies an abstract noun; in (8b) it classifies a count noun whose original classifier is unknown to the speaker; and in (8c) it quantifies an unspecified amount of mass.

(8) [N NUM *və*<sup>13</sup>]

- a. pɛ<sup>31</sup>            næ<sup>53</sup> *və*<sup>13</sup>  
       disease     two    CLF:versatile  
       ‘two diseases’

- b. ni<sup>55</sup>to<sup>55</sup>      næ<sup>53</sup> və<sup>13</sup>/læ<sup>31</sup>  
mouth      two    CLF:versatile/ CLF:inani  
'two mouths'
- c. yu<sup>33</sup>      næ<sup>53</sup> və<sup>13</sup>  
medicine    two    CLF:versatile  
'two (unspecified units) of medicine'

In addition to numeral phrases, və<sup>13</sup> can also appear in contexts that do not contain a numeral. Numeral classifiers and və<sup>13</sup> show strong functional overlap in the referential values they encode across noun phrases, while differing primarily in their numeric values. Noun phrases classified by sortal classifiers typically pick out specific individuals, as seen in (1)–(4) above. However, noun phrases involving və<sup>13</sup> remain underspecified in number, allowing for singular, plural or mass interpretations. Examples (9) through (11) illustrate the variable numeric values in NPs classified by və<sup>13</sup>.

- (9) [N və<sup>13</sup>] (1/>1/imprecise mass) (Def/Gen)
- a. la<sup>53</sup>yi<sup>31</sup>      və<sup>13</sup>      tu<sup>55</sup> wə<sup>13</sup> tō<sup>33</sup> [SG]  
swing      CLF:versatile    3SG play PROG  
'He is playing on the swing.' (CAL clips)
- b. pe<sup>33</sup>tsɿ<sup>31</sup>      və<sup>13</sup>      tēi<sup>13</sup>      ʔe<sup>31</sup> tēi<sup>13</sup>      kə<sup>13</sup> tēa<sup>53</sup> zə<sup>31</sup> [PL]  
cup      CLF:versatile    frighten    also frighten    drop all    PFV  
'(He) is frightened and makes the cups completely fall over.' (CAL clips)
- c. tu<sup>55</sup> eyi<sup>31</sup>      və<sup>13</sup>      tʰi<sup>33</sup> tʰu<sup>53</sup> tō<sup>55</sup> [mass]  
3SG blood      CLF:versatile    bleed exit    DUR  
'Her blood was coming out.' (Dangerous cases as a doctor)
- (10) [N DEM və<sup>13</sup>] (Def/Gen)(1/>1/imprecise mass)
- a. pə<sup>31</sup>mə<sup>31</sup>      tə<sup>55</sup> və<sup>13</sup>      pʰa<sup>13</sup>      to<sup>13</sup>      ʔə<sup>31</sup> [SG]  
plastic\_paper    DIST CLF:versatile    cover      COMPL    IMP  
'Cover the plastic paper (on the ground).' (Sowing interview 2)
- b. tə<sup>55</sup>-pʰiɛ<sup>33</sup>      ʔa<sup>55</sup>      və<sup>13</sup>      lə<sup>31</sup>nə<sup>55</sup>      ʔa<sup>31</sup>-pe<sup>33</sup>  
NMLZ-CLF:piece    PROX    CLF:versatile    COP      1-PL  
tse<sup>13</sup>mə<sup>33</sup>      næ<sup>53</sup> næ<sup>13</sup> sē<sup>31</sup>    næ<sup>13</sup> ta<sup>55</sup> tsu<sup>31</sup> [PL]  
Spring\_Festival    two    day    three day    fry    eat  
'As to those thin slices (of rice cakes), we will fry and eat them in several days during the Spring Festival.' (How to cook rice cake)

- c.     $\text{ʔa}^{31}$          $\text{ka}^{33}$   $\text{ʔa}^{31}\text{-pe}^{55}$     **( $\text{kə}^{31}\text{eyi}^5$ )**     $\text{ʔa}^{55}$          $\text{və}^{13}$                      $\text{ne}^{33}$   $\text{tə}^{31}$   $\text{va}^{51}$   
PROX    LOC 1-PL        grain        PROX        CLF:versatile    TOP one year  
 $\text{tu}^{55}$   $\text{tə}^{h31}$   $\text{tə}^{31}$   $\text{wu}^{33}$  [mass]  
3SG COP one CLF:season  
‘Here those grains of ours (harvest) one season per year.’ (Rice Harvest)

(11) [REL DEM  $\text{və}^{13}$ ](Def/Gen) (1/>1/imprecise mass)

- a.     $[\text{ʔa}^{55}$          $\text{sue}^{33}$   $\text{mæ}^{55}]_{\text{REL}}$      $\text{tə}^{13}$              $(\text{ni}^{55}\text{mu}^{33})$      $\text{ʔa}^{55}$          $\text{və}^{13}$  [SG]  
PROX    VCL tell        LINK        story        PROX        CLF:versatile  
‘the story that I tell this time’ (New Year Tradition)
- b.     **$[\text{tə}^{55}\text{-pe}^{33}$      $\text{k}^h\text{ẽ}^{53}$   $\text{to}^{53}]_{\text{REL}}$**      $\text{ʔa}^{55}$          $\text{və}^{13}$                      $\text{lə}^{31}\text{ne}^{33}$      $\text{ji}^{13}$      $\text{tsō}^{31}$   
3-PLsew DUR                    PROX        CLF: versatile    COP        one CLF:kind  
 $\text{sæ}^{53}\text{væ}^{53}$   $\text{sɿ}^{13}\text{ne}^{31}$     [PL]  
flower    like  
‘Those [the patterns on the waistband] that are sewed by them are one kind of flower-like (patterns).’ (Zauzou Clothes 2)
- c.     **$[\text{ŋə}^{31}$   $\text{pi}^{13}]_{\text{REL}}$**      $\text{tə}^{55}$   $\text{və}^{13}$                      $\text{ŋu}^{55}$   $\text{ʔa}^{55}$          $\text{næ}^{53}$   $\text{næ}^{13}$   $\text{ja}^{13}$   $\text{tea}^{55}$   
2    give        DIST CLF:versatile    1    PROX        two day SEQ finish  
 $\text{ji}^{55}$          $\text{xə}^{31}$  [mass]  
complete FP  
‘I just finished the thing [wine] you gave me these days.’ (Everyday Conversation 4)

In the bare classifier phrase (9),  $\text{və}^{13}$  is compatible with singular or plural reading, and may also convey an indefinite quantity of mass, as in (9c). A demonstrative word and  $\text{və}^{13}$  may constitute a complex demonstrative pronoun  $\text{ʔa}^{55} \text{və}^{13}$  ‘this/these’ or  $\text{tə}^{55} \text{və}^{13}$  ‘that/those’. A complex demonstrative pronoun occurs alone, as in (10), or serves as the head NP of a relative clause, as in (11). The number interpretation in [N DEM  $\text{və}^{13}$ ] and [REL DEM  $\text{və}^{13}$ ] is ambiguous between singular, plural and mass.

Second,  $\text{və}^{13}$  exhibits patterns similar to a pluralizer. Like the plural quantifier  $\text{tə}^{13}$  that signifies a group of animates in the indefinite ‘one’ phrase [N ‘one’  $\text{tə}^{13}$ ] (as in (7) above),  $\text{və}^{13}$  may appear in the indefinite phrase [N ‘one’  $\text{və}^{13}$ ] to convey a small quantity of count nouns (12a) or mass entities (12b), in addition to the singular reference in (12c).

(12) [N ‘one’  $\text{və}^{13}$ ] (1/>1/small quantity mass) (indef)

- a.     $\text{tea}^{55}\text{tsɿ}^{31}$      $\text{təo}^{55}$          $\text{tə}^{55}$      **$\text{wu}^{53}\text{vuu}^{13}$**      $\text{tə}^{31}$      $\text{və}^{13}$                      $\text{tə}^{55}$      $\text{to}^{53}$   
table        CLF:string up    egg        one CLF: versatile    put    DUR  
‘Some eggs are placed on the table.’ (CAL clips)
- b.     $\text{yɯ}^{33}$          $\text{tə}^{31}$      $\text{və}^{13}$   
medicine one CLF:versatile  
‘some medicine’ (Elicited)

- c.    **la<sup>55</sup>tɔ<sup>33</sup>**      **tə<sup>31</sup>**    **və<sup>13</sup>**                      wɛ<sup>53</sup>    to<sup>53</sup>  
       hole            one    CLF: versatile    dig    COMPL  
       ‘Dug a pit’ (Two brothers)

Moreover, *və<sup>13</sup>* may appear in clause complementation (i.e., [clause DEM +   ]) with the demonstrative *tə<sup>55</sup>*, yielding the complementizer *tə<sup>55</sup>-və<sup>13</sup>*. As shown in (13), *tə<sup>55</sup>-və<sup>13</sup>* functions as an optional complementizer where *və<sup>13</sup>* cannot be substituted by other sortal classifiers.

- (13) [**nə<sup>31</sup>** **ti<sup>33</sup>**                      **xua<sup>13</sup>**            **ve<sup>33</sup>**    **to<sup>53</sup>**]                      (**tə<sup>55</sup>** **və<sup>13</sup>**)                      ʔa<sup>31</sup>-pe<sup>33</sup>    ʔa<sup>31</sup>  
       2PL father            how\_many buy    COMPL    DIST CLF: versatile    1-PL            NEG  
       sa<sup>53</sup>  
       know  
       ‘We don’t know how much money your father-in-law spent to buy them.’ (Cattle transaction)

## 2.2 Grammatical nominalizations in Jiangmo Zauzou

### [1] Numerals

#### [1a] Cardinal numerals

Jiangmo Zauzou employs a base-ten numeral system comprising ten monosyllabic cardinal numerals for values ‘one’ through ‘ten’, with no lexicalized form for ‘zero’. The complete set of basic numerals is presented in Table 2.

Table 2. Jiangmo Zauzou simple cardinal numbers

| Simple Cardinal number          | Gloss |
|---------------------------------|-------|
| tə <sup>31</sup>                | 1     |
| næ <sup>53</sup>                | 2     |
| sɛ̃ <sup>33</sup>               | 3     |
| yi <sup>31</sup>                | 4     |
| ŋu <sup>33</sup>                | 5     |
| k <sup>h</sup> a <sup>53</sup>  | 6     |
| ne <sup>55</sup>                | 7     |
| ia <sup>55</sup>                | 8     |
| ku <sup>33</sup>                | 9     |
| ts <sup>h</sup> e <sup>55</sup> | 10    |

#### [1b] Ordinal numerals

Ordinal numerals basically share roots with cardinal numerals, with the exception of *tə<sup>31</sup>to<sup>55</sup>* ‘the first’ (literally ‘the former one’). The ordinal number marker *-ja<sup>31</sup>* is consistently suffixed to cardinal numerals to derive basic ordinal numbers, as demonstrated in Table 3.

Table 3. Jiangmo Zauzou ordinal numbers

| Basic ordinal number                                | Gloss              |
|-----------------------------------------------------|--------------------|
| tə <sup>31</sup> tə <sup>55</sup> -ja <sup>31</sup> | 1st/the former one |
| nə <sup>53</sup> -ja <sup>31</sup>                  | 2nd                |
| sẽ <sup>33</sup> -ja <sup>31</sup>                  | 3rd                |
| yi <sup>31</sup> -ja <sup>31</sup>                  | 4th                |
| ŋu <sup>33</sup> -ja <sup>31</sup>                  | 5th                |
| k <sup>h</sup> a <sup>53</sup> -ja <sup>31</sup>    | 6th                |
| ne <sup>53</sup> -ja <sup>31</sup>                  | 7th                |
| ja <sup>53</sup> -ja <sup>31</sup>                  | 8th                |
| ku <sup>33</sup> -ja <sup>31</sup>                  | 9th                |
| ts <sup>h</sup> e <sup>55</sup> -ja <sup>31</sup>   | 10th               |

## [2] Demonstratives

Jiangmo Zauzou employs three sets of bound deictic morphemes that constitute demonstratives. Set 1 and Set 2 are native Jiangmo Zauzou deictic morphemes, while Set 3 consists entirely of borrowings from Lisu. Table 4 displays these three sets of morphemes. As bound morphemes, these deictic forms cannot occur independently but must combine with either a numeral classifier, a plural quantifier, or the adverbial morpheme -sã<sup>31</sup> to form demonstrative pronouns (e.g., ʔa<sup>33</sup> ja<sup>55</sup> ‘this (human)’), demonstrative adjectives (e.g., tɛu<sup>33</sup> ʔa<sup>33</sup> ja<sup>55</sup> ‘this person’), or demonstrative adverbs (e.g., ʔa<sup>33</sup> ka<sup>33</sup> ‘here’, ʔa<sup>33</sup>sã<sup>31</sup> ‘(like) this’).

Table 4. Demonstratives in Jiangmo Zauzou

| Deictic distinctions |       | Distance to the speaker       | Native demonstratives (Set 1)       | Native demonstratives (Set 2) | Borrowed demonstratives (Set 3)    |
|----------------------|-------|-------------------------------|-------------------------------------|-------------------------------|------------------------------------|
| PROX (proximal)      |       | close to the speaker          | ʔa <sup>33</sup>                    | ʔa <sup>33</sup>              | thi <sup>55</sup>                  |
| DIST(distal)         | DIST1 | a little far from the speaker | tɛo <sup>33</sup>                   | tə <sup>55</sup>              | wu <sup>13</sup> /ku <sup>55</sup> |
|                      | DIST2 | far from the speaker          | mu <sup>33</sup> /mõ <sup>53</sup>  |                               |                                    |
|                      | DIST3 | very far from the speaker     | tɛo <sup>33</sup> tɛo <sup>33</sup> |                               |                                    |
| HIGH                 |       | higher than the speaker       | ʔa <sup>31</sup> mo <sup>55</sup>   | N/A                           | N/A                                |
| LOW                  |       | lower than the speaker        | ʔa <sup>31</sup> mi <sup>55</sup>   | N/A                           | N/A                                |

Set 1 deictic morphemes show a six-way semantic distinction. They are restricted to forming demonstrative adverbs when combined with the locative classifier ka<sup>33</sup> ‘CLF:place’ (e.g., tɛo<sup>55</sup> ka<sup>33</sup> ‘there (moderately distant from the speaker)’). The place deixis encoded by Set

1 morphemes includes both horizontal and vertical dimensions. Set 2 morphemes exhibit a simpler two-way proximal-distal contrast and primarily function as attributive demonstratives or pronominal demonstratives when used with numeral classifier (e.g.,  $tə^{55} ja^{55}$  ‘that (human)’) or plural quantifiers (e.g.,  $tə^{55} tə^{13}$  ‘that group’). The proximal  $ʔa^{55}$  and distal  $tə^{13}$  can also occur with the classifier  $ka^{33}$  or the adverbial bound morpheme  $-sə̃^{31}$  to form demonstrative adverbs meaning ‘here/there’ and ‘like this/that’. Finally, the Set 3 morphemes are loan words from Lisu, which are commonly found in the texts serving both Set 1 and Set 2 functions.

The two Set 2 deictic morphemes turn out to be the dominant forms in demonstrative pronouns and demonstrative adjectives.

### [3] Genitive noun phrases

Pronominal genitive noun phrases are formed by the possessive pronouns displayed in Table 5. Possessive pronouns in Jiangmo Zauzou essentially mirror the forms of personal pronouns, with pronominal roots encoding person distinctions and plural suffixes marking number distinctions.

Table 5. Jiangmo Zauzou possessive pronouns

|            | Singular                         | plural                              |
|------------|----------------------------------|-------------------------------------|
| 1st person | $\eta u^{33}$ ‘my’               | $\eta u^{33}/ʔa^{31}-pe^{33}$ ‘our’ |
| 2nd person | $nə^{33}/\eta ɔ^{31}$ ‘your’     | $nə^{33}-tə^{13}$ ‘your’            |
| 3rd person | $tə^{33}/təu^{33}$ ‘his/her/its’ | $tə^{55}-pe^{33}$ ‘their’           |

Nominal genitive noun phrases consist of a nominal possessor and a possessed noun. In both pronominal and nominal genitive NPs, the genitive enclitic  $=ze^{55}$  is optionally used after the possessor, indicating a possessive relationship between the two nominals.

In pronominal genitive phrases, such as examples in (14), the pronominal possessor can directly modify the possessed noun without marking or take the optional  $=ze^{55}$ . Similarly, in nominal genitive phrases where the possessor is a noun, as in (15),  $=ze^{55}$  is also optional.

#### (14) Pronominal genitive noun phrase

- a.  $\eta u^{55}=ze^{55}$        $ts^h a^{55} pə^{35}$        $ts^h u \tilde{e}^{55}$   
 1=(GEN)      book      CLF:book  
 ‘my book’
- b.  $tə^{55}-pe^{33}$        $k^h y i^{31} ʔə^{55}$   
 3-PL      dog      CLF:animal  
 ‘their dog’

#### (15) Nominal genitive noun phrase

- a.  $tə^{55}=ze^{55}$        $tə^{55}-pə^{55}$        $tə^h i^{13}$        $ʔa^{31}$        $t^h o^{53}$        $zə^{31}$   
 up;front=GEN      NMLZ-grow      have(ani)      NEG can;capable      FP:assr  
 ‘I am unable to have the previous growth.’



- b.    ɲu<sup>55</sup>    ti<sup>33</sup>                    mō<sup>53</sup>=pe<sup>55</sup>                    nu<sup>31</sup>    kō<sup>55</sup>  
          1       father            PL:inani=PL:ani    cattle CL:animal  
          ‘the cattle of my father’s family’ (Li 2020:229)

Alienable and inalienable possession in Jiangmo Zauzou is not morphologically differentiated. However, inalienable possession involving kinship terms seems to prohibit the presence of =ze<sup>55</sup>. As demonstrated in (16), examples of inalienable possession with kinship terms as the possessed noun do to allow =ze<sup>55</sup> after the pronominal possessor.

(16) Inalienable possession with kinship terms

- a.    ɲu<sup>55</sup>    mæ<sup>55</sup>  
          1       mother  
          ‘my mother’
- b.    nə<sup>55</sup>                    pɔ<sup>13</sup>  
          2PL;2POSS           father  
          ‘your father’
- c.    tə<sup>55</sup>                    zo<sup>31</sup>                    na<sup>53</sup>    ni<sup>55</sup>  
          3PL;3POSS           child                    two    CLF:sibling  
          ‘two sons of his.’

Possessums are often omitted when their referents are anaphoric or discourse-recoverable. In such cases, the possessive postposition =ze<sup>55</sup> becomes obligatory to maintain a possessive interpretation. Example (17a) demonstrates this with a pronominal possessor, while (17b) shows a nominal possessor. In both types of NPs, =ze<sup>55</sup> is obligatory when the possessum is unexpressed.

(17) [Possessor NP ze<sup>31</sup>]:

- a.    [tə<sup>55</sup>-pe<sup>33</sup>=ze<sup>13</sup>]                    na<sup>53</sup>    ɕi<sup>55</sup>    ʔa<sup>31</sup>    tɕ<sup>h</sup>i<sup>13</sup>                    ta<sup>55</sup>  
          3PL;3POSS-PL=GEN                    LOWside    NEG have(inani)                    connect  
          le<sup>35</sup>  
          come.towards\_DC  
          ‘Theirs (pipes) are not connected to the lower part.’
- b.    tsō<sup>33</sup>t<sup>h</sup>i<sup>33</sup>                    kə<sup>55</sup>                    tə<sup>55</sup>    ku<sup>33</sup>=ze<sup>55</sup>  
          general                    LINK                    DIST CLF:stream=GEN  
          ‘the (water) of that main stream.’

The possessum does not have to be a noun. It is often encoded by a bare classifier (18), or a demonstrative phrase (19).

(18) [Possessor CLF]

tə<sup>55</sup>-mu<sup>33</sup>=ze<sup>33</sup>      ku<sup>33</sup>      ne<sup>55</sup>    ʔa<sup>55</sup>-sə̃<sup>33</sup>    pã<sup>55</sup>      xu<sup>33</sup>  
 3PL;3POSS-PL=GEN CLF:stream      TOP PROX-?    release      disappear  
 tə<sup>55</sup>      zo<sup>33</sup>  
 PROG      FP:assr  
 ‘The stream of their family is continuously flowing out.’

(19) [Possessor DEM CLF]

th<sup>31</sup>a<sup>31</sup>ma<sup>13</sup>    ŋu<sup>55</sup> (=ze<sup>55</sup>)      tə<sup>55</sup>      læ<sup>31</sup>  
 hat      1=GEN      DIST      CLF:inani  
 ‘my hat’ (Li 2024)

#### [4] Adjectives

Attributive adjectives in Jiangmo Zauzou modify a noun or a classifier. The latter often serves as the nominal head when the head noun is unexpressed in adnominal noun phrases. Adjectives commonly co-occur with other adnominal modifiers such as numerals and demonstratives within noun phrases. Simple and complex adjectival phrases are distinguished based on the presence of absence of additional modifiers beyond the adjective. The word order between attributive adjectives and their head N/CLF is rigid in simple adjectival phrases but flexible in complex adjective phrases.

In simple adjectival phrases, adjectives typically appear as prenominal modifiers to either nouns or classifiers, as exemplified in (20) and (21). Importantly, the reversed \*[N Adj] and \*[CLF Adj] orders are either prohibited in simple adjectival phrases or is ambiguous between attributive and predicative readings (Li 2020: 533).

(20) [Adj N]

ve<sup>31</sup>jo<sup>55</sup>      pẽ<sup>35</sup>  
 far      place  
 ‘far place(s)/(a/the) far place’

(21) [(N) Adj CLF]

(kæ<sup>13</sup>)      jĩ<sup>31</sup>    ja<sup>55</sup>  
 Younger\_sibling      big    CLF:human  
 ‘a short younger brother/sister’

Attributive adjectives in complex adjectival phrases may occur either prenominally or postnominally. The relative ordering of adjective and CLF/N is determined by the presence or absence of the head noun.

Examples (22) and (23) demonstrate attributive adjectival phrases with an elided head noun, where a single classifier serves as the nominal head. The adjective is a prenominal modifier of the CLF in (22) and a postnominal modifier of CLF in (23). Also note that prenominal attributive adjectives tend to co-occur with linking particles such as tə<sup>31</sup> in (22a) or

the nominalizer =xo<sup>53</sup> in (22b). As Li (2020) observes, monosyllabic prenominal adjectives like kɛ̃<sup>55</sup> and ts<sup>h</sup>u<sup>31</sup> in (22a)–(22b) seem to require the linking particles tə<sup>13</sup>/=xo<sup>53</sup>.

(22) [Adj CLF]

a. [Adj-tə<sup>31</sup>-NUM-CLF]

kɛ̃<sup>55</sup>      tə<sup>31</sup>      na<sup>53</sup>      ja<sup>53</sup>      ʔo<sup>53</sup>      sa<sup>53</sup>  
 skinny      LINK      two      CLF:human      pig      kill  
 ‘The two skinny ones (brothers) are killing the pig.’

b. [Adj xo<sup>53</sup> NUM CLF]

ts<sup>h</sup>u<sup>55</sup> xo<sup>53</sup>      tə<sup>31</sup>      kɔ̃<sup>33</sup>  
 fat      NMLZ      one      CLF:animal  
 ‘a fat one’

c. [Adj DEM CLF]

pɔ̃<sup>35</sup>      ja<sup>55</sup>      tɔ̃<sup>55</sup>-ji<sup>33</sup>      tɔ̃<sup>55</sup>      ja<sup>33</sup>  
 father      CLF:human      NMLZ-big DIST      CLF:human  
 ‘Father is the oldest one.’ (Family Photo)

(23) [CLF Adj]

tɛo<sup>55</sup>      na<sup>53</sup>      ʔɔ̃<sup>55</sup>      ts<sup>h</sup>u<sup>55</sup>  
 DIST1      two      CLF:animal      fat  
 ‘those two fat ones (pigs)’

In examples of (24), there is an overt head noun in addition to CLF. Complex adjectival phrases involving an explicit head noun only display the [N Adj] order.

(24) [N Adj]

a. [DEM-CLF-N-Adj]

tɛo<sup>55</sup>      ʔɔ̃<sup>55</sup>      ʔo<sup>53</sup>      ts<sup>h</sup>u<sup>55</sup>  
 DIST1      CLF:animal      pig      fat  
 ‘that fat pig’

b. [N Adj DEM CLF]

k<sup>h</sup>yi<sup>31</sup> mǎ<sup>31</sup> tɛo<sup>55</sup>      ʔɔ̃<sup>33</sup>  
 dog      old      DIST1      CLF:animal  
 ‘that old dog’

c. [DEM N Adj CLF]

tɛo<sup>55</sup>      ʔo<sup>53</sup>      ts<sup>h</sup>u<sup>55</sup> ʔɔ̃<sup>55</sup>  
 DIST1      pig      fat      CLF:animal  
 ‘that fat pig’

## [5] Verb-based nominalizations

### [5a] Relative clauses

Li (2020) distinguishes two types of relative constructions based on the presence/absence of linking particles (i.e.,  $tə^{13}$  and  $=xo^{53}$ ): 1) zero-marked relatives clause and 2) marked relative clauses. Zero-marked relative clauses are directly joined to the head NP in the matrix without any linking particles. The relative clause consistently appears as a prenominal modifier of the head NP, as illustrated in (25). Marked relative clauses, by contrast, employ the constituency marker  $tə^{31}$  or  $=xo^{53}$  to connect the relative clause with its head noun, as in (26). Marked REL involving  $=xo^{53}$  can be a headless relative construction, in which the head NP can be unexpressed in the matrix clause. See (26c) as an example.

#### (25) Zero-marked REL

|                                  |                                                  |                  |                                  |                   |                                 |
|----------------------------------|--------------------------------------------------|------------------|----------------------------------|-------------------|---------------------------------|
| [ŋu <sup>55</sup> Ø <sub>i</sub> | te <sup>h</sup> a <sup>33</sup> pa <sup>55</sup> | ŋo <sup>55</sup> | pi <sup>13</sup> ] <sub>RC</sub> | [tə <sup>55</sup> | ja <sup>33</sup> ] <sub>i</sub> |
| 1                                | money                                            | lend/borrow      |                                  | APPL; AFF DIST    | CLF:human                       |

‘the person whom I lent money to.’

#### (26) Marked REL

##### a. [REL tə<sup>13</sup> N]

|                   |                                 |                                  |                    |                                      |                  |                                 |
|-------------------|---------------------------------|----------------------------------|--------------------|--------------------------------------|------------------|---------------------------------|
| [ʔa <sup>31</sup> | na <sup>13</sup> Ø <sub>i</sub> | mæ <sup>55</sup> ] <sub>RC</sub> | tə <sup>31</sup> [ | (ni <sup>55</sup> mō <sup>33</sup> ) | ʔa <sup>55</sup> | kō <sup>33</sup> ] <sub>i</sub> |
| PROX              | day                             | tell                             | LINK               | story                                | PROX             | CLF:segment                     |

‘the story that will be told today.’

##### b. [REL xo<sup>53</sup> N]

|                 |                  |                   |                                  |                  |                                    |                                 |
|-----------------|------------------|-------------------|----------------------------------|------------------|------------------------------------|---------------------------------|
| [Ø <sub>i</sub> | wu <sup>53</sup> | tʰu <sup>53</sup> | le <sup>13</sup> ] <sub>RC</sub> | xo <sup>53</sup> | [mi <sup>55</sup> le <sup>35</sup> | və <sup>13</sup> ] <sub>i</sub> |
|                 | weave            | exit              | come                             | NMLZ             | fabrics                            | CLF:versatile                   |

‘the fabrics that are weaved.’

##### c. [REL xo<sup>53</sup>]

|                 |                                  |                  |                |                  |
|-----------------|----------------------------------|------------------|----------------|------------------|
| [Ø <sub>i</sub> | ɕo <sup>13</sup> ] <sub>RC</sub> | xo <sup>53</sup> | Ø <sub>i</sub> | mi <sup>53</sup> |
|                 | rest                             | NMLZ             |                | INT1             |

‘There are many people who rest here.’

The classification of relative clauses can be determined by the position of the head noun relative to the restricting clause (Keenan & Comrie 1977; Keenan 1985; Song 2001). In Jiangmo Zauzou, the head noun in relative constructions can be expressed: (1) in its full form or as a reduced form (i.e., a classifier phrase) in the matrix clause while being absent from the relative clause (i.e., external-headed relative clauses), or (2) in its full form in both the relative clause and matrix clause (i.e., correlative relative clauses). Examples (27) demonstrate external-headed REL where the head NP is expressed exclusively in the matrix clause through either a noun or a classifier phrase. Examples (28) illustrate correlative REL that overtly express the head noun in the relative clause and the matrix clause. Notably, the head noun in correlative matrix clauses shows a strong preference for encoding via classifier phrases (28)

rather than full noun phrases. The classificatory function of the classifier in the coreferential CLP in (28) reveals that the relativized argument represents a person in (28a) and an animal in (28b).

(27) External-headed relative clause/gapping

- a. [Ø<sub>i</sub> ɣo<sup>53</sup> sa<sup>53</sup>]<sub>RC</sub> [ɲu<sup>55</sup> və<sup>31</sup>jo<sup>55</sup> tə<sup>55</sup> ja<sup>31</sup>]<sub>i</sub> nō<sup>33</sup>  
 pig kill 1 older\_brother DIST CLF:human sick  
 ‘My older brother who kills pigs is sick.’

- b. [Ø<sub>i</sub> tɛɛ<sup>13</sup>xũ<sup>55</sup>]<sub>RC</sub> [tə<sup>55</sup> nə<sup>53</sup> ja<sup>53</sup>]<sub>i</sub> xẽ<sup>31</sup> tsu<sup>33</sup> xu<sup>55</sup> zɔ<sup>31</sup>  
 marry DIST two CLF:human collect eat disappear PFV  
 ‘The two (people) who married received (the money).’ (Zauzou marriage)’

(28) Correlative relative clause/non-reduction

- a. [ɲu<sup>55</sup> və<sup>33</sup>jo<sup>55</sup> nō<sup>33</sup>]<sub>RC</sub> [tə<sup>55</sup> ja<sup>33</sup>]<sub>i</sub> ɲu<sup>55</sup> ɲə<sup>55</sup> ɲə<sup>35</sup> zɔ<sup>31</sup>  
 1 older\_brother sick DIST CLF :human 1 see see PFV  
 ‘I saw that older brother of mine who is sick.’

- b. [ɲu<sup>55</sup> xue<sup>33</sup> wu<sup>53</sup> sa<sup>53</sup>]<sub>RC</sub> [tə<sup>55</sup> ɣō<sup>55</sup>]<sub>i</sub> ɣa<sup>31</sup>-pe<sup>33</sup> =ze<sup>55</sup>  
 1 friend chicken kill DIST CLF:animal 1-PL=GEN  
 ɣō<sup>55</sup>  
 CLF:animal  
 ‘The chicken my friend killed is ours.’

In addition to the position of head noun, relative clauses can be classified based on the expression of the head noun within the relative clause. Keenan (1985) and Comrie (1989) identify four relativization strategies according to how the head noun is encoded internally in the relative clause. Jiangmo Zauzou displays both gapping and non-reduction strategies in this regard. Specifically, the external-headed relative clauses in (27) exhibit the gapping strategy, where the head noun is omitted within the relative clause, while the correlative relative clauses in (28) demonstrate the non-reduction strategy through their explicit use of full nouns within the relative clause.

### [5b] Complement clauses

Jiangmo Zauzou employs both the nominalizer =xo<sup>53</sup> and the demonstrative phrase tə<sup>55</sup> və<sup>13</sup> in complementation. They are optionally used to mark complement clauses.

As a nominalizer, =xo<sup>53</sup> is often encliticized to entire clauses as a clausal nominalizer to create a clausal argument (e.g., topic clause) or to complement the main predicate. Examples (29) illustrate complement clauses that have been nominalized by =xo<sup>53</sup>.

(29) Clause-xo<sup>53</sup> as complement

- a. [miã<sup>31</sup>    ŋa<sup>53</sup>            tsu<sup>31</sup> ji<sup>55</sup>    xo<sup>53</sup>]            sue<sup>55</sup> to<sup>53</sup>  
horse    gnaw            eat    go    NMLZ            think DUR  
'(The tiger) wanted to eat horses.'
- b. [sɿ<sup>13</sup>tuɔ<sup>55</sup>k<sup>h</sup>uɛ<sup>53</sup>            tɔ<sup>55</sup>            vɔ<sup>33</sup>            wu<sup>13</sup>]            xo<sup>53</sup>  
over\_10\_Yuan            DIST            CLF:versatile            receive            NMLZ  
xu<sup>55</sup>-ja<sup>33</sup>            sa<sup>53</sup>            tɕi<sup>33</sup>  
WH-CLF:human            know            Q;QUOTE;hearsay  
'Who knows that they only get more than 10 Yuan?'

It is also very common for a clause to be nominalized through combination of the distal demonstrative  $tə^{55}$  and the versatile classifier  $və^{13}$ . The resulting form  $tə^{55} və^{13}$  typically follows a clause as a complementizer, marking the preceding clause as the complement of the main predicate.

(30) Clause- $t\theta^{55}$   $v\theta^{13}$  as complement

- a. [se<sup>55</sup>jã<sup>35</sup>    se<sup>55</sup>jã<sup>35</sup>    ji<sup>33</sup>lã<sup>33</sup>    pã<sup>13</sup>    le<sup>13</sup>    tɔɔ<sup>13</sup>] tɔ<sup>55</sup>    vɔ<sup>13</sup>  
 what    what    stuff    bring come need DIST CLF:versatile  
 tɔ<sup>55</sup>    na<sup>13</sup>    ŋɔ<sup>55</sup>    tɛã<sup>55</sup>  
 DIST    day CAUS:say PURP  
 ‘People will agree upon everything that needs to be brought that day.’  
 (Zauzou Marriage)
- b. [tɛw<sup>33</sup>    tɔ<sup>55</sup>    jẽ<sup>33</sup>    ʔo<sup>31</sup>ji<sup>31</sup>    la<sup>31</sup>    ʔa<sup>31</sup>    ʔo<sup>31</sup>ji<sup>31</sup>] tɔ<sup>55</sup>  
 3SG    DIST house    big    and    NEG big    DIST  
 vɔ<sup>13</sup>    ŋɛ<sup>55</sup>    zo<sup>31</sup>    pe<sup>55</sup>  
 CLF:versatile    see    PFV    FP:assr  
 ‘(We) see if these beehives are big or not big.’

The above two nominalizers are not obligatory and zero-marked complement clauses in Jiangmo Zauzou are common. See (31) as an example.

- (31) [ʔa<sup>55</sup> kɔ̃<sup>55</sup> xe<sup>31</sup> tʰu<sup>53</sup> ji<sup>55</sup> zo<sup>31</sup>] ŋə<sup>55</sup> ja<sup>33</sup>  
 PROX CLF:animal AGT; INS; ABL exit go PFV CAUS;say SEQ  
 ‘He thought about taking this out.’

### 3. DATA

In this section, we present data illustrating the use of classifiers, pluralizers ( $t\partial^{13} / m\partial^{53}$ ), and the non-classifying nominalizer ( $=xo^{53}$ ) for grammatical nominalizations, with a particular focus on their roles in noun-phrase and modification uses in Jiangmo Zauzou. The data primarily derive from the Jiangmo Zauzou language materials compiled by Li & Li (1993), supplemented by additional examples from our native speaker language consultant, Li Shaoen, as well as grammatical illustrations in Li (2020).

We first demonstrate how the nominalization hierarchy proposed by Shibatani (2021:524, 2023:35) applies to classifiers, pluralizers, and non-classifying nominalizers in Jiangmo Zauzou, as depicted in Figures 3 through 6.

Figure 3. Classifier pattern in Jiangmo Zauzou

|          |       |       |                      |       |               |
|----------|-------|-------|----------------------|-------|---------------|
| NP use:  | NUM > | DEM > | <sup>OPT</sup> GEN > | ADJ > | V-based NMLZ  |
| Mod use: | NUM > | DEM > | *GEN >               | ADJ > | ?V-based NMLZ |

Figure 4. Pluralizer pattern in Jiangmo Zauzou

|          |        |       |                        |       |              |
|----------|--------|-------|------------------------|-------|--------------|
| NP use:  | *NUM > | DEM > | <sup>OPT</sup> GEN >   | ADJ > | V-based NMLZ |
| Mod use: | *NUM > | DEM > | ? <sup>OPT</sup> GEN > | ADJ > | V-based NMLZ |

Figure 5. Non-classifying Nominalizer pattern in Jiangmo Zauzou

|          |        |        |        |       |              |
|----------|--------|--------|--------|-------|--------------|
| NP use:  | *NUM > | *DEM > | *GEN > | ADJ > | V-based NMLZ |
| Mod use: | *NUM > | *DEM > | *GEN > | ADJ > | V-based NMLZ |

Figure 6. Non-classifying genitive marker pattern<sup>3</sup> in Jiangmo Zauzou

|          |        |        |                      |        |               |
|----------|--------|--------|----------------------|--------|---------------|
| NP use:  | *NUM > | *DEM > | GEN >                | *ADJ > | *V-based NMLZ |
| Mod use: | *NUM > | *DEM > | <sup>OPT</sup> GEN > | *ADJ > | *V-based NMLZ |

(\*=no marking or impossible; \*OPT=marking applies to a specific/definite referent; ?=no marking or impossible depending on consultants OPT=optional marking; unmarked=obligatory marking)

### 3.1 Classifiers

We first examine whether Jiangmo Zauzou classifiers can form noun phrases with numerals, demonstratives, genitive nouns, adjectives, and verbs, respectively, and then explore whether these noun phrases can be further extended to serve modification functions.

#### [1] Numerals

As discussed in section 2.1.1, Jiangmo Zauzou classifiers function as numeral classifiers, obligatorily co-occurring with numerals in counting constructions. The resulting [NUM CLF] phrase can either function as the head of a noun phrase or as a modifier of a noun. In (32), *ku<sup>31</sup>ja<sup>53</sup>* (comprising the numeral ‘9’ and the human classifier) forms a noun phrase where the classifier itself serves as the nominal head in the absence of an overt noun. In (33), *ne<sup>55</sup>ja<sup>53</sup>* (the numeral 9 and the classifier for human) modifies the preceding noun *mu<sup>31</sup>zo<sup>31</sup>* ‘god’, demonstrating its attributive function.

<sup>3</sup> The genitive marker =*ze<sup>55</sup>* is confined to genitive phrases and hence will not be surveyed in this section. See examples in (14) (15) and (17) above for the use of =*ze<sup>55</sup>* in genitive noun phrases.

NP use

- (32) ʔa<sup>33</sup>      ku<sup>31</sup> ja<sup>53</sup>      tə<sup>55</sup> pɔ<sup>13</sup>      tə<sup>55</sup> mə<sup>33</sup>      nɔ̃<sup>33</sup>  
 PROX      nine CLF:human      3SG father      3SG mother      illness  
 ɕi<sup>55</sup> xo<sup>53</sup>      nɔ̃<sup>55</sup> wu<sup>13</sup>      xe<sup>31</sup> ne<sup>33</sup>.....  
 die NMLZ      hear COMPL      ABL TOP  
 ‘When these nine heard that their parents had died of illness,’ (Li & Li 1993: 62-63)

Mod use

- (33) mu<sup>31</sup> zo<sup>31</sup> ne<sup>55</sup>      ja<sup>53</sup>      ne<sup>31</sup> mɔ̃<sup>13</sup>      so<sup>35</sup>      ji<sup>55</sup>      ŋə<sup>55</sup>  
 god child seven      CLF:human      TOP sun      create      go      say;CAUS  
 ‘The seven gods were sent to create the sun.’ (Li & Li 1993: 61)

## [2] Demonstratives

Classifiers are obligatorily used with demonstratives in Jiangmo Zauzou, occurring in both noun phrase (NP) head positions and modification constructions, as shown in examples (34)–(36). Crucially, when a CLF forms a constituent with a demonstrative, the head noun must precede the [DEM CLF] sequence, as illustrated by the grammatical patterns in (35a) and (35b). Conversely, the alternative [DEM CLF N] sequence, shown in the ungrammatical examples (35a’) and (35b’), is prohibited in the language.

- (35) a. k<sup>h</sup>yi<sup>31</sup> [ʔa<sup>33</sup> ʔɔ̃<sup>31</sup>]  
 Dog PROX CLF:animal  
 ‘this dog’  
 a’ \* [ʔa<sup>33</sup> ʔɔ̃<sup>31</sup>]      k<sup>h</sup>yi<sup>31</sup>  
 PROX CLF:animal      dog  
 ‘this dog’  
 b. k<sup>h</sup>yi<sup>31</sup> [wu<sup>35</sup> ʔɔ̃<sup>31</sup>]  
 dog DIST1 CLF:animal  
 ‘that dog’  
 b’. \* [wu<sup>35</sup> ʔɔ̃<sup>31</sup>]      k<sup>h</sup>yi<sup>31</sup>  
 DIST1 CLF:animal      dog  
 ‘that dog’ (Examples provided by Li Shaoen)

A demonstrative may precede a noun without an intervening classifier only when the noun is a classified noun, depending on the type of demonstrative. For addressee-proximal or distal demonstratives (referencing either speaker or addressee), the head noun may occur either before or after the demonstrative, as evidenced by (35b) and (36a). However, with speaker-proximal demonstratives, only the pre-demonstrative noun position is permitted, as shown in (35a). Our consultant Li Shaoen confirms that the alternative ordering in (36b) is



ungrammatical. Across these patterns, the structure with the head noun preceding the demonstrative is the unmarked option, while the structure with the demonstrative preceding the head noun is marked.

- (36) a.   wu<sup>35</sup> [k<sup>hyi</sup>i<sup>31</sup>       ʔō<sup>31</sup>]  
           DIST1       dog               CLF:animal  
           ‘that dog’
- b.   \*ʔa<sup>33</sup> [k<sup>hyi</sup>i<sup>31</sup>       ʔō<sup>31</sup>]  
           PROX       dog               CLF:animal  
           ‘this dog’ (Examples provided by Li Shaoen)

### [3] Genitive noun phrases

The possessive postposition =ze<sup>55</sup> is required to convey a possessive interpretation when the possessum is implicit (§2.2). In such NP-use constructions, classifiers may optionally co-occur with this genitive marker, as shown in (37). Crucially, the possessive marker =ze<sup>55</sup> becomes optional when the possessum is overtly expressed. See (14) above. However, in modification use, where the genitive marker modifies a noun, classifier co-occurrence is prohibited, as demonstrated by the ungrammaticality of (38).

NP use

- (37) tɛ<sup>hɛ</sup>53 ʔa<sup>33</sup>       ʔō<sup>31</sup>               nɛ<sup>31</sup> [[ŋu<sup>33</sup>-pe<sup>33</sup>=ze<sup>33</sup>]       (ʔō<sup>31</sup>)]  
       goat PROX       CLF:animal       TOP 1-PL [EXCL]= GEN CLF:animal  
       ‘This goat is ours.’ (Example provided by Li Shaoen)

- (38) [ŋu<sup>33</sup>-pe<sup>33</sup>=ze<sup>33</sup>       (\*ʔō<sup>31</sup>)]       tɛ<sup>hɛ</sup>53 ʔō<sup>31</sup>  
       1-PL[EXCL]= GEN   CLF:animal       goat CLF:animal  
       ‘our goat’ (Example provided by Li Shaoen)

### [4] Adjectives

Classifiers may co-occur with adjectives in both NP use and modification use, as evidenced by examples (39) and (40). While classifier is obligatory in its NP use, as in (39), it remains optional in modification uses. As established in the discussion of Jiangmo Zauzou noun phrase structure in Section 1, adjectives may appear in either prenominal or post-nominal position. In constructions with an overt head noun, postnominal adjectives, such as ŋā<sup>55</sup> in (40a), may be followed by a numeral classifier. By contrast, prenominal adjectives do not form classifier-containing constituents with the noun, as illustrated by (40b). Rather, prenominal adjectives tend to directly modify nouns without a classifier, as in (40c).

NP use

- (39) ŋā<sup>55</sup>       ʔō<sup>33</sup>               nɛ<sup>31</sup> ŋu<sup>33</sup>=ze<sup>33</sup>.  
       small       CLF:animal       TOP 1SG=GEN  
       ‘The small one (dog, cat, pig, etc.) is mine.’ (Example provided by Li Shaoen)

Mod use

- (40) a.  $k^{h}yi^{31}$   $[\eta\tilde{a}^{35}$   $\gamma\delta^{33}]$   
 dog small CLF:animal  
 ‘small dog’ (Example provided by Li Shaoen)
- b.  $*[\eta\tilde{a}^{35}$   $\gamma\delta^{33}]$   $k^{h}yi^{31}$   
 small CLF:animal dog (Shibatani 2021: 488)
- c.  $ve^{31}jo^{55}$   $p\tilde{e}^{35}$   
 far place  
 ‘far place(s)/(a/the) far place’ (Example provided by Li Yu)

## [5] Verb-based nominalizations

### [5a] Relative clauses

Jiangmo Zauzou zero-marked relative clauses display two relativizing strategies: external-headed relative clauses and correlative relative clauses (see 2.2). In external-headed relative constructions, the head NP may surface as a noun or as a classifier phrase. When the head NP is encoded as a classifier (i.e., NP use), it is required with the relative clause, as shown in example (41). When the head NP is a noun (i.e., modification use), a classifier is disfavored. However, the grammaticality of classifier use in modification uses, as in (42), remains indeterminate at present. While our consultant Li Shaoen rejects such constructions, extensive fieldwork by co-author Li Yu suggests their potential acceptability<sup>4</sup>. Resolving this discrepancy will require additional native speaker judgments. Crucially, the consensus on the grammaticality of classifier use in noun phrase uses and the variation in modification uses align with Shibatani’s theory, specifically, the proposed expansion from noun phrase usage to modification usage. For correlative relative clauses, where the head NP appears internal to the relative clause, only NP uses of classifiers are attested. A classifier is required to nominalize a relative clause into a nominal referent, as in (43). Notably, all attested examples of this construction involve demonstrative phrases as head NPs.

NP use: external-headed relative clause

- (41)  $[\gamma a^{33}ni^{33}$   $\epsilon i^{55}$   $\gamma\delta^{33}]$   $ne^{31}$   $\eta u^{33}=ze^{33}$ .  
 yesterday die CLF:animal TOP 1SG=GEN  
 ‘The animal which died yesterday is mine.’ (Example provided by Li Shaoen)

Mod use: external-headed relative clause

- (42)  $??u^{53}$   $p^{h}\tilde{o}^{35}$   $[\gamma a^{33}ni^{33}$   $\epsilon i^{55}$   $\gamma\delta^{33}]$   
 chicken M yesterday die CLF:animal  
 ‘The cock which died yesterday’ (Example provided by Li Shaoen)

<sup>4</sup> According to Li Yu, the reversed order  $[\gamma a^{33}ni^{33}\epsilon i^{55}\gamma\delta^{33}]\gamma u^{53}p^{h}\tilde{o}^{35}$  is ungrammatical.

NP use: correlative relative clause

- (43) [ŋu<sup>55</sup> və<sup>33</sup>jo<sup>55</sup>i nō<sup>33</sup>]<sub>RC</sub> [tə<sup>55</sup> ja<sup>33</sup>]<sub>I</sub> ŋu<sup>55</sup> ŋæ<sup>55</sup> ŋə<sup>35</sup> zə<sup>31</sup>  
 I older\_brother sick DIST CLF:human I see see PFV  
 ‘I saw that older brother of mine who is sick.’ (Example provided by Li Yu)

### [5b] Complement clauses

Classifiers can form a complement clause, as shown in (44). It should be noted that the versatile classifier və<sup>13</sup> is typically used with the distal demonstrative tə<sup>55</sup> in complement clauses. In this construction, classifiers are optionally used to generate noun phrases.

NP use

- (44) [[ʔə<sup>31</sup>pə<sup>13</sup> ka<sup>33</sup> ŋu<sup>31</sup> ni<sup>33</sup>] tə<sup>55</sup> və<sup>13</sup>] xo<sup>55</sup> ja<sup>33</sup> le<sup>35</sup> sa<sup>53</sup>.  
 Lake LOC fish have(ani) DIST CLF:versatile anybody also know  
 ‘Anybody knows that there are fish in the lake.’ (Example provided by Li Shaoen)

- (45) [ʔa<sup>55</sup> kə<sup>55</sup> xe<sup>31</sup> tʰu<sup>53</sup> ji<sup>55</sup> zə<sup>31</sup>] ŋə<sup>55</sup> ja<sup>33</sup>  
 PROX CLF:animal AGT; INS; ABL exit go PFV CAUS;say SEQ  
 ‘He thought about taking this out.’

## 3.2 Pluralizers

As explained in section 2.1.2, Jiangmo Zauzou employs two distinct pluralizers: tə<sup>13</sup> for animate nouns and mō<sup>53</sup> for inanimate nouns. The two pluralizers express a plural quantity in contexts where classifiers would usually express singular reference, as demonstrated by the alternation between sortal classifiers and pluralizers in bare classifier phrases and demonstrative phrases in (46). Despite that pluralizers can substitute classifiers in numerous contexts to signal different number interpretations, they are generally incompatible with co-occurring classifiers, with the sole exception of lexical nominalizations, as evidenced by (46d’).

- |         |                                             |                                |     |                                          |                              |
|---------|---------------------------------------------|--------------------------------|-----|------------------------------------------|------------------------------|
| (46) a. | təu <sup>33</sup><br>person<br>‘the person’ | ja <sup>33</sup><br>CLF:human  | a’. | təu <sup>33</sup><br>person<br>‘persons’ | tə <sup>13</sup><br>PL:ani   |
| b.      | kʰyi <sup>31</sup><br>dog<br>‘the dog’      | ʔə <sup>33</sup><br>CLF:animal | b’. | kʰyi <sup>31</sup><br>dog<br>‘dogs’      | tə <sup>13</sup><br>PL:ani   |
| c.      | jẽ <sup>33</sup><br>house<br>‘the house’    | læ <sup>33</sup><br>CLF:inani  | c’. | jẽ <sup>33</sup><br>house<br>‘houses’    | mō <sup>53</sup><br>PL:inani |

- |                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>d.    sɛ<sup>53</sup>-tsɛ̃<sup>33</sup>      tsɛ̃<sup>33</sup><br/>             tree-CL:tree      CLF:tree<br/>             ‘the tree’</p> <p>e.    [ʔa<sup>31</sup> so<sup>35</sup>] tə<sup>55</sup>    və<sup>13</sup><br/>             food cook DIST CLF:versatile<br/>             ‘to cook food’</p> | <p>d’.    sɛ<sup>53</sup>-tsɛ̃<sup>33</sup>      mō<sup>53</sup><br/>             tree-CLF:tree      PL:inani<br/>             ‘trees’</p> <p>e’.    [ʔa<sup>31</sup> so<sup>35</sup>] tə<sup>55</sup>    mō<sup>53</sup><br/>             food cook DIST PL:inani<br/>             ‘to cook food (several times)’<br/>             (Example provided by Li Shaoen)</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## [1] Numerals

Pluralizers are incompatible with numerals. The modification use examples (6) and (7) in section 2.1.2 are incorrect attestations. Consequently, NP use is not acceptable for nominalization with pluralizers, as shown in (47a’).

NP use

- |                                                                                                                                                               |                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(47) a.    ʔa<sup>33</sup>      ku<sup>31</sup> ja<sup>53</sup><br/>                 PROX      nine CLF:human<br/>                 ‘these nine (gods)’</p> | <p>a’.    *ʔa<sup>33</sup>      ku<sup>31</sup> tə<sup>13</sup><br/>                 PROX      nine PL:ani<br/>                 ‘these nine (gods)’</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|

In modification use, it is noteworthy that the animate pluralizer *tə13* exhibits semantic similarity to group classifiers like *ma<sup>33</sup>*. Crucially, they differ distributionally: the former cannot co-occur with numerals whereas the latter permits it, as shown in (48a)–(48b).

MOD use

- |                                                                                                                                                                                  |                                                                                                                                                             |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(48) a.    *nu<sup>31</sup>      sɛ̃<sup>33</sup>    tə<sup>13</sup><br/>                 cattle      three PL:ani<br/>                 Intended: ‘three heads of cattle’</p> | <p>b.    nu<sup>31</sup>      sɛ̃<sup>33</sup>    ma<sup>33</sup><br/>             cattle      three CLF:group<br/>             ‘three herds of cattle’</p> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|

## [2] Demonstratives

Pluralizers can co-occur with demonstratives in both NP and Mod uses, as shown in examples (49) and (50). Both animate and inanimate pluralizers are obligatory in the two contexts.

NP use

- (49) a.    tɕo<sup>55</sup>      mō<sup>53</sup>      nɛ<sup>31</sup>    zɔ<sup>33</sup>zɔ<sup>31</sup>    jɛ̃<sup>33</sup>      mō<sup>53</sup>.  
           DIST1      PL:inani    TOP Zauzou    house      PL:inani  
           ‘Those things in the distance are the houses of the Zauzou people.’

- b. **tɛo<sup>33</sup>**      **tə<sup>13</sup>**      nɛ<sup>31</sup>    zɔ<sup>33</sup>    zɔ<sup>31</sup>    tə<sup>13</sup>  
DIST1      PL:ani      TOP    Zauzou      PL:ani  
‘Those people are Zauzou people.’

(Example provided by Li Shaoen)

Mod use

- (50) a. jɛ̃<sup>33</sup>      mi<sup>55</sup>    mɔ̃<sup>53</sup>  
house      LOW PL:inani  
‘Those houses down there’

- b. tɛu<sup>33</sup>      ʔa<sup>33</sup>      tə<sup>13</sup>  
person      PROX      PL:ani  
‘these persons’

(Example provided by Li Shaoen)

### [3] Genitive noun phrases

Pluralizers can optionally co-occur with the genitive marker =ze<sup>33</sup> in its NP use, as shown in examples (51).

NP use

- (51) a. **tu<sup>55</sup> =ze<sup>33</sup>**      **mɔ̃<sup>53</sup>**      nɛ<sup>31</sup>    tsɿ<sup>31</sup>      ja<sup>53</sup>  
3SG=GEN      PL:inani      TOP    officer      CLF:human  
ʔo<sup>55</sup>ʔue<sup>31</sup>    tɛɔ<sup>31</sup>      thie<sup>31</sup>      tə<sup>33</sup>    nu<sup>35</sup>      pe<sup>33</sup>.  
pillow      CLF: pillow      under      put    FP:assr      FP:assr  
‘As to the officer’s belongings (=clothes), (the thief) put them under his pillow.’  
(Li & Li 1993: 116)

- b. tɛo<sup>33</sup>      sɿ<sup>33</sup>ŋã<sup>55</sup>      tə<sup>13</sup>      nɛ<sup>31</sup>    **ŋu<sup>33</sup>-pe<sup>33</sup>=ze<sup>33</sup>**      **(tə<sup>13</sup>)**  
DIST1      child      PL:ani      TOP    1-PL [EXCL]=GEN      PL:ani  
‘Those children are ours.’ (Example provided by Li Shaoen)

Regarding the modification use of pluralizers in genitive noun phrases with =ze<sup>33</sup>, their grammaticality remains unresolved. In (52), our consultant Li Shaoen judges pluralizer inclusion as ungrammatical, whereas the co-author Li Yu considers it grammatical. Parallel to the (42) case, systematic grammaticality judgments from additional Jiangmo Zauzou speakers are required. However, these variable judgments remain theoretically compatible with Shibatani’s theory of functional extension from NP use to Mod use.

Mod. use

- (52) a. [ŋu<sup>33</sup>-pe<sup>33</sup>=ze<sup>33</sup>      (?mɔ̃<sup>53</sup>) ]    mi<sup>33</sup>      mɔ̃<sup>53</sup>  
1-PL [EXCL]=GEN    PL:inani    clothes      PL:inani  
‘our clothes’ (Example provided by Li Shaoen)

- b.     $\eta u^{-33}pe^{33}=ze^{33}$              $(tə^{13})$              $nu^{31}$              $tə^{13}$   
       1-PL [EXCL]=GEN    PL:ani            cow/bull    PL:ani  
       ‘our cattle’

#### [4] Adjectives

Pluralizers may co-occur with adjectives in both NP and Mod uses, as evidenced by examples (53) and (54). Importantly, in modification use (54), pluralizers are obligatory when number specification is required, whereas bare [N Adj] combinations denote kind reference. Mirroring classifiers, the reversed [Adj N] order appears to prohibit pluralizer insertion.

NP use

- (53) a.     $ts^{h_1}_{31}$              $m\tilde{o}^{53}$              $ne^{31}$   $ts\phi^{33}$   $k^{h_5}_{35}$      $ja^{33}=ze^{33}$              $mi^{33}$   
           old            PL:inani    TOP old person    CLF:human=GEN    clothes  
            $m\tilde{o}^{53}$   
           PL:inani  
           ‘The old ones are the old person’s clothes.’ (Example provided by Li Shaoen)

- b.     $?a^{31}$              $lo^{35}$   $tə^{13}$              $ne^{31}$   $\eta u^{33}$   $tea^{53}$              $tə^{13}$   
       NEG            good PL:ani            TOP 1SG friend            PL:ani  
       ‘No good ones are my friends.’ (Example provided by Li Shaoen)

Mod use

- (54) a.     $mi^{33}$              $[ts^{h_1}_{31} m\tilde{o}^{53}]$   
           clothes    old    PL:inani  
           ‘old clothes’ (Example provided by Li Shaoen)
- b.     $teu^{33}$              $[?a^{31} lo^{35} tə^{13}]$   
           person    NEG good PL:inani  
           ‘bad persons’ (Example provided by Li Shaoen)

#### [5] Verb-based nominalizations

Pluralizers may co-occur with verbs in NP use, as exemplifies in (55). In (55b), the demonstrative  $tə^{55}$  preceding the pluralizer  $m\tilde{o}^{53}$  can be omitted. Both cases in (55) are external-headed relative clauses. For correlative relative clauses, pluralizer occurrence is similarly obligatory, as demonstrated by (56).

NP use: external-headed relative clause

- (55) a.     $ci^{55}$   $tə^{13}$              $ne^{31}$   $\eta\phi^{31}=ze^{33}$      $?u^{53}$              $tə^{13}$   
           die    PL:ani            TOP 2SG=GEN chicken    PL:ani  
           ‘The dead (animate) things are your chickens.’

- b. **va<sup>53</sup> (tə<sup>55</sup>) mō<sup>53</sup>**      ne<sup>31</sup> mi<sup>33</sup>      tsh<sup>h</sup><sub>1</sub><sup>31</sup> mi<sup>33</sup>      tẽ<sup>33</sup>,  
wear DIST PL:inani TOP clothes old clothes torn  
'Wearing things (clothes) are old and torn clothes.' (Li & Li 1993: 66)

NP use: correlative relative clause<sup>5</sup>

- (56) [tɛyẽ<sup>55</sup>kã<sup>55</sup> tə<sup>13</sup>]<sub>i</sub>      ʔa<sup>31</sup>-pe<sup>33</sup>      ʔɔ<sup>31</sup> ɲa<sup>53</sup> [tə<sup>55</sup> tə<sup>13</sup>]<sub>i</sub>  
mosquito PL:ani 1-PL PAT bite DIST PL:ani  
'the mosquitos that bite us.' (Example provided by Li Yu)

Speaker judgments again diverge regarding the acceptability of modification use in verb-based nominalizations. For instance, our consultant Li Shaoen rejects (57a) as ungrammatical, while Li Yu accepts it. Notably, Li Shaoen's own compiled materials nevertheless contain attested instances of modification use of verb-based nominalizations, as seen in (57b).

Mod use

- (57) a. ʔʔu<sup>53</sup>      [ɛi<sup>55</sup> tə<sup>13</sup>]  
chicken die PL:ani  
'the dead chickens' (Example provided by Li Shaoen)
- b. **xu<sup>31</sup>wa<sup>33</sup> [lyɛ<sup>33</sup> mō<sup>53</sup>]**      ʔẽ<sup>55</sup>      ya<sup>53</sup>      tɛɛ<sup>53</sup>  
bone carry PL:inani pick\_up collect all  
'(She) picked up and collect all the bones to carry.' (Li & Li 1993: 76)

### 3.3 Nominalizer

The nominalizing enclitic =xo<sup>53</sup> is the most productive nominalizer in Jiangmo Zauzou functioning in both participant and clausal nominalizations (Li 2020: 272). The enclitic =xo<sup>53</sup> is often used in headless relative clauses (i.e., lacking an overt head NP in the matrix clause) to nominalize participants. In (58a), =xo<sup>53</sup> nominalizes the preceding clause as the actor of 'ride', while in (58b) it derives the undergoer of 'eat'. =xo<sup>53</sup> is also used as a particle that links the relative clause with the head NP in the matrix. Example (59a) demonstrates its use in an external-headed relative clause. =xo<sup>53</sup> can be used in clausal nominalization as well. (59b) is an example where it is used in a complement clause.

- (58) a. **mɔ<sup>13</sup>tɕo<sup>53</sup>**      **tɛi<sup>55</sup>=xo<sup>53</sup>**      ʔɛ<sup>33</sup>, tə<sup>33</sup> tɛɛ<sup>55</sup>      lə<sup>31</sup>.  
Motorcycle ride=NMLZ; LINK also one CLF:pair FP:assr  
'There are a pair of people who are riding a motorcycle.' (Li 2020: 272)
- b. ʔə<sup>31</sup>tə<sup>55</sup>      və<sup>35</sup>      ne<sup>55</sup> [**ʔo<sup>53</sup> tə<sup>13</sup>**      **tsu<sup>31</sup>]=xo<sup>53</sup>**]  
maize CLF:versatile TOP pig PL:ani eat NMLZ; LINK  
'Maize is what pigs eat.' (Li 2020: 272)

<sup>5</sup> Like classifiers, only NP use is available for pluralizers.

- (59) a. [wu<sup>53</sup> tʰu<sup>53</sup> le<sup>13</sup>]<sub>i</sub> =xo<sup>53</sup> [mi<sup>55</sup>le<sup>35</sup> və<sup>13</sup>]<sub>i</sub>  
 weave exit come;towards\_DC=NMLZ; LINK fabrics CLF:versatile  
 ‘the fabrics that are weaved’ (Li 2020: 272)
- b. miā<sup>31</sup> ŋa<sup>53</sup> tsu<sup>31</sup> ji<sup>55</sup>=xo<sup>53</sup> sue<sup>55</sup> to<sup>53</sup>  
 horse gnaw eat go=NMLZ; LINK think DUR  
 ‘(The tiger) wanted to eat horses.’ (Li 2020: 274)

=xo<sup>53</sup> is the non-classifying nominalizer. In (60a), =xo<sup>53</sup> denotes grass, in (59b), it denotes an animal, and in (60c), it denotes a human.

- (60) a. ŋo<sup>33</sup> xo<sup>53</sup> tə<sup>31</sup> po<sup>13</sup>  
 blue NMLZ one CLF:clump  
 ‘a bush of blue ones’ (Li & Li 1993: 67)
- b. ʔa<sup>33</sup> ni<sup>33</sup> ei<sup>55</sup> xo<sup>53</sup> ne<sup>31</sup> ŋu<sup>33</sup>=ze<sup>33</sup>.  
 yesterday die NMLZ TOP 1SG=GEN  
 ‘The one (dog, chicken, pig etc.) that died yesterday is mine.’  
 (Example provided by Li Shaoen)
- c. na<sup>33</sup>te<sup>35</sup> la<sup>53</sup> xo<sup>53</sup> ʔa<sup>31</sup> ni<sup>33</sup>.  
 heart enough NMLZ NEG have(ani)  
 ‘There is no one satisfied.’ (Li & Li 1993: 143)

### [Adjectives]

When classifiers or pluralizers are absent in attributive adjective constructions, the nominalizer =xo<sup>53</sup> becomes obligatory with adjectives in both NP and Mod uses, as shown in (61) and (62). The semantic distinction between classifying and non-classifying nominalizers in adjectival phrases manifests such that classifying forms yields specific readings while their non-classifying counterparts produce nonspecific interpretations.

NP use

- (61) tsue<sup>55</sup>kuə<sup>31</sup>və<sup>31</sup> tə<sup>55</sup> ji<sup>31</sup> xo<sup>55</sup> ʔa<sup>31</sup> te<sup>h</sup>i<sup>13</sup>  
 Cun\_pipe CLF:versatile up;front big NMLZ;LINK NEG have(inani)  
 ‘There is no such thing that is bigger than the cun pipe.’ (Repair Pipe)  
 (Li 2020: 272-273)

Mod use

- (62) a. ja<sup>53</sup>yo<sup>35</sup> xo<sup>53</sup> sɿ<sup>33</sup>mi<sup>13</sup> zo<sup>31</sup>  
 beautiful NMLZ human.F child  
 ‘a beautiful girl’ (Example provided by Li Shaoen)
- b. \*sɿ<sup>33</sup>mi<sup>13</sup> zo<sup>31</sup> ja<sup>53</sup>yo<sup>35</sup> xo<sup>53</sup>  
 human.F child beautiful NMLZ



Note that in modification uses, adjectives nominalized by  $=xo^{53}$  come before the head noun (62a) and cannot follow it (62b). When compared with examples of adjectives nominalized by classifiers in (40) and by pluralizers in (54), this reveals a systematic word order reversal: classifiers/pluralizers pattern with [N Adj] order, while the non-classifying nominalizer  $=xo^{53}$  requires [Adj N] order.

### [Verb-based]

The nominalizer  $=xo^{53}$  can co-occur with verbs in both NP and Mod uses, as shown in (62) and (63a). Similar to adjectives in modification, verbs nominalized by  $=xo^{53}$  can only modify head nouns when preceding them.

NP use

- (62)  $\text{ʔa}^{33} \text{ni}^{33} \quad \text{ei}^{55} \quad \text{xo}^{53} \quad \text{ne}^{31} \quad \text{ŋu}^{33}=\text{ze}^{33}$ .  
yesterday die NMLZ TOP 1SG=GEN

‘The one (dog, chicken, pig etc.) that died yesterday is mine.’

(Example provided by Li Shaoen)

Mod use

- (63) a.  $[\text{ei}^{55} \text{xo}^{53}] \quad \text{ʔu}^{53}\text{p}^{\text{h}}\text{õ}^{35}$   
die NMLZ chicken.M  
‘the cock which died’ (Example provided by Li Shaoen)

- b.  $*\text{ʔu}^{53}\text{p}^{\text{h}}\text{õ}^{35} \quad [\text{ei}^{55} \text{xo}^{53}]$   
chicken-MSCL died NMLZ  
‘the cock which died’

Note that numerals, demonstratives, and genitive nouns cannot be attached to  $=xo^{53}$ , as shown in examples (64).

- (64) a.  $*\text{sẽ}^{33} \text{xo}^{53} \quad \text{nẽ}^{33} \quad \text{ŋu}^{33} \quad \text{zo}^{31} \quad \text{ja}^{53}$   
three NMLZ TOP 1SG child CLF:human  
‘The threes are my children.’
- b.  $*\text{ʔa}^{33} \quad \text{xo}^{53} \quad \text{nẽ}^{33} \quad \text{ŋu}^{33} \quad \text{zo}^{31} \quad \text{ja}^{53}$   
PROX NMLZ TOP 1SG child CLF:human  
‘This is my child.’
- c.  $*\text{ŋu}^{33}=\text{ze}^{33} \quad \text{xo}^{53} \quad \text{nẽ}^{33} \quad \text{ji}^{31} \quad \text{læ}^{31}$   
1SG=GEN NMLZ TOP big CLF:inani  
‘Mine is big.’

## 4. ANALYSIS AND DISCUSSION

As Figure 1 shows, Jiangmo Zauzou classifiers occupy the highest position in the hierarchical distribution and functional model of grammatical nominalization. Our data

substantially corroborate Shibatani's (2023) hierarchy. Regarding Jiangmo Zauzou pluralizers, their distributional behavior parallels that of classifiers except in numeral (NUM) nominalization, as illustrated in Figure 2. Conversely, Jiangmo Zauzou non-classifying nominalizer may have an inverse hierarchical expansion pattern, progressing from V-based nominalization to ADJ nominalization. From the viewpoints of nominalization, classifiers and pluralizers, which share both classifying functions and similar expansion patterns, constitute a unified category. However, Jiangmo Zauzou pluralizers lack NUM nominalization, which represents the most basic nominalization type for classifiers. This distinction raises theoretical questions about whether pluralizers should be classified within the same group as classifiers.

In addition, the variable behavior of GEN nominalization across classifiers, pluralizers and non-classifying nominalizers warrants careful examination, as evidenced by Figures 3, 4, and 5. The classifier pattern shows optional genitive marking in NP use but prohibits it in Mod use. Pluralizers permit optional genitive marking in both NP and Mod uses. The non-classifying nominalizer =xo<sup>53</sup> disallows genitive marking. At present, the underlying causes of this distribution remain unresolved. As previously discussed, possessive constructions in Jiangmo Zauzou are rather complicated, featuring both possessive pronouns and an optional genitive marker =ze<sup>31</sup>. Three distinct configurations exist: contexts requiring =ze<sup>31</sup>, those permitting it optionally, and those prohibiting it categorically (i.e., notably kinship terms). This complexity directly impacts our understanding of nominalization patterns, necessitating further syntactic and semantic analysis of how possessive constructions interact with classifier, pluralizer, and non-classifying nominalizations. Crucially, genitive nominalization displays unique properties compared to other nominalization types in the language.

## 5. CONCLUSION

The present analysis has demonstrated that the distribution and extension of nominalizations in Jiangmo Zauzou follows a clear hierarchy: classifiers exhibit the widest scope, followed by pluralizers, and finally non-classifying nominalizers. This pattern suggests a potential correlation between the classificatory strength of different types of nominalizers and their functional range in nominalization. Furthermore, non-classifying nominalizers show a distinct behavior by not participating in verbal and adjectival nominalizations, which raises the hypothesis that their expansion direction is opposite to that of classifiers and pluralizers. Future research should test whether these patterns hold crosslinguistically, both in neighboring languages and in typologically diverse language families. Additionally, consultant judgments reveal variability in noun-phrase and modification uses, particularly regarding verbal nominalization by classifiers and genitive nominalization by pluralizers. To resolve these discrepancies, further data collection with a larger group of consultants is necessary. Such research will help identify systematic patterns in these variations, providing critical evidence for evaluating Shibatani's theory.

## ABBREVIATIONS

|       |                               |
|-------|-------------------------------|
| 1     | first person                  |
| 2     | second person                 |
| 3     | third person                  |
| ABL   | ablative                      |
| ADJ   | adjective                     |
| AFF   | affective                     |
| AGT   | agent                         |
| ani   | animate                       |
| APPL  | applicative                   |
| assr  | assertive                     |
| CAUS  | causative                     |
| CLF   | classifier                    |
| COMPL | completive                    |
| COP   | copula                        |
| DC    | deictic center                |
| DEM   | demonstrative                 |
| DEF   | definite                      |
| DIST  | distal                        |
| DUR   | durative                      |
| F     | feminine                      |
| FP    | final particle                |
| EXCL  | exclusive                     |
| GEN   | genitive                      |
| IMP   | imperative                    |
| inani | inanimate                     |
| INS   | instrumental                  |
| INT   | intensifier                   |
| LOC   | locative                      |
| LOW   | lower than the deictic center |
| LINK  | linker                        |
| N     | nominal                       |
| NP    | noun phrase                   |
| M     | masculine                     |
| Mod   | modification                  |
| NEG   | negative                      |
| NMLZ  | nominalizer                   |
| NUM   | numeral                       |
| PFV   | perfective                    |
| PL    | plural                        |
| POSS  | possessive                    |
| PROG  | progressive                   |
| PROX  | proximal                      |

|       |                 |
|-------|-----------------|
| PURP  | purposive       |
| Q     | question        |
| QUOTE | quotative       |
| RC    | relative clause |
| REL   | relative clause |
| SEQ   | sequential      |
| SG    | singular        |
| TOP   | topic           |
| V     | verb            |
| VCL   | verb classifier |

## ACKNOWLEDGMENTS

We would like to express our deepest gratitude to our Jiangmo Zauzou consultant, Li Shaoen. The present study has been supported by grants from the Japan Society for the Promotion of Science (KAKENHI Project IDs: 22H00659 and 23K21931, ‘Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers’, PI: Kazuhiro Kawachi) and from the National Institute for Japanese Language and Linguistics (‘Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective’, PI: Prashant Pardeshi).

## REFERENCES

- Li, Shaoen, and Zhien Li. 1993. *Nuzu ruorou yuyan ziliaoji* [A collection of Zauzou language]. Kunming: Yunnan Nationality Publishing House.
- Li, Yu. 2020. A Grammar of Zauzou. Ph.D. dissertation. The State University of New York at Buffalo.
- Senft, Gunter. 1993. A grammaticalization hypothesis on the origin of Kilivila classificatory particles. *Sprachtypologie und Universalienforschung* 46, 100–112.
- Shibatani, Masayoshi. 2021. *Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni* (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)

# Classifiers and nominalization in Lan Hmyo

Yoshihisa Taguchi<sup>1</sup>

## ABSTRACT

This study describes the use of classifiers in Lan Hmyo, a Hmongic language spoken in Guizhou, China, with a special focus on its relation to grammatical nominalization, guided by Shibatani's nominalization theory. The findings indicate that although the behavior of classifiers in Lan Hmyo largely conforms to the expectations of the theory, it also exhibits a deviation in the behavior of genitive nominal and pronominal phrases.

## 1. INTRODUCTION

In this study, the author describes the behavior of classifiers in Lan Hmyo, a Hmongic language, in relation to grammatical nominalization under the theoretical framework of Masayoshi Shibatani's nominalization theory (Shibatani 2021, 2023)<sup>2</sup>. Grammatical nominalization denotes a process by which a syntagma functioning as a nominal phrase is derived. The 'NP use' of a nominalization (i.e., a derived nominal phrase) is the use of a nominalization as a noun phrase. The 'modification use' of a nominalization refers to the use of a nominalization in a noun phrase, in which it constitutes a syntagma with a noun to codify the relation of modification. In this context, the nominalization serves as a modifier and the noun as the modified element. A brief sketch of the grammar of Lan Hmyo, relevant to nominalization, is provided below.

### (1) Noun/classifier phrase structure:

The structure of noun phrases can be schematized as: [X Noun X] (X denotes a modifying phrase). In this schema, only nouns are obligatory. The structure of classifier phrases can be schematized as: [X NUM CLF X Noun X DEM] (X denotes a modifying phrase). In this schema, only classifiers are obligatory. Note that demonstratives and numerals can appear in sentences only when classifiers co-occur with them; thus, they require the presence of a classifier in a phrase to license their existence. This study assumes that numerals and demonstratives require a classifier phrase as their head.

### (2) Distinction between nouns and adjectives:

In Lan Hmyo, the distinction between nouns and adjectives is consistently clear, and this distinction is primarily evidenced through syntactic behavior. The category of 'adjective' as a word class can be identified, but only as a subclass of the category of verb<sup>3</sup>. Verbs in Lan Hmyo

---

<sup>1</sup> Graduate School of Human Sciences, Chiba University, 1-33 Yayoi, Inage-ku, Chiba City, 263-8522, Japan.  
email: ytaguchi@chiba-u.jp

<sup>2</sup> Lan Hmyo is one of the West Hmongic languages (aka, Chuanqiandian) spoken in Guizhou Province in China. Lan Hmyo has lexical tones, designated as A, B, and C.

<sup>3</sup> For more information on the category of adjectives and their behavior in relative clauses in Lan Hmyo, see Taguchi (2019).

are defined as a class of words that can be directly negated using a negative marker, which always precedes them. In contrast, nouns must co-occur with a copula to undergo negation.

### (3) Definiteness/specificity distinction:

Demonstratives can function as a definite marker. The use of classifiers without numerals and demonstratives, ‘bare classifier construction’, can be interpreted as definite in some environments.

### (4) Non-classifying nominalizers:

Lan Hmyo has two non-classifying nominalizers: *taA/təA-* and *-moA*<sup>4</sup>. *-moA* both functions as a linking device and a nominalizer.

This study draws data from narrative texts (Taguchi and Song 2024) and those gathered through elicitation. The title of the text is indicated in the examples from narrative text data. Examples without titles were from those gathered through elicitation.

## 2. OVERVIEW OF NOUN CLASSIFICATION AND GRAMMATICAL NOMINALIZATIONS IN LAN HMYO

### 2.1 Noun classification in Lan Hmyo

Nouns are classified into several classes by classifiers: animate (no distinction between human and non-human), linear (one-dimensional), planar (two-dimensional), spherical (three-dimensional), and some classes based on functions. These classifiers can be regarded as classifiers often described under the name of ‘sortal classifiers’. Lan Hmyo also has ‘mensural classifiers’, as described in the literature. These include classifiers denoting units of measures and noun-derived classifiers denoting certain types of container concepts, such as bottles, buckets, and bags. Common features that characterize sortal and mensural classifiers are reduplication, denoting distributional meaning and constituting a syntagma with numerals and demonstratives.

### 2.2 Grammatical nominalizations in Lan Hmyo

#### [1] Numerals<sup>5</sup>

Cardinal numerals, at least those lower than ten, cannot stand alone and must co-occur with classifiers to appear in sentences. A numeral and a classifier constitute a classifier phrase, which follows a strict numeral-classifier order. Classifier phrases constituted by a numeral and a classifier behave like noun phrases; they can appear in the topic, subject, and object positions of a sentence. In the second line of the example below, the numeral *pluA-* ‘four’ appears with a classifier *ðanB* ‘CLF:animate’ in the object position of the verb *ŋoB* ‘bear.’

---

<sup>4</sup> Before most of the verbs (including adjectives), *taA/təA-* appears as *təA-*; however, the other form *taA-* appears just before a restricted number of verbs, especially before those expressing colors.

<sup>5</sup> This study only includes information on cardinal numerals and does not mention ordinal numerals.

- (1) *ŋoB muAδuB- δaNB -waC.*  
 bear IGN- CLF -PTCL  
 ‘How many pigs were born?’

*ŋoB pluA- δaNB.*  
 bear four- CLF  
 ‘Four.’ (Soul Summoning)

The obligatoriness of co-occurrence with classifiers is somewhat relaxed in higher numbers. See the example below in which the speaker first used the numeral *tsiA-tʃheNA* ‘three thousand’ without a classifier, then rephrased it with the classifier *ʔlaNA*.

- (2) *təA-nthuC ntoB nəA-piB ... təA-aC tɛaB niA-piB aAtaNA -niB,*  
 NMLZ-write paper ANPH-house NMLZ-do official ANPH-house child -PTCL  
*nthuC szB-tuC tsiA-tʃheNA təAtsC... tsiA-tʃheNA ʔlaNA təAtsC.*  
 write PURP-acquire three-thousand letter three-thousand CLF letter  
 ‘The son of the official must write three thousand characters.’ (Fetching a Drum from Sichuan)

## [2] Demonstratives

Lan Hmyo has two demonstratives: the proximal *-nonB* and the distal *-eB*. Demonstratives obligatorily co-occur with classifiers to appear in sentences. While classifiers can stand alone, demonstratives require a classifier as their head in a classifier phrase. The syntactic order within a classifier phrase follows the pattern: classifier-demonstrative. In the third line of the example below, the demonstrative *-eB*, together with a classifier *δaNB*, is placed in the topic position of the sentence.

- (3) *naB, ʁoNA naB ʁoNA -niB.*  
 this.way 2PL PTCL 2SG -PTCL  
 ‘He asked one of the thieves, “What’s your name?”’

*kaNB jonC ... kaNB jonC joA ngoA.*  
 1SG OP 1SG COP oil slippery  
 ‘One thief said, “My name is Oil-slippery.”’

*δaNB -eB -niB.*  
 CLF DEM PTCL  
 ‘What is his name?’ (Go-In-Oil-Then-Not-Sticky)

## [3] Genitive noun phrases

In Lan Hmyo, genitive noun phrases and genitive pronominal phrases are formed by adding the modification marker *-moA*. The genitive noun or pronominal phrases appear before

the modified noun. Example (4) illustrates a genitive noun phrase, while example (5) shows a genitive pronominal phrase.

- (4) *niB tsoC- puC ηteaC ηtonAhmonC θhuC niB-piB aAniB -moA*  
 3 CLNK1- sleep arrive midnight pull.out 3-house brother -MOD  
*təAntaBwiB nonA -loA.*  
 finger eat -PTCL  
 ‘When they slept until midnight, the old woman pulled out one of the boy’s fingers and ate it.’ (The Mountain Witch 1)

- (5) *təBqwenAnonB χonA- ?miA aAntenB hleA niB -moA taA-mbuAsenA,*  
 thereafter CLNK1-DISP1 knife cut 3 -MOD NMLZ-liver  
*niB χonA- δaC -loA.*  
 3 CLNK1- die -PTCL  
 ‘Then he stabbed her in the liver, and the old woman died.’ (The Mountain Witch 1)

Genitive noun and pronominal phrases, formed by adding a modification marker *-moA*, can function as NPs. In the examples below, the genitive pronominal phrase, *kanB -moA* ‘mine,’ is in the object position of the verb *nonA* ‘to eat.’

- (6) *niB tsoC- zoCseA, konA nonA kanB -moA -loA.*  
 3 CLNK1- say 2SG eat 1SG -MOD -PTCL  
 ‘Then the man said to them, “You ate my thing!”’ (A Story of a Silly Man and his Smart Wife)

Only when the modified noun is a noun denoting a kin term, it is optionally possible not to use *-moA* (see Example 7).

- (7) *niB δanB aAtanA -eB tsoC- qhliA ηgoA khuC-ntuA qoC taA-ζonA.*  
 3 CLF child -DEM CLNK1- take.off CLF shoe-cloth GOAL NMLZ-gate  
 ‘Then the boy took off his cloth shoes and put them at the gate of the well.’ (The Dragon’s Head)

#### [4] Adjectives

In Lan Hmyo, for an adjective to behave like a noun, it must be nominalized by a general nominalizer *təA-/taA-*, which is placed before adjectives. In the following example, the adjective ‘big’ is nominalized by *təA-* and co-occurs with a classifier *?lanA*. The order of the classifier and nominalized adjective is classifiers-adjectives.

- (8) *toC kanB tiB ?lanA təA- ndanC luB.*  
 BEN 1SG take CLF NMLZ-big come  
 ‘Take a big one for me.’



When an adjective modifies a noun, it must also be nominalized by a general nominalizer *təA-/taA-*. Nominalized adjectives can appear on either side of a noun. The following two examples express approximately the same meaning:

- (9) *kanB maB tuC təiA- phoNA aAzuC taA- qloA.*  
 1SG buy acquire three- CLF clothes NMLZ-white
- (10) *kanB maB tuC təiA- phoNA taA- qloA aAzuC.*  
 1SG buy acquire three- CLF NMLZ-white clothes  
 ‘I bought three white clothes.’

## [5] Verb-based nominalizations

### [5a] Relative clauses

Relative clauses in Lan Hmyo can be classified by two conditions: whether the relative clause is pre-head or post-head and whether the nominalizer used is *taA/təA-* or *-moA*, or no nominalizer is used. These conditions allow us to classify relative clauses into five types: (1) HEAD [*taA/təA-* ...], (2) [*taA/təA-* ...] HEAD, (3) [... *-moA*] HEAD, (4) [*taA/təA-* ... *-moA*] HEAD, and (5) [...] CLF N DEM. The strategy of forming relative clauses is gapping. In the examples below, relative clauses are indicated in square brackets ([ ]).

#### [5a-1] Post-head relative clause with *taA/təA-*

Relative clauses of this type are constructed by placing verb phrases nominalized by *təA-/taA-* after head nouns.

- (11) *naA [təA- tshiA lanB]*  
 human NMLZ-speak speech  
 ‘person who speaks/is speaking’<sup>6</sup>

#### [5a-2] Pre-head relative clause with *taA/təA-*

Relative clauses of this type are constructed by placing verb phrases nominalized by *təA-/taA-* before head nouns.

- (12) [*təA- tshiA lanB*] *naA*  
 NMLZ- speak speech human  
 ‘person who speaks/is speaking’

#### [5a-3] Pre-head relative clauses with *-moA*

Relative clauses of this type are constructed by placing verb phrases nominalized by *-moA* before head nouns. They tend to be interpreted as describing a specific event.

<sup>6</sup> Lan Hmyo does not have tenses as a verbal category. The examples (11)–(15) can be interpreted as describing a past event.

- (13) [*tshiA lanB -moA*] *naA*  
 speak speech -MOD human  
 ‘person who is speaking’

#### [5a-4] Pre-head relative clause with *taA-* ... *-moA*

Relative clauses of this type are constructed by placing verb phrases nominalized by both *təA-/taA-* and *-moA* before head nouns.

- (14) [*təA- tshiA lanB -moA*] *naA*  
 NMLZ- speak speech -MOD human  
 ‘person who speaks/is speaking’

#### [5a-5] Pre-head relative clause without nominalizers

Relative clauses of this type are constructed by placing verb phrases before the head formed with a classifier and a demonstrative (a noun can be optionally added). They are used to describe specific events.

- (15) [*tshiA lanB*] *ðanB naA -eB*  
 speak speech CLF human -DEM  
 ‘the person who is speaking’

In summary, relative clauses formed with *-moA* are always the pre-head type, and relative clauses with *təA-/taA-* can be either the pre-head or the post-head type. Presumably, prehead relative clauses without nominalizers (5a-5) are a variant of the type 5a-3, from which *-moA* is deleted under the condition that the head is formed with a classifier and a demonstrative. In the following sections, I will not mention this type.

#### [5b] Complement clauses

The strategy of forming complement clauses is juxtaposition: verbs and complement clauses are juxtaposed without a specific marker signaling their relationship. The author does not mention complement clauses in this study.

### 3. DATA

This section presents data on the classifier usage in grammatical nominalizations in their noun-phrase use and modification use. Figure 1 summarizes the distributions of the classifier use for grammatical nominalization in Shibatani’s (2021, 2023) implicational hierarchy.

Figure 1. Classifier use pattern in Lan Hmyo

NP use: CARD.NUM > DEM > \*GEN > <sup>OPT</sup>ADJ > <sup>OPT</sup>V-based NMLZ

Mod use: CARD.NUM > \*DEM > \*GEN > \*ADJ > \*V-based NMLZ

(\*=impossible; OPT=optional marking; unmarked=obligatory marking)

### [1] Numerals

Classifier phrases consisting of a cardinal numeral and a classifier indicate a distribution similar to that of noun phrases. See (1) for a relevant example. The classifier phrases of this type can co-occur with a noun as in (16), in which a classifier phrase, *ʔwuA-ðanB* ‘two-CLF:animate’, appears just before a noun, *aAtanA* ‘child.’

- (16) *εeA -eB maA ʔwuA- ðanB aAtanA niA-paB niA-miC muB nonA*  
time-DEM exist two- CLF child ANPH-father ANPH-mother go eat  
*ŋiA tsonC ŋdaB -loA.*  
buffalo release drum -PTCL  
‘Once upon a time, there were two children. They were staying at home while their parents went away to attend a ceremony.’ (The Mountain Witch 1)

### [2] Demonstratives

A classifier phrase formed with a demonstrative and a classifier can be used as an NP in sentences (NP use). See (3) for the relevant example. However, the classifier phrases of this type cannot co-occur with a noun, regardless of the order of the two constituents. See examples below<sup>7</sup>.

- (17) *\*naA ðanB -eB*  
person CLF -DEM  
(Intended meaning ‘that person’)

- (18) *\*ðanB -eB naA*  
CLF -DEM person  
(Intended meaning ‘that person’)

### [3] Genitive noun phrases

As noted in Section 2, in forming a genitive noun phrase, *-moA* is always used, except when the modified noun is a noun denoting a kin term. Therefore, we conclude that classifiers do not form genitive noun phrases with nouns. Note that classifiers cannot precede genitive nouns formed by *-moA* (see 19).

<sup>7</sup> The only possible sequence to express the intended meaning is *ðanB naA -eB*. Nouns must appear between classifiers and demonstratives.

- (19) \**ʔanA kanB -moA -eB*  
 CLF 1SG -MOD -DEM  
 ‘that of mine (Intended meaning)’<sup>8</sup>

#### [4] Adjectives

As noted in Section 2, adjectives can co-occur with a classifier when they are nominalized by a general nominalizer *təA-/taA-*, placed before adjectives. Therefore, when we examine the behavior of adjectives in Shibatani’s hierarchy, we examine nominalized adjectives (or adjective-based nominalization). A syntagma comprising a classifier and a nominalized adjective can be used as an NP in sentences. See (20), in which the classifier *phoNA* denotes an entity categorized as a kind of clothes.

- (20) *tsiA- phoNA taA- qloA*  
 three- CLF NMLZ- white  
 ‘three white clothes’

Although the sequence of a classifier and a nominalized adjective can appear before a noun, as in (21), this can be interpreted as the co-occurrence of a classifier phrase (*tsiA- phoNA*) and a noun phrase (*taA- qloA aAzuC*). Therefore, it is not regarded as an example of the modification-use of the syntagma comprising a classifier and a nominalized adjective. The author returns to this issue in Section 4.

- (21) *tsiA- phoNA taA- qloA aAzuC -eB*  
 three- CLF NMLZ- white clothes -DEM  
 ‘those three white clothes’

#### [5] Verb-based nominalizations

##### [5a] Relative clauses

As illustrated in Section 2, relative clauses in Lan Hmyo are formed with nominalizers *təA-/taA-* or *-moA*. These clauses co-occur with a classifier that constitutes a phrase functioning as an NP. There are two patterns for this. The first is formed from the post-head or pre-head relative clause with *taA/təA-* (5a-1 and 5a-2) by deleting the head noun. The second is formed with a pre-head relative clause with *taA- ... -moA* (5a-4) by deleting the head noun<sup>9</sup>. See (22) and (23) for an example of the first and second pattern, respectively.

[5a-1]/[5a-2] Post-head/pre-head relative clause with *taA/təA-*

- (22) *ðanB [təA- tshiA lanB] -eB*  
 CLF NMLZ- speak speech -DEM  
 ‘that person who speaks/is speaking’

<sup>8</sup> The only possible sequence to express the intended meaning is *kanB -moA ʔanA -eB*., where a genitive noun phrase *kanB -moA* modifies a classifier phrase *ʔanA -eB*.

<sup>9</sup> The phrase formed by the similar procedure in the case of [5a-3], \**ðanB [tshiA lanB -moA] -eB* is not acceptable.

[5a-4] Pre-head relative clause with *taA- ... -moA*

- (23)  $\delta anB$  [*təA- tshiA lanB -moA*] *-eB*  
CLF NMLZ- speak speech -MOD -DEM  
'that person who speaks/is speaking'

As in the case of adjectives, although the two sequences of a classifier and a relative clause can co-occur with a noun (see Examples 24 and 25, respectively), this is presumably not the case for the modification use of the syntagma comprising a classifier and a relative clause.

- (24)  $\delta anB$  [*təA- tshiA lanB*] *naA -eB*  
CLF NMLZ- speak speech human -DEM  
'the person who speaks/is speaking'

- (25)  $\delta anB$  [*təA- tshiA lanB -moA*] *naA -eB*  
CLF NMLZ- speak speech -MOD human -DEM  
'the person who speaks/is speaking'

Rather than as the combination of a classifier phrase and a noun, examples (24) and (25) are better interpreted as a sequence of a classifier,  $\delta anB$ , and a noun phrase, in which a noun, *naA*, is modified by a relative clause. This issue is discussed in Section 4.

#### 4. ANALYSIS AND DISCUSSION

The pattern exhibited by the classifiers of Lan Hmyo in relation to the five types of grammatical nominalization seems to largely conform to Shibatani's hierarchy. However, there is a salient deviation from this hierarchy. Whereas classifiers cannot co-occur with genitive noun or pronominal phrases as they do with nouns, they can co-occur with adjective- and verb-based nominalizations as optional constructions, which are positioned lower in the hierarchy. Additionally, the author addresses the argument to explain the facts noted in subsections [4] and [5] of Section 4. The author maintained that the alleged examples of the modification use of adjective- and verb-based nominalizations should be analyzed as a sequence of a classifier and a noun phrase. This analysis is supported by the fact that demonstratives require a classifier to appear in sentences. If we assume that this means that demonstratives require a classifier phrase as their head, we can offer the following explanation: let us consider the case of verb-based nominalization in (24) as an example. If we assume that the sequence of a classifier,  $\delta anB$ , and a relative clause, *təA- tshiA lanB*, in (24), is an example of the modification-use and that this sequence modifies the following noun *naA*, then, the demonstrative appearing at the final position cannot have a classifier phrase as its head. The sequence in question must have the structure in which a relative clause, *təA- tshiA lanB*, modifies the following noun *naA*, constituting an NP, which, in turn, modifies the preceding classifier  $\delta anB$ , constituting a classifier phrase. Then, the demonstrative felicitously has a classifier phrase as its head. Therefore, the structure is not \* $[[_{NP} \delta anB [təA- tshiA lanB]] naA]_{NP} -eB$ ] but should be  $[_{CLFP} \delta anB [_{NP} [təA- tshiA lanB] naA]_{NP}]_{CLFP} -eB]_{CLFP}$ .

## 5. CONCLUSION

This study examines the use of classifiers in Lan Hmyo in relation to grammatical nominalization under the guidance of Shibatani's nominalization theory. The results suggest that although the data of Lan Hmyo largely conform to the expectations of the theory, they also exhibit a deviation in the behavior of genitive nominal or pronominal phrases. Further studies should explore the reasons for this anomaly. Additionally, in Lan Hmyo, demonstratives placed in the phrase-final position suggest that the sequences that are seemingly the modification uses of adjective-based and verb-based nominalizations should be analyzed as a sequence of a classifier and a noun phrase.

## ABBREVIATIONS

|          |                            |
|----------|----------------------------|
| 1        | first person               |
| 2        | second person              |
| 3        | third person               |
| ADJ      | adjective                  |
| ANPH     | anaphoric                  |
| BEN      | benefactive                |
| CARD.NUM | cardinal numeral           |
| CLF      | classifier                 |
| CLNK     | clause linker              |
| COP      | copular                    |
| DEM      | demonstrative              |
| DISP     | disposal                   |
| GEN      | genitive noun phrase       |
| IGN      | ignorative                 |
| MOD      | modification marker        |
| NMLZ     | nominalizer/nominalization |
| NUM      | numeral                    |
| PL       | plural                     |
| PTCL     | particle                   |
| PURP     | purposive                  |

## ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my Lan Hmyo consultant, Mr. Yun Song. The present study has been supported by grants from the Japan Society for the Promotion of Science (KAKENHI Project ID: 24K03886, 'A typological study on the grammar of Hmongic languages: with a focus on nominalization and serial verb construction', PI: Yoshihisa Taguchi) and from the National Institute for Japanese Language and Linguistics ('Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective', PI: Prashant Pardeshi).

## REFERENCES

- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)
- Taguchi, Yoshihisa. 2019. Rahakuga myaogo no kankeisetsu nituite (Relative clauses in Lan Hmyo). *Journal of Chiba University Eurasian Society* 21, 33–54.
- Taguchi, Yoshihisa and Yun Song. 2024. *Stories of Hmyo: Thirty-five narrative texts in Lan Hmyo*. Chiba: Department of Eurasian Languages and Cultures, Faculty of Letters, Chiba University.

# Classifiers as nominalizers in Gouliang Miao

Fang Wu<sup>1</sup> and Hongyong Liu<sup>2</sup>

## ABSTRACT

This study investigates the use of classifiers as nominalizers in noun-phrase (NP) and modifier (Mod) constructions in Gouliang Miao, a sub-dialect of the Xiangxi variety of the Miao language. In this dialect, classifiers combine with various elements—such as numerals, demonstratives, genitives, adjectives, and verb-based nominalizations—to form structures that both denote and categorize nominal entities (e.g., ‘two animate entities’, ‘this fruit-class thing’). The study examines whether the rich classifier-marking patterns in Gouliang Miao conform to the constraints posited in the Shibatani hierarchy, which predicts the distribution and development of classifiers. Our findings reveal that classifier marking on adjectives—ranked higher in the hierarchy—is more restricted than on lower-ranked verb-based nominalizations, contrary to the hierarchy’s predictions. Identifying and accounting for such exceptions offers valuable insights not only into the structure of Gouliang Miao but also into broader typological patterns of classifier systems.

## 1. INTRODUCTION

Nominalization is generally understood as the process of converting non-noun elements into nouns or syntactic units that function as noun phrases. Cross-linguistic comparisons reveal that nominalization can be divided into lexical nominalization and clausal nominalization. The former involves deriving nouns from verbs, adjectives, among others, while the latter involves nominalizing entire clauses such as relative clauses, complement clauses, or adverbial clauses (Comrie & Thompson 2007: 334, LaPolla 2008, Huang 2020). Common nominalization strategies in the world’s languages include morphological and syntactic means. Morphological strategies include affixation, the use of case markers or plural markers as nominalization markers, reduplication, and zero-marking (i.e., no morphological change or other marking); syntactic strategies include the use of particles, numerals, demonstratives, or nominalized words as nominalization markers (Li & Thompson 1981: 575, Comrie & Thompson 2007: 334–342, Yap & Grunow-Härsta 2010; Yap et al. 2011: 9–21; Gerner 2012).

Wu et al. (2023) describe three nominalization means in Fenghuang Miao (which Gouliang Miao belongs to): adding prefixes, using auxiliary words as markers, and using classifiers as nominalizers. And the nominalization function of classifiers is precisely their ‘relationalization’ function, which adds to verb-object structures, verb-complement structures, coordinated predicate structures, serial verb constructions, adverbial-predicate structures, and subject-predicate structures, transforming them into pre-posed relative clauses.

Contrary to these traditional understandings of nominalizations, Shibatani (2019, 2021, 2023) argues that (i) nominalizations also apply to nominals, such as nouns (e.g., village >

---

<sup>1</sup> College of Foreign Studies, Jinan University. Email: wufangwinne@163.com

<sup>2</sup> Faculty of Arts and Humanities, University of Macau. Email: hongyongliu@um.edu.mo



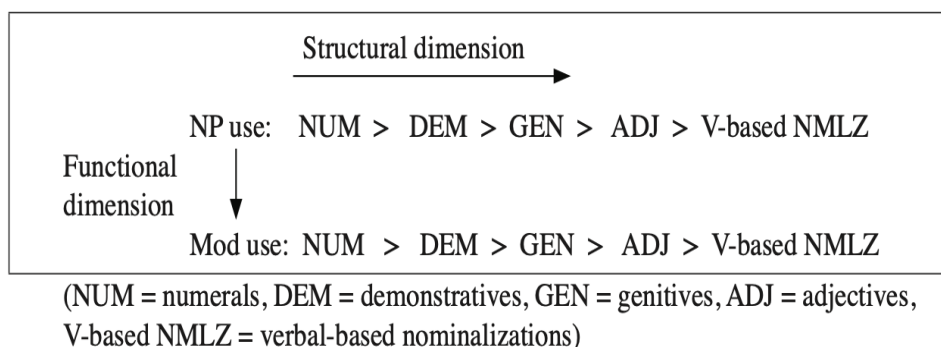
villager, Berlin > Berliner), numerals, and demonstratives (see below), and that (ii) grammatical nominalizations have two major uses or functions of NP use and Modification use. In the former they head a noun phrase and play a referential function of pointing out a thing in a real or imagined world. In the Mod use, they either restrict or identify the denotation of a head noun (Shibatani 2021, 2023). The same ideas apply to those nominalizations derived by classifier marking, as shown by the Gouliang Miao data, in which the nominalization functions of classifiers are seen not only in the predicate structures listed in Wu et al. (2023), but also in the other syntactic structures such as numerals, demonstratives, genitive noun phrases, adjectives, and verbal-based nominalizations.

The present study specifically focuses on the question of how classifiers function in the following types of grammatical nominalizations in Gouliang Miao, a sub-dialect of the Xiangxi dialect of the Miao language (Yang 2004).

- [1] Numerals
- [2] Demonstratives
- [3] Genitive noun phrases
- [4] Adjectives
- [5] Verb-based nominalizations

The study also tests if the patterns of classifier marking in Gouliang Miao conform to Shibatani's (2021, 2023) hypothesis. By means of the implicational hierarchy (Figure1), he hypothesizes that classifying nominalization markers, including classifiers and gender markers, are likely to occur in the structures higher in the hierarchy in both structural and functional dimensions. For example, the hierarchy predicts that numerals are marked by classifiers more prevalently than demonstratives and other structures positioned lower in the hierarchy and that NP-use contexts call for classifier marking more prevalently than modification-use contexts.

Figure 1. Hierarchy constraining classifier/gender marking (Shibatani 2023: 35)



The research question of the present study is how this hypothesis holds in Gouliang Miao. In Section 2, we will examine the functions of classifiers in the NP use and the Mod use in Gouliang Miao. Section 3 will explore the semantics of classifiers in the NP use and in the modification use. Section 4 is the conclusion.

## 2. NOMINALIZATION FUNCTION OF CLASSIFIERS IN GOULIANG MIAO

### 2.1 Numeral (NUM)

Gouliang Miao is a classifier language in which numerals cannot directly modify nouns; instead, a classifier must intervene between the numeral and the noun. Thus, the sequence NUM-CL-N is obligatory. In this structure, the NUM-CL phrase functions as a modifier of the head noun, a usage referred to as the modification use of classifiers, as illustrated in (1a). Additionally, classifiers may combine with numerals to form a NUM-CL sequence that can serve as a syntactic argument, as demonstrated in (1b). In this context, the classifier phrase (NUM-CL) is analogous to an argumental noun phrase (NP), and this usage is termed the NP use of classifiers. In both the modification and NP uses, the presence of classifiers is mandatory.

- (1) a.       $\text{vu}^{22}$     $\text{m}\epsilon^{53}$     $[\text{pu}^{214} \text{ l}\epsilon^{214}]$ .                      (NP use)  
              3SG   have   three   CL  
              ‘He has three ones.’
- b.       $\text{vu}^{22}$     $\text{m}\epsilon^{53}$     $[[\text{pu}^{214} \text{ l}\epsilon^{214}] \text{ pju}^{42}]$ .                      (Mod use)  
              3SG   have   three   CL   house  
              ‘He has three houses.’

### 2.2 Demonstrative (DEM)

Classifiers may also co-occur with demonstratives. As illustrated in (2b), the CL-DEM sequence can independently function as a syntactic argument. In the modification use, the appropriate word order is CL-N-DEM, as shown in (2c). In both the NP use and the modification use, the presence of classifiers remains obligatory.

- (2) a.       $\text{mu}^{53}$     $\text{li}^{33}$     $\eta\text{x}^{33}$     $\eta\text{x}^{22}$     $\text{qa}^{214}$                        $\text{ti}^{214?}$   
              2SG   want   buy   CL   chicken                      which  
              ‘Which chicken do you want to buy?’
- b.       $\text{v}\epsilon^{22}$     $\text{li}^{33}$     $\eta\text{x}^{33}$     $[\eta\text{x}^{22} \text{ n}\tilde{\epsilon}^{42}]$ .                      (NP use)  
              1SG   want   buy   CL   this  
              ‘I want to buy this one.’
- c.       $\text{v}\epsilon^{22}$     $\text{li}^{33}$     $\eta\text{x}^{33}$     $[[\eta\text{x}^{22} \text{ qa}^{214}] \text{ n}\tilde{\epsilon}^{42}]$ .                      (Mod use)  
              1SG   want   buy   CL   chicken                      this  
              ‘I want to buy this chicken.’

### 2.3 Genitive (GEN)

In Gouliang Miao, genitive noun phrases may be constructed either by employing the genitive marker  $\text{nn}\eta^{42}$  or through the use of classifiers. For example,

- (3) a. mu<sup>53</sup> nɒŋ<sup>42</sup> ʃ<sup>42</sup> ŋi<sup>33</sup> ŋhã<sup>214</sup> ti<sup>214</sup>?  
 2SG GEN clothes COP CL which  
 ‘Which one is your clothes?’
- b. [vɛ<sup>22</sup> nɒŋ<sup>42</sup>] ŋi<sup>33</sup> ŋhã<sup>214</sup> nẽ<sup>42</sup>.  
 1SG GEN COP CL this  
 ‘Mine is this one.’
- c. [[vɛ<sup>22</sup> nɒŋ<sup>42</sup>] ʃ<sup>42</sup>] ŋi<sup>33</sup> ŋhã<sup>214</sup> nẽ<sup>42</sup>.  
 1SG GEN clothes COP CL this  
 ‘My clothes is this one.’
- (4) a. [[vɛ<sup>22</sup> nɒŋ<sup>42</sup>] ŋhã<sup>214</sup>] ŋi<sup>214</sup> nẽ<sup>42</sup>. (NP use)  
 1SG GEN CL COP here  
 ‘Mine is here.’
- b. [[[vɛ<sup>22</sup> nɒŋ<sup>42</sup>] ŋhã<sup>214</sup>] ʃ<sup>42</sup>] ŋi<sup>214</sup> nẽ<sup>42</sup>. (Mod use)  
 1SG GEN CL clothes COP here  
 ‘My clothes is here.’
- (5) a. [vɛ<sup>22</sup> ŋhã<sup>214</sup>] ŋi<sup>214</sup> nẽ<sup>42</sup>. (NP use)  
 1SG CL COP here  
 ‘Mine is here.’
- b. [[vɛ<sup>22</sup> ŋhã<sup>214</sup>] ʃ<sup>42</sup>] ŋi<sup>214</sup> nẽ<sup>42</sup>. (Mod use)  
 1SG CL clothes COP here  
 ‘My clothes is here.’

Examples (3)–(5) demonstrate that when the possessive marker *nɒŋ*<sup>42</sup> is present, the use of classifiers is optional (cf. (3c) and (4b)). In contrast, in the absence of the genitive marker, classifiers become obligatory for the formation of possessive constructions, as illustrated in (5).

## 2.4 Adjective (ADJ)

In Gouliang Miao, classifiers cannot be used to directly nominalize adjectives, either within noun phrase contexts or in modification contexts, as illustrated in examples (6)–(8).

NP use

- (6) a. \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [lɛ<sup>214</sup> lo<sup>53</sup>].  
 1SG PROG look\_for CL big  
*to mean*, ‘I am looking for a big one.’

- b. \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [[lo<sup>53</sup> lɛ<sup>214</sup>].  
 1SG PROG look\_for big CL  
*to mean, 'I am looking for a big one.'*

Mod use (prenominal)

- (7) a. \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [lɛ<sup>214</sup> lo<sup>53</sup>] qp<sup>214</sup>-ti<sup>42</sup>].  
 1SG PROG look\_for CL big basket  
*to mean, 'I am looking for a big basket.'*

- b. \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [[[lo<sup>53</sup> lɛ<sup>214</sup>] qp<sup>214</sup>-ti<sup>42</sup>].  
 1SG PROG look\_for big CL basket  
*to mean, 'I am looking for a big basket.'*

Mod use (postnominal)

- (8) a. \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [qp<sup>214</sup>-ti<sup>42</sup> [lɛ<sup>214</sup> lo<sup>53</sup>]].  
 1SG PROG look\_for basket CL big  
*to mean, 'I am looking for a big basket.'*

- b. \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [qp<sup>214</sup>-ti<sup>42</sup> [lo<sup>53</sup> lɛ<sup>214</sup>]].  
 1SG PROG look\_for basket big CL  
*to mean, 'I am looking for a big basket.'*

However, when adjectives and classifiers are combined with different nominalizers, a range of syntactic patterns can be observed. Gouliang Miao employs two dedicated nominalizers: *mv*<sup>53</sup>, which is head-initial, and *nmɿ*<sup>42</sup>, which is head-final. As shown in examples (9)–(11), *mv*<sup>53</sup> is used to nominalize the adjective lo<sup>53</sup> ‘big’ in both NP and modification contexts.

NP use

- (9) vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [mv<sup>53</sup> lo<sup>53</sup>].  
 1SG PROG look\_for NMLZ big  
*‘I am looking for a big one/big ones.’*

Mod use

- (10) (mv<sup>53</sup> + A + head noun)  
 vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [[mv<sup>53</sup> lo<sup>53</sup>] qp<sup>214</sup>-ti<sup>42</sup>].  
 1SG PROG look\_for NMLZ big basket  
*‘I am looking for a big basket/big baskets.’*

- (11) (mv<sup>53</sup> + A + head noun)  
 vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [qp<sup>214</sup>-ti<sup>42</sup> [mv<sup>53</sup> lo<sup>53</sup>]].  
 1SG PROG look\_for basket NMLZ big  
*‘I am looking for a big basket/big baskets.’*

When combined with adjectives and the nominalizer *mv*<sup>53</sup>, classifiers can serve a nominalizing function in both NP and modification contexts. However, this nominalizing function is limited to specific syntactic sequences: CL-*mv*<sup>53</sup>-A, CL-*mv*<sup>53</sup>-A-head noun, and *mv*<sup>53</sup>-A-CL-head noun. These patterns are illustrated in examples (12)–(17).

NP use

- (12) (Cl- *mv*<sup>53</sup>- A)  
 vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [lɛ<sup>214</sup> [mv<sup>53</sup> lo<sup>53</sup>]].  
 1SG PROG look\_for CL NMLZ big  
 ‘I am looking for a big one.’

- (13) (*mv*<sup>53</sup>- A- Cl)  
 \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [[mv<sup>53</sup> lo<sup>53</sup>] lɛ<sup>214</sup>].  
 1SG prog look\_for NMLZ big CL  
*to mean*, ‘I am looking for a big one.’<sup>3</sup>

Mod use

- (14) (Cl-*mv*<sup>53</sup>-A-head noun)  
 vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [lɛ<sup>214</sup>[[mv<sup>53</sup> lo<sup>53</sup>] qv<sup>214</sup>-ti<sup>42</sup>]]  
 1SG PROG look\_for CL NMLZ big basket  
 ‘I am looking for a big basket.’
- (15) (*mv*<sup>53</sup>-A-CL-head noun)  
 vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [[[mv<sup>53</sup> lo<sup>53</sup>] lɛ<sup>214</sup>] qv<sup>214</sup>-ti<sup>42</sup>].  
 1SG PROG look\_for NMLZ big CL basket  
 ‘I am looking for a big basket.’

- (16) (head noun- Cl- *mv*<sup>53</sup>- A)  
 \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [qv<sup>214</sup>-ti<sup>42</sup> [lɛ<sup>214</sup> [mv<sup>53</sup> lo<sup>53</sup>]]].  
 1SG PROG look\_for basket CL NMLZ big  
*to mean*, ‘I am looking for a big basket.’

- (17) (head noun - *mv*<sup>53</sup>- A- Cl)  
 \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [qv<sup>214</sup>-ti<sup>42</sup> [[mv<sup>53</sup> lo<sup>53</sup>] lɛ<sup>214</sup>]].  
 1SG PROG look\_for basket NMLZ big CL  
*to mean*, ‘I am looking for a big basket.’

In Gouliang Miao, the nominalizer *nvŋ*<sup>42</sup> can be employed within modification phrases. As illustrated in (18), *nvŋ*<sup>42</sup> appears after *qv*<sup>42</sup> *no*<sup>214</sup> ‘what’, which functions as a modifier of the

<sup>3</sup> In Gouliang Miao, measure adjectives—such as lo<sup>53</sup> ‘big’, cu<sup>214</sup> ‘small’, ʂã<sup>214</sup> ‘tall/high’, lɛ<sup>42</sup> ‘short’, tu<sup>42</sup> ‘long’, hẽ<sup>42</sup> ‘heavy’, ɛa<sup>214</sup> ‘light’, to<sup>214</sup> ‘deep’, mjẽ<sup>42</sup> ‘shallow’, ta<sup>214</sup> ‘thick’, ŋẽ<sup>22</sup> ‘thin’—can combine with a classifier to form a compound adjective. For example, the phrase vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [mv<sup>53</sup> [lo<sup>53</sup> lɛ<sup>214</sup>]] ‘I am looking for a big one’ demonstrates this construction, where lo<sup>53</sup> lɛ<sup>214</sup> functions as a compound adjective meaning ‘big in size’.

head noun  $qp^{214}-ti^{42}$  ‘basket’. In this construction,  $no\eta^{42}$  occurs in the sequence modifier– $no\eta^{42}$ –modified.

- (18)  $mu^{53}$   $li^{33}$   $qp^{42}$   $no^{214}$   $no\eta^{42}$   $qp^{214}-ti^{42}?$   
 2SG want what NMLZ basket  
 ‘What kind of basket do you want?’

Unlike  $mv^{53}$ ,  $no\eta^{42}$  cannot be used to directly nominalize adjectives, as demonstrated in examples (19)–(21).

NP use

- (19) (A+  $no\eta^{42}$ )  
 $*v\epsilon^{22}$   $tai^{35}\eta i^{214}$   $tsh\eta^{35}$   $[lo^{53} no\eta^{42}]$ .  
 1SG PROG look\_for big NMLZ  
*to mean*, ‘I am looking for a big one/big ones.’

Mod use

- (20) (A+  $no\eta^{42}$ +head noun)  
 $*v\epsilon^{22}$   $tai^{35}\eta i^{214}$   $tsh\eta^{35}$   $[lo^{53} no\eta^{42}] qp^{214}-ti^{42}$ .  
 1SG PROG look\_for big NMLZ basket  
*to mean*, ‘I am looking for a big basket/big baskets.’

- (21) (head noun +A+  $no\eta^{42}$ )  
 $*v\epsilon^{22}$   $tai^{35}\eta i^{214}$   $tsh\eta^{35}$   $[qp^{214}-ti^{42} [lo^{53} no\eta^{42}]]$ .  
 1SG PROG look\_for basket big NMLZ  
*to mean*, ‘I am looking for a big basket/big baskets.’

Furthermore, as illustrated in (22)–(23), classifiers are also unable to nominalize sequences consisting of adjectives followed by  $no\eta^{42}$ .

NP use

- (22) (CL+A+  $no\eta^{42}$ )  
 $*v\epsilon^{22}$   $tai^{35}\eta i^{214}$   $tsh\eta^{35}$   $[le^{214} [lo^{53} no\eta^{42}]]$ .  
 1SG PROG look\_for CL big NMLZ  
*to mean*, ‘I am looking for a big one.’

- (23) (A+  $no\eta^{42}$ + CL)  
 $*v\epsilon^{22}$   $tai^{35}\eta i^{214}$   $tsh\eta^{35}$   $[[lo^{53} no\eta^{42}] le^{214}]$ .  
 1SG PROG look\_for big NMLZ CL  
*to mean*, ‘I am looking for a big one.’

Mod use

- (24) (CL+ A+ nɔŋ<sup>42</sup>+head noun)  
 \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɔ<sup>35</sup> [[lɛ<sup>214</sup> [lo<sup>53</sup> nɔŋ<sup>42</sup>]] qɔ<sup>214</sup>-ti<sup>42</sup>].  
 1SG PROG look\_for CL big NMLZ basket  
*to mean*, ‘I am looking for a big basket.’
- (25) (head noun + CL+ A+ nɔŋ<sup>42</sup>)  
 \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɔ<sup>35</sup> [qɔ<sup>214</sup>-ti<sup>42</sup> [lɛ<sup>214</sup> [lo<sup>53</sup> nɔŋ<sup>42</sup>]]]  
 1SG PROG look\_for basket CL big NMLZ  
*to mean*, ‘I am looking for a big basket.’
- (26) (A+ nɔŋ<sup>42</sup>+ CL+ head noun)  
 \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɔ<sup>35</sup> [[[lo<sup>53</sup> nɔŋ<sup>42</sup>] lɛ<sup>214</sup>] qɔ<sup>214</sup>-ti<sup>42</sup>].  
 1SG PROG look\_for big NMLZ CL basket  
*to mean*, ‘I am looking for a big basket.’
- (27) (head noun + A+ nɔŋ<sup>42</sup>+ CL)  
 \*vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɔ<sup>35</sup> [qɔ<sup>214</sup>-ti<sup>42</sup> [[[lo<sup>53</sup> nɔŋ<sup>42</sup>] lɛ<sup>214</sup>]].  
 1SG PROG look\_for basket big NMLZ CL  
*to mean*, ‘I am looking for a big basket.’

However, when both nominalizers co-occur, or when both nominalizers appear together with a classifier, the resulting sequences are grammatical, as demonstrated in examples (28)–(34).

NP use

- (28) (mɔ<sup>53</sup>+A+ nɔŋ<sup>42</sup>)  
 vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɔ<sup>35</sup> [mɔ<sup>53</sup> lo<sup>53</sup> nɔŋ<sup>42</sup>].  
 1SG PROG look\_for NMLZ big NMLZ  
 ‘I am looking for a big one/big ones.’
- (29) (mɔ<sup>53</sup>+A+ nɔŋ<sup>42</sup>+ CL)  
 vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɔ<sup>35</sup> [[mɔ<sup>53</sup> lo<sup>53</sup> nɔŋ<sup>42</sup>] lɛ<sup>214</sup>].  
 1SG PROG look\_for NMLZ big NMLZ CL  
 ‘I am looking for a big one.’
- (30) (CL+ mɔ<sup>53</sup>+A+ nɔŋ<sup>42</sup>)  
 vɛ<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɔ<sup>35</sup> lɛ<sup>214</sup> [mɔ<sup>53</sup> lo<sup>53</sup> nɔŋ<sup>42</sup>].  
 1SG PROG look\_for CL NMLZ big NMLZ  
 ‘I am looking for a big one.’

Mod use

- (31) (mv<sup>53</sup>+A+ nv<sup>42</sup>+head noun)  
vε<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [[mv<sup>53</sup> lo<sup>53</sup> nv<sup>42</sup>] qp<sup>214</sup>-ti<sup>42</sup>].  
1SG PROG look\_for NMLZ big NMLZ basket  
‘I am looking for big baskets/a big basket.’
- (32) (head noun + mv<sup>53</sup>+A+ nv<sup>42</sup>)  
vε<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [qp<sup>214</sup>-ti<sup>42</sup> [mv<sup>53</sup> lo<sup>53</sup> nv<sup>42</sup>]].  
1SG PROG look\_for basket NMLZ big NMLZ  
‘I am looking for big baskets/a big basket.’
- (33) (CL+ mv<sup>53</sup>+A+ nv<sup>42</sup>+head noun)  
vε<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [[le<sup>214</sup> [mv<sup>53</sup> lo<sup>53</sup> nv<sup>42</sup>]] qp<sup>214</sup>-ti<sup>42</sup>].  
1SG PROG look\_for CL NMLZ big NMLZ basket  
‘I am looking for a big basket.’
- (34) (mv<sup>53</sup>+A+ nv<sup>42</sup>+CL+head noun)  
vε<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [[[mv<sup>53</sup> lo<sup>53</sup> nv<sup>42</sup>] le<sup>214</sup>] qp<sup>214</sup>-ti<sup>42</sup>].  
1SG PROG look\_for NMLZ big NMLZ CL basket  
‘I am looking for a big basket.’

But if the head noun precedes the sequence CL+ mv<sup>53</sup>+A+ nv<sup>42</sup> or the sequence mv<sup>53</sup>+A+ nv<sup>42</sup>+CL, the resultant combinations are unacceptable.

- (35) (head noun+ CL+ mv<sup>53</sup>+A+ nv<sup>42</sup>)  
\*vε<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [qp<sup>214</sup>-ti<sup>42</sup> [le<sup>214</sup> [mv<sup>53</sup> lo<sup>53</sup> nv<sup>42</sup>]]].  
1SG PROG look\_for basket CL NMLZ big NMLZ  
to mean, ‘I am looking for a big basket.’
- (36) (head noun+ mv<sup>53</sup>+A+ nv<sup>42</sup>+CL)  
\*vε<sup>22</sup> tai<sup>35</sup>ŋi<sup>214</sup> tʂhɒ<sup>35</sup> [qp<sup>214</sup>-ti<sup>42</sup> [[mv<sup>53</sup> lo<sup>53</sup> nv<sup>42</sup>] le<sup>214</sup>]].  
1SG PROG look\_for basket NMLZ big NMLZ CL  
to mean, ‘I am looking for a big basket.’

Nominalized phrases containing mv<sup>53</sup> or/and nv<sup>42</sup> without a classifier exhibit a number-neutral interpretation. In contrast, when an adjective is nominalized by a classifier, the resulting phrase receives either a singular reading (with individual classifiers) or a plural reading (with collective classifiers).

## 2.5 V-based nominalization (V-based NMLZ)

In Gouliang Miao, the nominalizers nv<sup>42</sup> or mv<sup>53</sup> can be used to nominalize verb phrases. For example, in (37a), nv<sup>42</sup> nominalizes the verb phrase ‘she feeds’. The resulting nominalized verb phrase can then function as a modifier of the head noun, as illustrated in (37b).



- (37) a.     $\text{vu}^{22}$      $[\text{su}^{42} \text{ nɒŋ}^{42}]$   $\text{lo}^{53}$      $\text{ta}^{42}$ .  
          3SG    feed    NMLZ big    very  
          ‘The one(s) s/he feeds is/are big.’
- b.     $\text{vu}^{22}$      $[[\text{su}^{42} \text{ nɒŋ}^{42}] \text{ tɔ}^{214}\text{-mpa}^{35}]$   $\text{lo}^{53}$      $\text{ta}^{42}$ .  
          3SG    feed    NMLZ KP-pig    big    very  
          ‘The pig(s) that s/he feeds is/are big.’

In (38a),  $\text{mv}^{53}$  nominalizes the verb phrase ‘she feeds’. The resulting nominalized verb phrase can then function as a modifier of the head noun, as demonstrated in (38b).

- (38) a.     $\text{pu}^{214}$   $\text{ŋi}^{214}$   $[\text{mv}^{53} \text{ tʂhɒu}^{44}\text{toŋ}^{214}]$ .  
          1PL    COP    NMLZ farm  
          ‘We are farmers.’
- b.     $\text{pu}^{214}$   $\text{ŋi}^{214}$   $[\text{mje}^{53} [\text{mv}^{53} \text{ tʂhɒu}^{44}\text{toŋ}^{214}]]$ .  
          1PL    COP    people NMLZ farm  
          ‘We are farmers.’

In (39),  $\text{nvŋ}^{42}$  and  $\text{mv}^{53}$  co-occur to nominalize the verb phrase ‘she feeds’, and the resulting nominalized verb phrase can function as the modifier of the head noun, as illustrated in (39b-39c).

- (39) a.     $\text{pu}^{214}$   $\text{ŋi}^{214}$   $[\text{mv}^{53} \text{ tʂhɒu}^{44}\text{toŋ}^{214} \text{ nɒŋ}^{42}]$ .  
          1PL    COP    NMLZ farm    NMLZ  
          ‘We are farmers.’
- b.     $\text{pu}^{214}$   $\text{ŋi}^{214}$   $[\text{mje}^{53} [\text{mv}^{53} \text{ tʂhɒu}^{44}\text{toŋ}^{214} \text{ nɒŋ}^{42}]]$ .  
          1PL    COP    people NMLZ farm    NMLZ  
          ‘We are farmers.’
- c.     $\text{pu}^{214}$   $\text{ŋi}^{214}$   $[[\text{mv}^{53} \text{ tʂhɒu}^{44}\text{toŋ}^{214} \text{ nɒŋ}^{42}] \text{ mje}^{53}]$ .  
          1PL    COP    NMLZ farm    NMLZ    people  
          ‘We are farmers.’

In Gouliang Miao, a classifier can directly nominalize verb phrases when it follows the verb phrase, as illustrated in (40)–(41). In (40a), the classifier  $\text{ŋ}^{22}$  nominalizes the verb phrase ‘she feeds’, and the resulting nominalized verb phrase can function as a modifier of the head noun, as shown in (40b). In (41), classifiers cannot precede a verb phrase. Example (42) demonstrates that, in modification contexts, a classifier can nominalize a verb phrase only when the nominalized verb phrase is appositive.

- (40) a. vu<sup>22</sup> [su<sup>42</sup> ŋɿ<sup>22</sup>] lo<sup>53</sup> ta<sup>42</sup>. (NP use)  
 3SG feed CL big very  
 ‘The one that she feeds is big.’
- b. vu<sup>22</sup> [[su<sup>42</sup> ŋɿ<sup>22</sup>] tɔ<sup>214</sup>-mpa<sup>35</sup>] lo<sup>53</sup> ta<sup>42</sup>. (Mod use)  
 3SG feed CL pig big very  
 ‘The pig that she feeds is big.’
- (41) a. \*vu<sup>22</sup> [ŋɿ<sup>22</sup> su<sup>42</sup>] lo<sup>53</sup> ta<sup>42</sup>.  
 3SG CL feed big very  
*to mean*, ‘The one that she feeds is big.’
- b. \*vu<sup>22</sup> [[ŋɿ<sup>22</sup> su<sup>42</sup>] tɔ<sup>214</sup>-mpa<sup>35</sup>] lo<sup>53</sup> ta<sup>42</sup>.  
 3SG CL feed pig big very  
*to mean*, ‘The pig that she feeds is big.’
- (42) a. tɔ<sup>214</sup>-mpa<sup>35</sup>, [vu<sup>22</sup> su<sup>42</sup> ŋɿ<sup>22</sup>], lo<sup>53</sup> ta<sup>42</sup>. (NP use)  
 pig 3SG feed CL big very  
*to mean*, ‘The pig, which she feeds, is big.’
- b. \*tɔ<sup>214</sup>-mpa<sup>35</sup> [vu<sup>22</sup> su<sup>42</sup> ŋɿ<sup>22</sup>] lo<sup>53</sup> ta<sup>42</sup>.  
 pig 3SG feed CL big very  
*to mean*, ‘The pig that she feeds is big.’

Classifiers in Gouliang Miao may co-occur with the nominalizers *noŋ*<sup>42</sup> or/and *mv*<sup>53</sup>. In such cases, *noŋ*<sup>42</sup> or *mv*<sup>53</sup> nominalizes the verb phrase, while the subsequent classifier semantically categorizes the referent of the nominalized phrase and marks it as singular. Nominalized phrases containing *noŋ*<sup>42</sup> or/and *mv*<sup>53</sup> without a classifier exhibit a number-neutral interpretation. In contrast, when a verb phrase is nominalized by a classifier, the resulting phrase receives either a singular reading (with individual classifiers) or a plural reading (with collective classifiers).

- (43) a. vɛ<sup>22</sup> tʂhɔ<sup>35</sup> [mẽ<sup>22</sup> [tʂhɔu<sup>44</sup>toŋ<sup>21</sup> noŋ<sup>42</sup>]]. (NP use)  
 1SG look\_for CL farm NMLZ  
 ‘I look for a farmer.’
- b. vɛ<sup>22</sup> tʂhɔ<sup>35</sup> [[[tʂhɔu<sup>44</sup>toŋ<sup>21</sup> noŋ<sup>42</sup>] mẽ<sup>22</sup>] mǰẽ<sup>53</sup>].  
 1SG look\_for farm NMLZ CL people  
 ‘I look for a farmer.’
- (44) a. vɛ<sup>22</sup> tʂhɔ<sup>35</sup> [mẽ<sup>22</sup> [mv<sup>53</sup> tʂhɔu<sup>44</sup>toŋ<sup>214</sup>]]. (NP use)  
 1SG look\_for CL NMLZ farm  
 ‘I look for a farmer.’

- b.       $ve^{22}$      $t\dot{s}h\dot{v}^{35}$              $[m\ddot{e}^{22} [mj\ddot{e}^{53} [mp^{53} t\dot{s}h\dot{v}^{44}to\eta^{214}]]]$ .  
 1SG   look\_for      CL    people NMLZ farm  
 ‘I look for a farmer.’
- (45) a.       $ve^{22}$      $t\dot{s}h\dot{v}^{35}$              $[m\ddot{e}^{22} [mp^{53} t\dot{s}h\dot{v}^{44}to\eta^{214} n\dot{v}\eta^{42}]]$ .            (NP use)  
 1SG   look for      CL    NMLZ farm                      NMLZ  
 ‘I look for a farmer.’
- b.       $ve^{22}$      $t\dot{s}h\dot{v}^{35}$              $[[m\ddot{e}^{22} [mp^{53} t\dot{s}h\dot{v}^{44}to\eta^{214} n\dot{v}\eta^{42}]] mj\ddot{e}^{53}]$ . (Mod use)  
 1SG   look\_for      CL    NMLZ farm                      NMLZ people  
 ‘I look for a farmer.’

In summary, the syntactic distribution of classifier nominalization in Gouliang Miao is presented in Table 1.

Table 1. The Distribution of Classifiers as Nominalizers in NP and Modification Contexts

|                              | NP use                                                                           | Modification use                                                                                                                                                                                                                                                             |
|------------------------------|----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Number                       | The classifier is obligatory.                                                    | The classifier is obligatory.                                                                                                                                                                                                                                                |
| Dem                          | The classifier is obligatory.                                                    | The classifier is obligatory.                                                                                                                                                                                                                                                |
| GEN                          | With GEN, the classifier is optional.                                            | With GEN, the classifier is optional.                                                                                                                                                                                                                                        |
|                              | Without GEN, the classifier is obligatory.                                       | Without GEN, the classifier is obligatory.                                                                                                                                                                                                                                   |
| Adj                          | ① $mp^{53}+A+ n\dot{v}\eta^{42}+ CL$ ;<br>② $CL+ mp^{53}+A+ n\dot{v}\eta^{42}$ . | ① CL- $mp^{53}$ A -head noun;<br>② $mp^{53}$ - A- CL- head noun;<br>③ $mp^{53}+A+ n\dot{v}\eta^{42}+head\ noun$ ;<br>④ head noun + $mp^{53}+A+ n\dot{v}\eta^{42}$ ;<br>⑤ $CL+ mp^{53}+A+ n\dot{v}\eta^{42}+head\ noun$ ;<br>⑥ $mp^{53}+A+ n\dot{v}\eta^{42}+CL+head\ noun$ . |
|                              | <b>Classifiers cannot be used to directly nominalize adjectives.</b>             |                                                                                                                                                                                                                                                                              |
| <b>V-<br/>based<br/>NMLZ</b> | With NMLZ, the classifier is optional.                                           | With NMLZ, the classifier is optional.                                                                                                                                                                                                                                       |
|                              | Without NMLZ, the classifier is obligatory.                                      | Without NMLZ, the classifier is obligatory.                                                                                                                                                                                                                                  |

### 3. SEMANTICS OF CLASSIFIERS AS NOMINALIZERS IN NP USE AND MOD USE

In (1a), the function of the classifier is to use the cardinal number  $pu^{214}$  ‘three’ as the criterion to build up a set of entities, and each member in the set is made up of three entities. In this sense, classifiers are used as entity-creating nominalizers. In this nominalizing process, the Same-kind Condition must be satisfied.

**The initial definition** of the Same-kind Condition: if two objects are counted by using the same classifiers, then the two objects satisfy the Same-kind Condition.

The Same-kind Condition will turn the set of (46) to the set of (47), in which all the atoms in each member of the set are counted by using the same classifier.

$$(46) \quad \{\{x1, x2, x3\}, \dots, \{x1, y1, z1\}, \{x1, x2, y3\}, \{x3, y1, z3\}\}$$

$$(47) \quad \begin{aligned} \text{a.} \quad & [[\text{three-Cl}]] = \{\{x1, x2, x3\}, \{y1, y2, y3\}, \{z1, z2, z3\}\} = \{X, Y, Z\} \\ & X = \{x1, x2, x3\} = \lambda x \in X. \text{ every } x \text{ is counted by using the same classifier } \&|X|=3 \\ & Y = \{y1, y2, y3\} = \lambda y \in Y. \text{ every } y \text{ is counted by using the same classifier } \&|Y|=3 \\ & Z = \{z1, z2, z3\} = \lambda z \in Z. \text{ every } z \text{ is counted by using the same classifier } \&|Z|=3 \\ \text{b.} \quad & \text{Classifiers as entity-creating nominalizers: } \parallel \text{pu}^{214} \text{m}\tilde{\text{e}}^{22} \parallel = \{X, Y, Z, M\} \end{aligned}$$

For example, if flowers and clouds can be counted by using the same classifier, then flowers and clouds satisfy the Same-kind Condition.

In (1b), the function of the classifier in the modification use is to use the noun  $te^{214}$  ‘children’ as the criterion to build up a set of entities, and each member in the set is made up of three atoms, and each atom has the property of being a child. In this sense, classifiers are used as classifying nominalizers. In this classification process, the Same-kind Condition is stricter.

**The revised definition** of the Same-kind Condition: if two objects can be named under a shared term, and can be counted by using the same classifier, then the two objects satisfy the Same-kind Condition.

The revised Same-kind Condition will turn (48) to (49), in which all the atoms in each member are of the same kind.

$$(48) \quad \{\{x1, x2, x3\}, \dots, \{x1, y1, z1\}, \{x1, x2, y3\}, \{x3, y1, z3\}\}$$

$$(49) \quad \begin{aligned} \text{a.} \quad & [[\text{three-Cl-N}]] = \{\{x1, x2, x3\}, \{y1, y2, y3\}, \{z1, z2, z3\}\} = \{X, Y, Z\} \\ & X = \{x1, x2, x3\} = \lambda x \in X. \text{CHILD}(x) \&|X|=3 \\ & Y = \{y1, y2, y3\} = \lambda y \in Y. \text{FISH}(y) \&|Y|=3 \\ & Z = \{z1, z2, z3\} = \lambda z \in Z. \text{CHICKEN}(z) \&|Y|=3 \\ \text{b.} \quad & \text{Classifiers as classifying nominalizers: } \parallel \text{pu}^{214} \text{m}\tilde{\text{e}}^{22} \text{te}^{214} \parallel = X \end{aligned}$$

The contrast between (47a) and (49a) shows that classifiers in the NP use, as shown in (47a), turns a cardinal number to a set of entities which can be counted by the same classifier, and classifiers in the modification use, as shown in (49a), are able to generate a set of sum entities of the same kind.

#### 4. CONCLUSION

The present study examines the role of classifiers in grammatical nominalizations within both noun phrase (NP) and modification contexts in Gouliang Miao. In NP contexts, classifiers combine with numerals, demonstratives, possessors (NP or NP+NMLZR), adjectives, and verb-based nominalizations (VP or VP+NMLZR). Here, classifiers serve to establish a set of entities defined by these elements, enabling them to be counted. In modification contexts,

classifiers link the head noun with numerals, demonstratives, possessors (NP or NP+NMLZR), adjectives, and verb-based nominalizations (VP or VP+NMLZR). In this function, classifiers generate a set of composite entities that share two properties.

Overall, as shown in Figure 2, the data from Gouliang Miao support Shibatani’s hypothesis regarding the distribution of classifying nominalization markers along the functional dimension. However, Gouliang Miao exhibits a slightly different pattern from the proposed hierarchy: classifiers can be used with verb-based nominalizations in both NP and modification contexts, but they cannot directly nominalize adjectives when non-classifying nominalizers are absent. Classifiers can nominalize adjectives in both NP and modification contexts only when combined with non-classifying nominalizers.

Figure 2: Classifier marking in Gouliang Miao

Without non-classifying nominalizers

|          |            |       |       |        |         |
|----------|------------|-------|-------|--------|---------|
| NP use:  | CARD.NUM > | DEM > | GEN > | *ADJ > | V-based |
| Mod use: | CARD.NUM > | DEM > | GEN > | *ADJ > | V-based |

With non-classifying nominalizers

|          |            |       |          |          |            |
|----------|------------|-------|----------|----------|------------|
| NP use:  | CARD.NUM > | DEM > | OPTGEN > | OPTADJ > | OPTV-based |
| Mod use: | CARD.NUM > | DEM > | OPTGEN > | OPTADJ > | OPTV-based |

- \*: no marking or impossible
- OPT: optional
- unmarked: obligatory marking

The slight deviation from Shibatani’s hypothesis, along with several other phenomena, warrants further investigation. To better understand the diachronic development of classifiers in Gouliang Miao, it is necessary to conduct a detailed analysis of data from both related Miao languages and neighboring languages.

## ABBREVIATIONS

|        |               |
|--------|---------------|
| 1      | first person  |
| 2      | second person |
| 3      | third person  |
| ADJ    | adjective     |
| ASP    | aspect        |
| DAT    | dative        |
| DEM    | demonstrative |
| GEN    | genitive      |
| Mod    | modification  |
| NMLZ   | nominalizer   |
| NP     | noun phrase   |
| NUM    | number        |
| PL     | plural        |
| PRED   | predicate     |
| PROP.N | proper noun   |
| PRON   | pronoun       |
| SG     | singular      |

## REFERENCES

- Comrie, Bernard, and Sandra A. Thompson. 2007. Lexical nominalization. In Timothy Shopen. (ed.) *Language typology and syntactic description (2nd edition) Vol. III*, 334–381. Cambridge: Cambridge University Press.
- Gerner, Matthias. 2012. The Typology of Nominalization. *Language and Linguistics* 13(4), 803–844.
- Huang Yang. 2020. Zhabayu de Mingwuhua he Guanxihua [Nominalization and Relativization in Zaba Language]. *Minzu Yuwen [Minority languages Of China]*, Issue 4, 29–42.
- LaPolla, Randy J. 2008. Nominalization in Rawang. *Linguistics of the Tibeto-Burman Area* 31(2), 45–65.
- Li, Charles N., and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley & Los Angeles & London: University of California Press.
- Luo, Anyuan. 2005. *Songtao Miaohua Miaoxie Yufaxue [A descriptive grammar of Songtao Miao]*. Beijing: Minzu University of China Publishing House.
- Mao, Zongwu, and Yunbing Li. 2007. *Younuoyu Yanjiu [A study of Younuo]*. Beijing: The Ethnic Publishing House.
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani, (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52. [https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en).

- Xiang, Rizheng. 1999. *Jiwei Miaoyu Yanjiu [A study of Jiwei Miao]*. Chengdu: Sichuan Minzu Publishing House.
- Wu, Xiuju, and Yunsha Li. 2023. Xiangxi Miaoyu de Mingwuhua Shouduan [On nominalization strategies in the Xiangxi dialect of Miao]. *Minzu Yuwen [Minority languages Of China]* Issue 5, 55–64.
- Yang, Zaibiao. 2004. *Miaoyu Dongbu Fangyan Tuyu Bijiao [Comparative dialectology of Eastern Miao]*. Beijing: Minzu University of China Publishing House.
- Yap, Foong Ha, and Karen Grunow-Hårsta. 2010. Non-referential uses of nominalization constructions: Asian perspectives. *Language and Linguistics Compass* 4(12), 1154–1175.
- Yap, Foong Ha, Karen Grunow-Hårsta, and Janick Wrona (eds.). 2011. *Nominalization in Asian languages: Diachronic and typological perspectives*. Amsterdam: John Benjamins.

# So-called gender markers as classifying nominalizers in Swedish

Tatjana Ilic<sup>1</sup> and Kazuhiro Kawachi<sup>2</sup>

## ABSTRACT

This study examines so-called gender marking—more accurately, classifier-like marking—on grammatical nominalizations in their noun-phrase and modification uses in Swedish. It shows that the distribution of classifying grammatical nominalization markers in this language across different grammatical structures deviates from Shibatani’s (2021: 524; 2023: 35) proposed implicational hierarchy ‘NUM (Numerals) > DEM (Demonstratives) > GEN (Genitive noun phrases) > ADJ (Adjectives) > V-based NMNLs (Verb-based nominalizations)’. Swedish employs two strategies for marking with classifying grammatical nominalization markers: the use of distinct forms (or suffixes) and the use of different singular articles. With the former strategy, a distinction is made for demonstratives, but only to a limited extent or not at all for other types of elements. With the latter, marking occurs with demonstratives, adjectives, and verb-based nominalizations (relative clauses), but not with numerals or genitive noun phrases. This deviation observed for genitive noun phrases is hypothesized to result from the interaction between the identifiability of the possessum relative to the possessor and the redundancy of marking the possessum’s category when it is likely more identifiable than the possessor.

## 1. INTRODUCTION

Similar to other Indo-European languages, Swedish (North Germanic; spoken in Sweden and Finland) is often described as having a gender system. However, the Swedish system is not based on biological gender but instead resembles an animacy-based classifier system. This paper treats it more generally as a classification system and refers to the so-called gender markers as ‘classifying grammatical nominalization markers’ (classifying nominalizers, henceforth) (Shibatani 2023, Shibatani and Ferreira this issue). We investigate which types of grammatical structures take classifying nominalizers to derive grammatical nominalizations in Swedish. We first describe definiteness markers as classifying nominalizers in Swedish, which are the primary indicators of classification in this language. We then discuss the use of the definiteness markers on cardinal numerals, ordinal numerals, demonstratives, genitive noun phrases, adjectives, and relative clauses in both their noun-phrase and modification uses. We show that the definiteness markers are used with relative clauses—the lowest on Shibatani’s (2021; 2023) proposed implicational hierarchy for classifying nominalizers—but not with genitive noun phrases, which are ranked higher on the hierarchy. We argue that this deviation from the hierarchy (Shibatani 2021: 524; 2023: 35) results from the joint expression of definiteness and classification in Swedish. Specifically, we hypothesize that the redundancy of

---

<sup>1</sup> Faculty of Business and Commerce, Keio University, 4-1-1 Hiyoshi, Koohoku-ku, Yokohama, Kanagawa 223-8521, Japan. email: tatjana@keio.jp

<sup>2</sup> Faculty of Business and Commerce, Keio University, 4-1-1 Hiyoshi, Koohoku-ku, Yokohama, Kanagawa 223-8521, Japan. email: kawachi@keio.jp



definiteness marking—when the referent is already uniquely specified as definite—prevents the category from being expressed.

Although Swedish is not an inflectionally rich language, its noun phrases are relatively complex. The basic word order in Swedish is SVO and the typical word order within a noun phrase is determiner–adjective–noun. Noun phrase constituents are marked for classifying nominalizer-based category, number, and definiteness, and the expression of these categories is closely intertwined.

Swedish classifying nominalizers distinguish between two grammatical categories, traditionally called common (*utrum*) and neuter (*neutrum*). The formal distinction between these two categories is made using indefinite and definite markers (free articles, suffixes, or a combination of the two), which assume different forms for common (C) and neuter (N) categories. Each noun is assigned to one of the two categories. The category of a noun is indicated by the indefinite marker, which is a free prenominal article etymologically related to numeral ‘one’: *en* for common and *ett* for neuter category (e.g., *en hund* ‘a dog’ and *ett hus* ‘a house’). In indefinite noun phrases with adjectival modification, category information is also indicated on the adjective, in agreement with the referent of the noun phrase. In the following sections, we examine the use of classifying nominalizers with different types of grammatical elements in the derivation of grammatical nominalizations in Swedish and show how classifying nominalizer-based categories are marked differently, depending on the definiteness and number of the noun phrase.

Swedish has non-classifying nominalizers, which are affixes—specifically, prefixes and suffixes—involving lexical nominalization. Noun derivation through affixation is a very productive process that can produce complex derivations (e.g., *arbet-s-lös-het-s-för-säkr-ing* ‘unemployment insurance’). One of the most common suffixes is the deverbal suffix *-ing/-ning*, which is used to derive nouns indicating actions or processes from verbs denoting actions (e.g., *utveckla* ‘develop’ > *utveckl-ing* ‘development’ and *mobba* ‘bully’ > *mobb-ning* ‘bullying’). Another common one is the deadjectival suffix *-het*, which derives nouns denoting quality or status from adjectives (e.g., *svag* ‘weak’ > *svag-het* ‘weakness’ and *sjuk* ‘sick’ > *sjuk-het* ‘illness’). A common suffix deriving nouns from other nouns is the denominal suffix *-skap* (e.g., *vän* ‘friend’ > *vän-skap* ‘friendship’ and *fader* ‘father’ > *fader-skap* ‘paternity’). Nouns can also be derived through zero suffixation, usually from verbs, whereby a zero suffix is added to the stem of a verb following deletion of the infinitive ending *-a* (e.g., *fuska* ‘cheat’ > *fusk-ø* ‘cheating’ and *flyta* ‘flow’ > *flyt-ø* ‘fluency’). In some cases, nominalization may involve a change of the stem vowel (e.g., *lång* ‘long’ > *längd* ‘length’ and *tung* ‘heavy’ > *tyngd* ‘weight’). Table 1 summarizes some of the most productive deadjectival, deverbal, and denominal noun suffixes (adapted from Holmes and Hinchliffe 2013: 622–24).

Table 1. Some common non-classifying nominalization markers in Swedish

| Denotation | Suffix                                              | Stem: verbal (V),<br>adjectival (A),<br>nominal (N)                                                                                | Nominalization                                                                                                                                                                                                         |
|------------|-----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| people     | -are<br>-ande<br>-ende<br>-an<br>-är                | lära ‘teach’ (V)<br>studera ‘study’ (V)<br>gå ‘walk’ (V)<br>Amerika ‘America’ (N)<br>funktion ‘function’ (N)                       | lär-are ‘teacher’ (agent)<br>studer-ande ‘student’ (agent)<br>gå-ende ‘pedestrian’ (agent)<br>Amerik-an ‘American’ (nationality)<br>function-är ‘official’ (occupation)                                                |
| Activity   | -an<br>-ande<br>-ende<br>-else<br>-(n)ing<br>-ition | börja ‘begin’ (V)<br>skriva ‘write’ (V)<br>le ‘smile’ (V)<br>jämfora ‘compare’ (V)<br>skriva ‘write’ (V)<br>informera ‘inform’ (V) | börj-an ‘beginning’ (abstract)<br>skriv-ande ‘writing’ (ongoing; result)<br>le-ende ‘smile’ (result)<br>jämfor-else ‘comparison (abstract)<br>skriv-ning ‘writing’ (activity)<br>informa-tion ‘information’ (abstract) |
| Status     | -het<br>-lek<br>-dom<br>-skap                       | brottslig ‘criminal’ (A)<br>kär ‘dear’ (A)<br>sjuk ‘sick’ (A)<br>vän ‘friend’ (N)                                                  | brottslig-het ‘crime’ (status)<br>kär-lek ‘love’ (quality)<br>sjuk-dom ‘sickness’ (status)<br>vän-skap ‘friendship’ (condition)                                                                                        |

Some deverbal nouns derived from action verbs can take arguments and form noun phrases. Although both external argument (Arg1) and internal argument (Arg2) can be expressed as part of a deverbal nominal phrase, there is a cross-linguistic tendency to express only the internal argument (Arg2), while the external argument (Arg1) remains optional (Koptjevskaja-Tamm 1993). This holds for Swedish as well. The two main suffixes that derive deverbal action nominals in Swedish are *-ande/-ende* and *-(n)ing* (Table 2 above). When this type of deverbal nominal is derived from a transitive base, Arg1 appears in the genitive case, and Arg2 occurs inside a prepositional phrase introduced by the preposition *av* ‘of’.<sup>3</sup> The order of the arguments is the same as in a finite clause (e.g., FINITE: *Polis-en grep tjuven* ‘The police arrested the thief’, DEVERBAL NOMINAL: *Polis-en-s gripande av tjuven* ‘the police’s arrest of the thief/the arrest of the thief by the police’). With nominalizations derived from transitive verb bases, the external argument (Arg1) is commonly omitted (e.g., *Grip-ande av tjuven var dramatiskt* ‘The arrest of the thief was dramatic’).<sup>4</sup>

In general, however, deverbal noun phrases are much less common in Swedish than in English. Instead, compounding (or incorporating) is used. Compounding is a valency-reducing mechanism by which the internal argument (Arg2) becomes part of the compound, and the external argument (Arg1) is genitivized (or is omitted altogether) (e.g., FINITE: *barnen spelar*

<sup>3</sup> The preposition *av* ‘of’ is also used to introduce the agent in passive clauses.

<sup>4</sup> A far less common pattern is one in which the pre-posed genitive corresponds to the internal argument (Arg2) and the external argument (Arg1) is introduced by the preposition *av* ‘of’. This pattern follows the argument order of a passive sentence and is potentially ambiguous between the two interpretations and is therefore very rare (PASSIVE: *Post-en övertogs av banken* ‘The post was taken over by the bank’, DEVERBAL NOMINAL PHRASE: *Bank-en-s övertagande av posten* ‘the bank’s taking over by the post’).

*boll* ‘the children play ball’, COMPOUND: *barn-en-s bollspel-ande* ‘the children’s ball-playing’).

The data in this chapter are drawn primarily from the Swedish Academy Grammar Book (Teleman et al. 1999) and a comprehensive grammar of Swedish language (Holmes and Hinchliffe 2013), as well as from the first author’s consultants.

## 2. OVERVIEW OF THE USE OF CLASSIFYING GRAMMATICAL NOMINALIZATION MARKERS AND GRAMMATICAL NOMINALIZATIONS IN SWEDISH

The classification system in Modern Swedish developed from an older system inherited from Proto-Indo-European, which distinguished masculine, feminine, and neuter genders. During the Middle Ages, masculine and feminine genders merged, forming the present-day common category. As a result, almost two-thirds of Swedish nouns belong to the common category, while the remainder are neuter (Bohnacker 2003: 199). Although category assignment is generally considered arbitrary (e.g., Teleman 1987, Josefsson 2002, 2005, 2006, 2010, Källström 1995, 1996), recent studies emphasize the role of animacy in this process both synchronically (Kilarski and Krynicki 2005, Åkerblom 2012) and diachronically (Trosterud 2006, Van Epps et al. 2021, Skrzypek 2021). Semantic motivation for category assignment in Swedish involves a combination of animacy and individuation, following Åkerblom’s (2012: 74) Animacy Hierarchy (human > animate > inanimate > mass/abstract), with a clear tendency for more animate, individuated entities (especially humans) to be classified as common, and for inanimate, less individuated entities (especially those with mass, collective, or abstract denotations) to be classified as neuter (Åkerblom 2012, Skrzypek 2021).<sup>5</sup> Thus, the classification system in Swedish is better analyzed as an animacy-based classifier system. This exemplifies the difficulty in categorizing classifying grammatical nominalization markers strictly as either classifiers or gender markers (Craig 1986, Grinevald 2000, 2002, Grinevald and Seifart 2004, Shibatani 2023, Shibatani and Ferreira this issue). Accordingly, the present paper adopts the term ‘classifying grammatical nominalization markers’ (henceforth, classifying nominalizers) rather than ‘gender markers’, contrary to the traditional terminology applied to Swedish grammar. It avoids the terms ‘gender distinction’ and ‘gender categories’ and instead uses ‘common/neuter distinction’ and ‘common/neuter categories’.

### 2.1 Use of classifying grammatical nominalization markers in Swedish

The common/neuter category of Swedish nouns is marked using definiteness markers as classifying nominalizers. In singular noun phrases, the indefinite marker is a free prenominal article (*en* for common and *ett* for neuter), while the definite marker is a suffix (*-en* for common and *-et* for neuter). In plural noun phrases, indefiniteness is indicated by a plural suffix with five allomorphs that signal the category of the referent. Nouns are therefore classified into five declension types based on the allomorph of the plural suffix they take. The definite marker in plural noun phrases is also a suffix, which attaches to a stem consisting of the noun root plus the plural suffix, providing an overall definite interpretation. This suffix likewise has distinct

---

<sup>5</sup> Similar analyses of gender, based on the concepts of animacy and individuation, have been articulated by Luraghi (2011) and Van Epps et al. (2021).

allomorphs that indicate category distinctions.

The choice of the particular allomorph of the indefinite plural suffix depends primarily on the noun's category, but also on its phonological or morphological properties. Some forms of the suffix are exclusively used for common category nouns of the first and second declension type (*-or* and *-ar*, respectively), while another form is exclusively used for neuter category nouns of the fourth declension type (*-n*). There are also forms that can be used for either category with nouns of the third and fifth declension (*-er/-r* and *-ø*, respectively). Nevertheless, even in these cases, the category of the noun can easily be inferred. For example, most nouns in the third declension type belong to the common category, and those that do not are primarily loanwords with specific stress patterns and endings (e.g., *ett museum* (SG)/*muse-er* (PL) 'museum', *ett bageri* (SG)/*bageri-er* (PL) 'bakery'). Similarly, most nouns in the fifth declension type are neuter, and those that are not have a human denotation (e.g., *en lärare* (SG)/*lärare-ø* (PL) 'teacher', *en resande* (SG)/*resande-ø* (PL) 'traveler') (see Holmes and Hinchliffe 2013: 16–44 for details).

The plural definite suffix has fewer allomorphs, which likewise provide clues about the category of the noun that they attach to. The choice of a particular allomorph similarly depends on the noun's category as well as its phonological or morphological properties. The plural definite suffix form *-na* occurs almost exclusively with common category nouns, and *-a* and *-en* occur only with neuter category nouns (Holmes and Hinchliffe 2013: 56–65, 102–114).

In addition to nouns, category distinctions are also marked on pronouns, which take different forms depending on the category of the referent of the noun phrase (e.g., *någon bil* (indefinite), *ingen bil* (negative), *vilken bil* (interrogative; relative), *min bil* 'my car' (possessive) with nouns that belong to the common category, and *något hus* (indefinite); *inget hus* (negative), *vilket hus* (interrogative; relative) *mitt hus* 'my house' (possessive) with nouns that belong to the neuter category). The plural forms of all these pronouns do not mark category and invariably end in *-a* (e.g., *någr-a bilar/någr-a hus* (indefinite), *ing-a bilar/ing-a hus* (negative), *vilk-a bilar/vilk-a hus* (interrogative; relative), *min-a bilar/min-a hus-ø* (possessive)).

Demonstrative pronouns in singular noun phrases also have two different category forms, *den här/den där* 'this/that' for common and *det här/det där* for neuter (e.g., *den här bil-en* 'this car'/*den där bil-en* 'that car' (common) and *det här hus-et* 'this house'/*det där huset* 'that house' (neuter)). In plural noun phrases, demonstrative pronouns do not distinguish categories and use the same form regardless of the referent of the noun phrase, *de här/de där* 'these'/'those' (e.g., *de här bil-ar-na/de där bil-ar-na* 'these cars'/'those cars' and *de här hus-ø-en/de där hus-ø-en* 'these houses'/'those houses').

In noun phrases with adjectival modification, adjectives take different endings, depending on the definiteness and number of the noun phrase. In indefinite singular noun phrases, adjectives are marked following the so-called 'strong' inflectional pattern, which indicates category distinctions (*-ø* for common and *-t* for neuter). In all other cases, adjectives are marked following the so-called 'weak' inflectional pattern, which makes no category distinction and uses the suffix *-a* (Table 2).

Table 2. Definiteness, number, and common/neuter distinction in Swedish noun phrases with adjectival modification

| Category | Singular                        |                                      | Plural                               |                                            |
|----------|---------------------------------|--------------------------------------|--------------------------------------|--------------------------------------------|
|          | Indefinite                      | Definite                             | Indefinite                           | Definite                                   |
| Common   | en stor-ø bil<br>'a big car'    | den stor-a bil-en<br>'the big car'   | två stor-a bil-ar<br>'two big cars'  | de stor-a bil-ar-na<br>'the two big cars'  |
| Neuter   | ett stor-t hus<br>'a big house' | det stor-a hus-et<br>'the big house' | två stor-a hus-ø<br>'two big houses' | de stor-a hus-ø-en<br>'the two big houses' |

Notice that in definite noun phrases with adjectival modification, definiteness is marked twice, following the so-called 'double-definiteness' marking pattern, which involves a preposed definite article—an independent word homophonous with the distal demonstrative 'that'—in addition to the nominal definite suffix discussed above. The common/neuter category distinction on the preposed definite article is made only in singular noun phrases (*den* for common vs. *det* for neuter), while the plural form remains the same (e.g., *de*) regardless of the category of the referent (Table 3).<sup>6</sup>

Table 3. Definiteness and common/neuter distinction in double-definiteness marking pattern in Swedish

|         |        |     |   |          |
|---------|--------|-----|---|----------|
| Swedish |        |     | + | noun-DEF |
| SG      | common | den |   |          |
|         | neuter | det |   |          |
| PL      |        | de  |   |          |

Summing up, the common/neuter distinction in Swedish noun phrases is indicated by definiteness markers as classifying nominalizers, which are free articles, suffixes, and a combination of the two. In singular noun phrases, the indefinite marker is a free prenominal article, and the definite marker is a suffix. In plural noun phrases, both the plural/indefinite and definite markers are suffixes, each with multiple allomorphs indicating the common/neuter category of the noun. In definite noun phrases with adjectival modification, definiteness is marked with a combination of the free prenominal definite article and the definite suffix. This pattern is commonly referred to as double-definiteness marking. The common/neuter distinctions made by indefinite and definite markers in Swedish are summarized in Table 4.

<sup>6</sup> The double-definiteness marking also occurs in noun phrases with demonstrative pronouns, as illustrated in the previous paragraph (e.g., SG.C: *den här bil-en*; SG.N: *det här hus-et*; PL.C: *de här bil-ar-na*; PL.N: *de här hus-ø-en*).

Table 4. Common/neuter distinctions made by indefinite and definite markers in Swedish

| Noun phrases without adjectival modification |                   |                                            |                              | Noun phrases with adjectival modification |                                |                                            |                                   |
|----------------------------------------------|-------------------|--------------------------------------------|------------------------------|-------------------------------------------|--------------------------------|--------------------------------------------|-----------------------------------|
| Free article                                 | Suffix            |                                            |                              | Suffix                                    | Free article + suffix          | Suffix                                     | Free article + suffix             |
| Indefinite (SG)                              | Definite (SG)     | Indefinite (PL)                            | Definite (PL)                | Indefinite (SG)                           | Definite (SG)                  | Indefinite (PL)                            | Definite (PL)                     |
| C: en,<br>N: ett                             | C: -en,<br>N: -et | C: -or/-ar/<br>-er/-ø,<br>N: -er/-n/<br>-ø | C: -na,<br>N: -en/<br>-na/-a | C: en,<br>N: et                           | C:<br>den -e,<br>N:<br>det -et | C: -or/-ar/<br>-er/-ø,<br>N: -er/-n/<br>-ø | C: de -na,<br>N: de<br>-en/-na/-a |

## 2.2 Grammatical nominalizations in Swedish

### [1] Numerals

#### [1a] Cardinal numerals

Cardinal numerals make no common/neuter distinction (e.g., *två* ‘two’ in *två fågl-ar* ‘two birds’ (C)/*två hus-ø* ‘two houses’ (N); *tre* ‘three’ in *tre fågl-ar* ‘three birds’ (C) and *tre hus-ø* ‘three houses’), except for the numeral ‘one’, which has two category forms, one for common (*en*) and one for neuter (*ett*). These forms are homophonous with the two forms of the indefinite article in singular noun phrases (e.g., *en* ‘one’ in *en fågel* ‘one/a bird’ (C) and *ett* ‘one’ in *ett barn-ø* ‘one/a child’ (N)).

Higher numerals containing ‘one’ (e.g., *tjugoen* (C)/*tjuget* (N) ‘21’) make a limited use of the common/neuter distinction. Typically, the common form (*tjugoen*) is used regardless of the category of the noun that it modifies (e.g., *tjugoen fågl-ar* ‘twenty-one birds’ (C)/*tjugoen barn-ø* ‘twenty-one children’ (N), except with years, certain measure and monetary nominals, in which case the neuter form (*tjuget*) is used (e.g., *tjuget år-ø* ‘twenty-one years’ (N)). The list of cardinal numerals in Swedish is provided in Table 5.

Table 5. Cardinal numerals in Swedish

|    |        |    |         |    |              |
|----|--------|----|---------|----|--------------|
| 1  | en/ett | 11 | elva    | 21 | tjugoen/ett  |
| 2  | två    | 12 | tolv    | 22 | tjugoen två  |
| 3  | tre    | 13 | tretton | 23 | tjugoen tre  |
| 4  | fyra   | 14 | fjorton | 24 | tjugoen fyra |
| 5  | fem    | 15 | femton  | 25 | tjugoen fem  |
| 6  | sex    | 16 | sexton  | 26 | tjugoen sex  |
| 7  | sju    | 17 | sjutton | 27 | tjugoen sju  |
| 8  | åtta   | 18 | arton   | 28 | tjugoen åtta |
| 9  | nio    | 19 | nitton  | 29 | tjugoen nio  |
| 10 | tio    | 20 | tjugo   | 30 | trettio      |

## [1b] Ordinal numerals

Ordinal numerals are derived from cardinal numerals using the suffix *-de/-te* or *-nde*, occasionally involving sound changes in the stem or the suffix (e.g., *tre* ‘three’ → *tre-dje* ‘third’, *fyra* ‘four’ → *ffär-de* ‘fourth’, *fem* ‘five’ → *fem-te* ‘fifth’), as shown in Table 6. Ordinal numerals for the multiples of ten are formed using the suffix *-nde* (e.g., *tjugo-nde* ‘twentieth’, *trettio-nde* ‘thirtieth’). The first two ordinal numerals are exceptions, having special forms (e.g., *första* ‘first’ and *andra* ‘second’).

Table 6. Ordinal numerals in Swedish

|    |         |    |           |    |              |
|----|---------|----|-----------|----|--------------|
| 1  | första  | 11 | elfte     | 21 | tjugoförsta  |
| 2  | andra   | 12 | tolfte    | 22 | tjugoandra   |
| 3  | tredje  | 13 | trettonde | 23 | tjugotredje  |
| 4  | fjärde  | 14 | fjortonde | 24 | tjugofjärde  |
| 5  | femte   | 15 | femtonde  | 25 | tjugofemte   |
| 6  | sexte   | 16 | sextonde  | 26 | tjugosexte   |
| 7  | sjätte  | 17 | sjuttonde | 27 | tjugosjätte  |
| 8  | åttonde | 18 | artonde   | 28 | tjugoåttonde |
| 9  | nionde  | 19 | nittonde  | 29 | tjugonionde  |
| 10 | tionde  | 20 | tjugonde  | 30 | trettionde   |

## [2] Demonstratives

There are two basic types of demonstratives in Swedish: single-word demonstratives—including the formal pronoun *denna/detta/dessa* (proximal) and the short pronoun *den/det/de* (distal)—and two-word demonstratives, which come in two forms, *den här/det här/de här* for proximal referents and *den där/det där/de där* for distal ones. The common/neuter distinction is made only on demonstratives used with singular referents, while demonstratives used with plural referents do not distinguish common and neuter, regardless of the pronoun type (Table 7).

Table 7. Demonstrative pronouns in Swedish

| Number       | Singular (C) |           | Singular (N) |           | Plural      |           |
|--------------|--------------|-----------|--------------|-----------|-------------|-----------|
| Pronoun type | single word  | two words | single word  | two words | single word | two words |
| Proximal     | denna        | den här   | detta        | det här   | dessa       | de här    |
| Distal       | den          | den där   | det          | det där   | de          | de där    |

All demonstratives in Table 7 are inherently definite. The informal demonstratives *denna* and *detta* can be analyzed as containing the definite suffix *-en* (C)/*-et* (N). The two-word demonstratives (*den här*, *det här*, *de här*, etc.) contain the definite article *den* (C)/*det* (N)/*de* (PL). The short demonstratives *den/det/de* are homophonous with the personal pronouns *den/det* ‘it’ (3SG) and *de* ‘they’ (3PL), but differ from them by carrying an obligatory emphatic stress.

Because demonstratives are used for definite referents, they occur in definite noun phrases and cannot appear in indefinite noun phrases. In noun phrases with singular referents,

the common/neuter distinction is marked on the demonstrative itself as well as on the noun (following the double-definiteness marking pattern) (e.g., *den här bil-en* ‘this car’ (C) and *det här hus-et* ‘this house’ (N); *den bil-en* ‘that house’ (C) and *det hus-et* ‘that house’ (N)). An exception to this pattern is the formal demonstrative *denna/detta/dessa*, which occurs with bare nouns even though the noun phrase as a whole is definite (e.g., *denna bil* ‘this car’ (C) and *detta hus* ‘this house’ (N)).

### [3] Genitive noun phrases

Genitive noun phrases with pronominal possessors are marked for definiteness and the common/neuter distinction. Although the third-person genitive pronouns are invariant in form regardless of the category, the first-person and second-person genitive pronouns have distinct common and neuter forms, which indicate the category of the whole noun phrase. Like the demonstratives, these pronouns, at least the first-person and second-person singular ones, can be analyzed as containing the definite suffix. A similar analysis can be extended to the first-person and second-person plural pronouns *våran* (C)/*vårat* (N) and *eran* (C)/*erat* (N), which are used in colloquial Swedish (Holmes and Hinchliffe 2013: 163). Table 8 shows the forms of the genitive pronouns used for different person-number and common/neuter combinations.

Table 8. Pronominal possessors in Swedish

|                 | SG.C      | SG.N       | PL   |
|-----------------|-----------|------------|------|
| 1SG             | min       | mitt       | mina |
| 2SG             | din       | ditt       | dina |
| 3SG (F)         | hennes    |            |      |
| 3SG (M)         | hans      |            |      |
| 3SG (inanimate) | dess      |            |      |
| 1PL             | vår/våran | vårt/vårat | våra |
| 2PL             | er/eran   | ert/erat   | era  |
| 3PL             | deras     |            |      |

### [4] Adjectives

Adjectives cannot stand alone to form a noun phrase.

Adjectives carry different suffixes for common and neuter only with indefinite singular referents (e.g., INDEF.SG.C: *en stor-ø bil* ‘a big car’, INDEF.SG.N: *ett stor-t hus-ø* ‘a big house’, INDEF.PL: *stor-a bil-ar/stor-a hus-ø*). Adjectives derived from the perfect participle forms of verbs also decline with singular indefinite referents according to the number and the common/neuter category of the noun they modify (e.g., INDEF.SG.C: *en stängd-ø dörr* ‘a closed door’, INDEF.SG.N: *ett stäng-t fönster* ‘a closed window’, INDEF.PL: *stängd-a dörrar/stängd-a fönster-ø*). However, adjectives derived from present participle forms do not decline for common/neuter or number (e.g., INDEF.SG.C: *en leende kvinna* ‘a smiling woman’, INDEF.SG.N: *ett leende barn* ‘a smiling child’, INDEF.PL: *leende kvinn-or/leende barn-ø*).



## [5] Relative clauses

Swedish relative clauses are formed with the relative pronouns *som* (informal) and *vilken* (formal), employing the gap strategy. The relative pronoun *som* is used in spoken language, as well as in informal written language. It does not decline for common/neuter or number and can be used with both animate and inanimate referents. *Som* can function as a subject (e.g., *den pojke som går därborta* ‘the boy who is walking over there’), an object (e.g., *det pojke som vi såg igår* ‘the boy who we saw yesterday’), or a prepositional complement (e.g., *en soffa som är bekväm att sitta i* ‘a couch that is comfortable to sit on’), in which case the preposition is placed at the end of the clause.

In contrast to the informal relative pronoun *som*, which makes no common/neuter distinction, the formal relative pronoun *vilken* makes a common/neuter distinction in singular noun phrases (SG.C: *vilken*, SG.N: *vilke-t*), while its plural form always takes the suffix *-a* regardless the common/neuter category of the referent (PL: *vilk-a*) (e.g., *den rapport (C) vilken (C) nu inlämnats* ‘the report that has now been submitted’, *det underlag (N) vilke-t (N) nu inlämnats* ‘the document that has now been submitted’, *de report-er (PL.C) vilk-a (PL) nu inlämnats* ‘the reports that have now been submitted’). Both *som* and *vilken* can freely be omitted unless they function as the subject of the relative clause, in which case their use is obligatory.

When a relative clause is preceded by the definite article *den/det/de* and followed by a bare modified noun (i.e., a noun without the definite suffix), it is restrictive (e.g., *Den kusin som han hade mest kontakt med bodde i västra Nyland* ‘The cousin who he had the most contact with lived in western Nyland’). Non-restrictive relative clauses commonly occur without the prenominal definite article and with the modified noun bearing the definite suffix (e.g., *Kusin-en, som han ofta träffade, bodde i västra Nyland*. ‘The cousin, who he often met, lived in western Nyland’).<sup>7</sup>

## 3. DATA

The data discussed in this section reveal the following pattern of common/neuter marking on grammatical nominalizations with classifying nominalizers (Figure 1).

---

<sup>7</sup> It is also possible for a non-restrictive relative clause to follow the double-definiteness marking pattern, i.e., with the prenominal definite article and with a modified noun bearing the definite suffix, though this is less typical (e.g., *Den där bil-en, som du nog inte har sett förut, tillhör sågverket* ‘That car, which you have not seen before, I guess, belongs to the mill’). Relative clauses in indefinite noun phrases can be either restrictive or non-restrictive, depending on whether the noun phrase is interpreted as specific or not (e.g., *Jag vill ha en bil som jag kan lita på* ‘I want to have a car that I can trust’). The form of the modified noun phrase does not change regardless of whether the relative clause is interpreted as restrictive or not; the only difference is the use of the comma separating the non-restrictive relative clause (e.g., *Igår pratade jag med en student, som läser musik på Universitetet*).

Figure 1. Overview of the pattern of common/neuter marking with classifying grammatical nominalization markers in Swedish

|             |                                                                                                                 |       |          |       |        |
|-------------|-----------------------------------------------------------------------------------------------------------------|-------|----------|-------|--------|
| NP use:     | LMT('one')CARD.NUM >                                                                                            | DEM > | (*)GEN > | ADJ > | REL.CL |
|             | LMT(INDEF)ORD.NUM                                                                                               |       |          |       |        |
| Mod use:    | LMT('one')CARD.NUM >                                                                                            | DEM > | (*)GEN > | ADJ > | REL.CL |
|             | LMT(INDEF)ORD.NUM                                                                                               |       |          |       |        |
| *:          | no marking or impossible                                                                                        |       |          |       |        |
| (*):        | limited (marking applies only to first- and second-person possessor pronouns with singular possessum referents) |       |          |       |        |
| LMT('one'): | limited (marking applies only to 'one' and cardinal numerals containing 'one')                                  |       |          |       |        |
| LMT(INDEF): | limited (marking applies only to indefinite noun phrases)                                                       |       |          |       |        |
| unmarked:   | obligatory marking                                                                                              |       |          |       |        |

However, Figure 1 oversimplifies the ways in which the common/neuter distinction is marked. This distinction can be indicated through different common/neuter forms of grammatical nominalizations or different common/neuter suffix forms (Figure 2) or through different common/neuter forms of the singular articles (Figure 3). An asterisk in parentheses (\*) in Figures 1 and 2 indicates that the first- and second-person pronouns show a common/neuter distinction based on the category of the possessed noun. The use of demonstratives is restricted to definite referents. In contrast, common/neuter marking on ordinal numerals via different forms of the singular article (in Figure 3) is limited to indefinites, whereas common/neuter marking on adjectives and relative clauses using different common/neuter forms of the singular articles is possible with both indefinite and definite referents.

Figure 2. Pattern of marking with different common/neuter forms or suffixes in Swedish

|             |                                                                                                                 |       |          |        |         |
|-------------|-----------------------------------------------------------------------------------------------------------------|-------|----------|--------|---------|
| NP use:     | LMT('one')CARD.NUM >                                                                                            | DEM > | (*)GEN > | *ADJ > | *REL.CL |
|             | *ORD.NUM                                                                                                        |       |          |        |         |
| Mod use:    | LMT('one')CARD.NUM >                                                                                            | DEM > | (*)GEN > | *ADJ > | *REL.CL |
|             | *ORD.NUM                                                                                                        |       |          |        |         |
| *:          | no marking or impossible                                                                                        |       |          |        |         |
| (*):        | limited (marking applies only to first- and second-person possessor pronouns with singular possessum referents) |       |          |        |         |
| LMT('one'): | limited (marking applies only to 'one' and cardinal numerals containing 'one')                                  |       |          |        |         |
| unmarked:   | obligatory marking                                                                                              |       |          |        |         |

Figure 3. Pattern of marking with different common/neuter forms of singular articles in Swedish

NP use: \*CARD.NUM > DEM > \*GEN > ADJ > REL.CL  
 LMT(INDEF)ORD.NUM

Mod use: \*CARD.NUM > DEM > \*GEN > ADJ > REL.CL  
 LMT(INDEF)ORD.NUM

\*: no marking or impossible

LMT(INDEF): limited (marking applies only to indefinite noun phrases)

unmarked: obligatory marking

The major findings are as follows. First, the common/neuter distinction in cardinal and ordinal numerals is marked only in indefinite noun phrases. For cardinal numerals, this distinction appears only on the numeral ‘one’ and on numerals containing it. For ordinal numerals, the distinction is expressed through different forms of the indefinite article in the singular. Second, the common/neuter distinction in singular demonstratives, which invariably form definite noun phrases, appears only in definite contexts. Third, the common/neuter distinction is not made in genitive noun phrases with full-noun possessors, although it is indicated by the first- and second-person possessive pronouns for singular possessum referents. Finally, the common/neuter distinction is marked with different forms of the definite and indefinite articles for singular referents on adjectives and relative clauses both in their noun-phrase use (henceforth, NP use) and modification use.

## [1] Numerals

Numerals have both NP use and modification use.

### [1a] Cardinal numerals

In both their NP and modification uses, cardinal numerals make a common/neuter distinction only with the numeral ‘one’ or numerals containing it and only in noun phrases with indefinite referents.

#### [1a-1] Cardinal numerals: Indefinites

The numeral ‘one’ has two forms, *en* for common and *ett* for neuter, which are homophonous with the two forms of the indefinite article in singular noun phrases (e.g., (1), (2), (4), (5)). Noun phrases involving numerals higher than ‘one’ therefore occur without the indefinite article, and the common/neuter distinction is marked on the noun they modify (e.g., (6), (7)). However, in NP use, such numerals do not mark the common/neuter distinction (e.g., (3)).

#### NP use

(1) En kom på besök.  
 one.C come.PST on visit  
 ‘One (C) came to visit.’

(2) Ett kom på besök.  
 one.N come.PST on visit  
 ‘One (N) came to visit.’

(3) Tre kom på besök.  
 three come.PST on visit  
 ‘Three came to visit.’

#### Mod use

(4) En fågel kom på besök.  
 One.C bird come.PST on visit  
 ‘One (C) bird came to visit.’

(5) Ett barn kom på besök.  
 one.N child come.PST on visit  
 ‘One (N) child came to visit.’

(6) Tre fågl-ar kom på besök.  
 three bird-PL.C come.PST on visit  
 ‘Three birds (C) came to visit.’

(7) Tre barn-ø kom på besök.  
 three child-PL.N come.PST on visit  
 ‘Three children (N) came to visit.’

### [1a-2] Cardinal numerals: Definites

Definite noun phrases with numerals cannot take the definite suffix *-en/-et*, which is a nominal suffix (e.g., *\*en-en*, *\*två-en*).<sup>8</sup> Numerals occurring in definite noun phrases are therefore not marked for the common/neuter category (e.g., (8), (9), (10)). When the referent is definite, noun phrases involving numerals require the double-definiteness marking pattern—with the preposed definite article (SG.C: *den*, SG.N: *det*, PL: *de*) and with the definite suffix attached to the noun. Because the numeral ‘one’ is identical in form to the indefinite article *en/ett* and therefore cannot occur after the definite article *den/det* (e.g., *\*den en fågel-en/bil-en* to mean, ‘the one bird/car’ (C) and *\*det ett barn-et/bi-et* to mean, ‘the one child/bee’ (N)), the double-definiteness marking pattern is only possible in noun phrases with plural referents (e.g.,

<sup>8</sup> Only numerals that have been nominalized through *-a* suffixation can take the definite suffix (e.g., *en tre-a/trea-n*, meaning ‘the number three’).

*de tre barn-ø-en* (N)), where the common/neuter distinction is marked only on the noun (e.g., (9), (10)).

NP use

- (8) De tre kom på besök.  
 DEF.PL three come.PST on visit  
 'The three came to visit.'

Mod use

- (9) De tre fågl-ar-na kom på besök.  
 DEF.PL three bird-PL.C-PL.DEF.C come.PST on visit  
 'The (C) three birds (C) came to visit.'

- (10) De tre barn-ø-en kom på besök.  
 DEF.PL three child-PL.N-PL.DEF.N come.PST on visit  
 'The (N) three children (N) came to visit.'

### [1b] Ordinal numerals

Ordinal numerals have both NP use and modification use.

#### [1b-1] Ordinal numerals: Indefinites

Ordinal numerals can occur in indefinite noun phrases, in which case the indefinite article makes the common/neuter distinction in both their NP and modification uses, as shown in (11), (12), and (13).

- |      |            |                   |                 |                     |
|------|------------|-------------------|-----------------|---------------------|
| (11) | NP use     |                   | Mod use         |                     |
|      | en tredje  | 'a (C) third one' | en tredje bil   | 'a (C) third car'   |
|      | ett tredje | 'a (N) third one' | ett tredje barn | 'a (N) third child' |

NP use

- (12) Jag vill ha ett tredje.  
 I want have INDEF.N third.  
 'I want to have a (N) third (one).'

Mod use

- (13) Jag vill ha ett tredje barn.  
 I want have INDEF.N third child.SG  
 'I want to have a (N) third child.'

Ordinal numerals can occur in indefinite plural noun phrases, as in (14) and (15), although this is rare.

NP use

- (14) Hon har vunnit en andra plats och tre hedrande fjärde.  
She has won INDEF.C second place and three honorary fourth  
'She has won a (C) second place and three honorary fourth places.'

Mod use

- (15) Hon har vunnit en andra plats och tre hedrande fjärde  
She has won INDEF.C second place and three honorary fourth  
platser.  
place.PL.C  
'Over the years she has won a (C) second place and three honorary fourth places.'  
(Teleman et al. 1999 SAG2: 495)

### [1b-2] Ordinal numerals: Definites

Because ordinal numerals typically have a uniquely identifiable reference, they are predominantly used in definite noun phrases. They can occur in the definite-suffix marking pattern, but in both their NP use (e.g., (16) and (17)) and modification use (e.g., (18) and (19)), they are not marked for the common/neuter category. In modification use, the definite suffix occurs not on the ordinal numerals but on the noun.

NP use

- (16) första 'the first' tredje 'the third'
- (17) Bor han i (det) tredje hus-et härifrån? Eller  
live he in (DEF.N) third house-DEF.N from.here? Or  
femte?  
fifth?  
'Does he live in the (N) third house (N) from here? Or in the fifth (one)?'

Mod use

- (18) första fågel-n 'the first bird (C)' tredje fågel-n 'the third bird (C)'  
första hus-et 'the first house (N)' tredje hus-et 'the third house (N)'
- (19) Han bor i (det) tredje hus-et härifrån.  
He lives in (DEF.N) third house-DEF.N from.here  
'He lives in the (N) third house (N) from here.'

In the double-definiteness marking pattern, the common/neuter distinction is indicated on the definite article, both when the ordinal numeral is in NP use (e.g., (20) and (21)) and in modification use (e.g., (22) and (23); in addition to the definite suffix on the noun).

NP use

- (20) den (C) första ‘the (C) first’  
det (N) tredje ‘the (N) third’

- (21) Bor han i det tredje hus-et härifrån? Eller det  
live he in DEF.N third house-DEF.N from.here? Or DEF.N  
femte?  
fifth?  
‘Does he live in the (N) third house (N) from here? Or in the (N) fifth (one)?’

Mod use

- (22) den första fågel-n ‘the (C) first bird (C)’  
det tredje hus-et ‘the (N) third house (N)’
- (23) Han bor i (det) tredje hus-et härifrån.  
He lives in (DEF.N) third house-DEF.N from.here  
‘He lives in the (N) third house (N) from here.’

## [2] Demonstratives

Demonstratives have both NP use and modification use.

Singular demonstratives both in their NP use (demonstrative pronouns) and in their modification use (adnominal demonstratives) are marked for the common/neuter category. For example, the items in (24) can each serve as an independent noun phrase (e.g., the object of *Jag såg* ‘I saw’). When used with plural referents, demonstratives are not marked for the common/neuter category in either their NP or modification use.

- |      |                      |                                  |
|------|----------------------|----------------------------------|
| (24) | NP use               | Mod use                          |
|      | denna ‘this (C) one’ | denna fågel ‘this (C) bird’      |
|      | detta ‘this (N) one’ | detta hus ‘this (N) house’       |
|      | dessas ‘these ones’  | dessas fågl-ar ‘these birds (C)’ |

## [3] Genitive noun phrases

Genitive noun phrases with pronominal possessors in Swedish behave differently from those with full-noun possessors with respect to common/neuter marking and are therefore discussed separately. Genitive noun phrases with full-noun possessors are not marked for the common/neuter category or definiteness in either their NP or modification use. In contrast, genitive noun phrases with pronominal possessors are marked for the common/neuter category and definiteness in both their NP and modification uses. This applies only to singular possessums; with plural possessums, the common/neuter distinction is marked not on the pronominal genitive noun phrase but on the possessed noun.

Furthermore, there is a crosslinguistic tendency for a possessed noun to denote a definite entity and the referent of the noun phrase made up of the possessed noun and a possessor noun is usually definite (Haspelmath 1999). This holds true for Swedish as well.

Swedish genitive noun phrases have definite referents and therefore cannot be marked for indefiniteness, regardless of whether they have full-noun possessors or pronominal possessors.

### [3a] Full-noun possessors

The full-noun possessor noun phrase precedes the possessed noun and takes the genitive suffix *-s*, while the possessed noun remains unmarked. Any definiteness markers that occur in the genitive noun phrase, such as preposed indefinite and definite articles, indicate the common/neuter category of the possessor noun, rather than that of the possessed noun. A full-noun possessor noun phrase has both NP and modification uses. However, it does not indicate the common/neuter category (or the (in)definiteness) of the entire noun phrase in either use, as in (25), where the common/neuter markings, both indefinite and definite, reflect the common/neuter category of the possessor noun.

#### (25) NP use

|                                   |                  |
|-----------------------------------|------------------|
| ett (INDEF.N) barn-s (-POSS)      | ‘a child’s’      |
| barn-et-s (-DEF.SG.N-POSS)        | ‘the child’s’    |
| barn-ø-s (-PL.N-POSS)             | ‘children’s’     |
| barn-ø-en-s (-PL.N-DEF.PL.N-POSS) | ‘the children’s’ |

#### Mod use

|                                                |                         |
|------------------------------------------------|-------------------------|
| ett (INDEF.N) barn-s (-POSS) skola (C)         | ‘a child’s school’      |
| barn-et-s (-DEF.N-POSS) hus (C)                | ‘the child’s house’     |
| barn-ø-s (-PL.N-POSS) skola (SG.C)             | ‘children’s school’     |
| barn-ø-en-s (-PL.N-DEF.PL.N-POSS) skola (SG.C) | ‘the children’s school’ |

A full-noun possessor noun phrase may take the definite article (*den/det/de*), but again, the article does not mark the definiteness of the whole noun phrase; rather, it indicates the definiteness of the possessor noun, as in (26).

#### (26) NP use

|                              |
|------------------------------|
| det (DEF.N) lilla            |
| barn-et-s (-DEF.SG.N-POSS)   |
| ‘the (N) little child’s (N)’ |

#### Mod use

|                                         |
|-----------------------------------------|
| det (DEF.N.) lilla                      |
| barn-et-s (-DEF.SG.N-POSS) skola (C)    |
| ‘the (N) little child’s (N) school (C)’ |

In genitive noun phrases with plural referents in modification use, the common/neuter category of the possessed noun is marked by the plural suffix, but not by the possessor noun, as in (27).

#### (27) NP use

|                            |
|----------------------------|
| barn-et-s (-DEF.SG.N-POSS) |
| ‘the child’s (N)’          |

#### Mod use

|                                         |
|-----------------------------------------|
| barn-et-s (-DEF.SG.N) leksak-er (-PL.C) |
| ‘the child’s (N) toys (C)’              |



### [3b] Pronominal possessors

In contrast to genitive noun phrases with full-noun possessors, genitive noun phrases with pronominal possessors are marked for both definiteness and the common/neuter category. The first- and second-person genitive pronouns (Table 8) in both their NP and modification uses make the common/neuter distinction for singular possessum referents. With plural referents, the common/neuter category is marked on the noun when it is modified by these pronouns, as in (28).

| (28) | NP use |             | Mod use              |                        |
|------|--------|-------------|----------------------|------------------------|
|      | min    | ‘mine (C)’  | min fågel            | ‘my (C) bird’          |
|      | mitt   | ‘mine (N)’  | mitt hus             | ‘my (N) house’         |
|      | mina   | ‘mine (PL)’ | mina fågl-ar (-PL.C) | ‘my (PL) birds (PL.C)’ |

Because pronominal possessors cannot co-occur with the definite article, they cannot appear in the double-definiteness marking pattern (e.g., \**det min fågel*).

### [4] Adjectives

Adjectives are marked for the common/neuter category in both their NP and modification uses when the referent is an indefinite singular. In the plural, the common/neuter distinction is marked on the noun when it is modified by an adjective. When the referent is definite, adjectives indicate the common/neuter category only in their NP use.

#### [4a] Adjectives: Indefinites

When an indefinite noun phrase formed with an adjective is singular, the adjective in either its NP or modification use follows the indefinite article *en* (common)/*ett* (neuter) and takes the suffix *-ø* (common)/*-t* (neuter) (following the strong marking pattern). The indefinite article is repeated after the adjective in NP use, and the noun does not receive common/neuter marking.

When the resulting noun phrase is plural, a free article *ena* occurs before the adjective in NP use and is also repeated after it, whereas this article never occurs when the adjective is used in modification. The adjective in both NP and modification use receives the plural suffix *-a* and does not indicate the common/neuter category, although a plural noun modified by an adjective does. Examples are given in (29).

| (29) | NP use         |                     | Mod use          |                       |
|------|----------------|---------------------|------------------|-----------------------|
|      | en stor-ø en   | ‘a (C) big (C) one’ | en stor-ø fågel  | ‘a (C) big (C) bird’  |
|      | ett stor-t ett | ‘a (N) big (N) one’ | ett stor-t hus   | ‘a (N) big (N) house’ |
|      | ena stor-a ena | ‘big ones’          | ø stor-a fågl-ar | ‘big birds (C)’       |

#### [4b] Adjectives: Definites

As shown in (30), for definite referents, adjectives modifying a noun take the definite adjectival suffix *-a*, which does not indicate the common/neuter category. (This is the same as the suffix *-a* used to mark plural in indefinite phrases, as in (29).) In definite phrases, adjectives

are almost always preceded by the definite article, which makes the common/neuter distinction in the singular (*den* for common and *det* for neuter), but not in the plural (*de*). The noun carries the definite suffix, resulting in the double-definiteness marking pattern.

Mod use

- |      |                               |                         |
|------|-------------------------------|-------------------------|
| (30) | den stor-a fågel-n/*fågel     | ‘the (C) big bird (C)’  |
|      | det stor-a hus-et/*hus        | ‘the (N) big house (N)’ |
|      | de stor-a fågl-ar-na/*fågl-ar | ‘the big birds (C)’     |

In NP use, the common/neuter distinction is made only on the definite article in the singular, as in (31). In modification use, due to the double-definiteness marking, the common/neuter category is marked both on the preposed singular definite article (*den/det*) and on the definite suffix of the noun. With plural referents, the category distinction is indicated only on the noun, as the plural definite article *de* does not distinguish between the common and neuter categories.

- |      |                    |                   |                      |                         |
|------|--------------------|-------------------|----------------------|-------------------------|
| (31) | NP use             |                   | Mod use              |                         |
|      | den stor-a/*stor-ø | ‘the (C) big one’ | den stor-a/*stor-ø   | ‘the (C) big bird (C)’  |
|      |                    |                   | fågel-n              |                         |
|      | det stor-a/*stor-t | ‘the (N) big one’ | det stor-a/*stor-t   | ‘the (N) big house (N)’ |
|      |                    |                   | hus-et               |                         |
|      | de stor-a          | ‘big ones’        | de stor-a fågl-ar-na | ‘the big birds (C)’     |

However, the definite article and the adjective in modification use do not seem to form a single constituent in the way that they do in NP use. For example, in cases when the referent is uniquely identifiable, definite noun phrases with adjectival modification occur without the definite article (e.g., *Vit-a hus-et* ‘the White house’, *Röd-a hav-et* ‘the Red Sea’, *svensk-a språk-et* ‘the Swedish language’, *högr-a hand-en* ‘the right hand’, *förra år-et* ‘last year’, *först-a gång-en* ‘the first time’, etc.). The definite article therefore seems to mark the whole structure comprising the adjective and the modified noun (e.g., [*det* [*vit-a hus-et*], rather than [[*det vit-a hus-et*]). Hence, we conclude that the adjectives in definite noun phrases do not indicate the common/neuter category in their modification use.

## [5] Relative clauses

Relative clauses occur in both NP and modification uses in indefinite noun phrases. The common/neuter distinction is made only in the singular.

### [5a] Relative clauses: Indefinites

Relative clauses accompanied by the indefinite article, which indicates the indefiniteness of the noun phrase, occur in both NP use (e.g., (32)) and modification use (e.g., (33)), in both singular and plural. The common/neuter distinction is made only with singular referents; in the plural, the common/neuter category is indicated on the noun modified by the relative clause in modification use.

NP use

- |      |                    |                      |
|------|--------------------|----------------------|
| (32) | en som jag känner  | ‘one (C) who I know’ |
|      | ett som jag känner | ‘one (N) who I know’ |
|      | två som jag känner | ‘two who I know’     |

Mod use

- |      |                             |                            |
|------|-----------------------------|----------------------------|
| (33) | en flicka som jag känner    | ‘a (C) girl who I know’    |
|      | ett barn som jag känner     | ‘a (N) child who I know’   |
|      | två flick-or som jag känner | ‘two girls (C) who I know’ |

### [5b] Relative clauses: Definites

Relative clauses in noun phrases with definite referents are commonly preceded by the definite article, both in their NP and modification use. The common/neuter distinction is indicated on the definite article in singular noun phrases, and additionally on the noun modified by the relative clause in modification use. Examples are given in (34), (35), and (36).

NP use

- |      |                    |                          |
|------|--------------------|--------------------------|
| (34) | den som jag känner | ‘the (C) one who I know’ |
|      | det som jag känner | ‘the (N) one who I know’ |
|      | de som jag känner  | ‘the ones who I know’    |

Mod use (restrictive meaning)<sup>9</sup>

- |      |                                        |                                |
|------|----------------------------------------|--------------------------------|
| (35) | den flicka/flicka-n som jag känner     | ‘the (C) girl (C) who I know’  |
|      | det barn/barn-et som jag känner        | ‘the (N) child (N) who I know’ |
|      | de flick-or/flick-or-na som jag känner | ‘the girls (C) who I know’     |

Mod use (non-restrictive meaning)

- |      |                                 |                             |
|------|---------------------------------|-----------------------------|
| (36) | flicka-n/*flicka som jag känner | ‘the girl (C), who I know’  |
|      | barn-et/*barn som jag känner    | ‘the child (N), who I know’ |
|      | flick-or-na som jag känner      | ‘the girls (C), who I know’ |

## 4. ANALYSIS AND DISCUSSION

The data discussed in this paper show that common/neuter marking in grammatical nominalizations in Swedish closely interacts with the definiteness of the noun phrase referent. The effects of this interaction are evident in the patterns of common/neuter marking on cardinal numerals, demonstratives, and genitive noun phrases. In the case of cardinal numerals, it seems straightforward to assume that the restriction of common/neuter marking to noun phrases with indefinite referents stems from the fact that the numeral ‘one’, which is the only numeral marked for the common/neuter category, is also the diachronic source of the indefinite article and is formally identical to it. As a result, ‘one’ can only occur in noun phrases with indefinite

---

<sup>9</sup> In the modification use, the noun usually does not take the definite suffix, but it can carry it in casual speech.

referents, and the common/neuter distinction in cardinal numerals with definite referents is consequently not marked.

With respect to the restriction of common/neuter marking to demonstratives in definite noun phrases, it is important to note, first, that the primary function of demonstrative pronouns is to point out specific entities, and second, that Swedish demonstratives can be analyzed as containing definite markers in their singular forms (e.g., *denna/detta*, etc.). They can therefore be viewed as inherently definite, and thus incompatible with indefinite referents.

Moreover, Swedish cannot use either definiteness marker to indicate the common/neuter category of the possessed noun (only that of the possessor noun) in genitive noun phrases with full-noun possessors. This inability to mark the common/neuter distinction on genitive noun phrases constitutes a gap in the common/neuter marking of grammatical nominalizations in Swedish, which otherwise aligns with the implicational hierarchy of gender and classifier marking on grammatical nominalizations proposed in Shibatani (2021: 524; 2023: 35). We interpret this gap in light of a crosslinguistic tendency for possessed nouns to denote definite entities and for genitive noun phrases to have overall definite reference (Haspelmath 1999). Specifically, we argue that this gap results from a redundancy in definiteness marking: since possessed nouns are already uniquely specified as definite by the genitive possessor, marking them again as definite—and thus for common/neuter—would be redundant.

Nevertheless, the common/neuter distinction is made by first- and second-person genitive pronouns, but not by third-person genitive pronouns, in Swedish. This suggests that the identifiability of the possessum is relative to that of the possessor. When the possessor is first or second person, it is always identifiable to the discourse participants, and the possessum may be less so, making the marking of its category informative. In contrast, when the possessor is third person, it can be unidentifiable relative to the possessum. Therefore, the interaction between the identifiability of the possessor and that of the possessum appears to be a factor in the common/neuter marking of genitive noun phrases.

Finally, the common/neuter distinction is obligatorily made by one of the definite markers when adjectives are used to form noun phrases in Swedish, but not in their modification use. As discussed in the previous section, the data suggest that the definite article and the adjective in modification use do not form a single constituent in the way they do in NP use. We concluded that adjectives in definite noun phrases indicate common/neuter in their NP use, but not in their modification use, supporting Shibatani and Shigeno's (2013) hypothesis regarding the development of nominalizations from their NP to modification use.

## 5. CONCLUSION

In this paper, we discussed the effects of the joint expression of the common/neuter classification and definiteness on the use of classifying grammatical nominalization markers (classifying nominalizers) on grammatical nominalizations in Swedish. We demonstrated that common/neuter marking on grammatical nominalizations in Swedish deviates from the proposed hierarchy of marking with classifying nominalizers (Shibatani 2021, 2023) and attributed this deviation to the influence of definiteness marking. Further research into the diachronic development of marking on grammatical nominalizations with classifying

nominalizers could provide additional insights into the proposed directionality of nominalization development from their NP to modification use (Shibatani and Shigeno 2013). A comparative analysis of data from other Scandinavian languages could also help identify the pattern of marking with classifying nominalizers that emerge due to language-specific factors. Further investigation into the relationship between classification and definiteness is needed, both in languages like Swedish, where the two categories are expressed jointly, and in languages where they are expressed independently.

## ABBREVIATIONS

|             |                      |
|-------------|----------------------|
| 1           | first person         |
| 2           | second person        |
| 3           | third person         |
| ADJ         | adjective            |
| C           | common category      |
| CARD.NUM    | cardinal numeral     |
| DEF         | definite             |
| DEM         | demonstrative        |
| GEN         | genitive noun phrase |
| GEN.NP      | genitive noun phrase |
| INDEF       | indefinite           |
| N           | neuter category      |
| NMNL        | nominalization       |
| NUM         | numeral              |
| ORD.NUM     | ordinal numeral      |
| PL          | plural               |
| POSS        | possessive           |
| REL, REL.CL | relative clause      |
| SG          | singular             |

## ACKNOWLEDGMENTS

We would like to acknowledge the first author's Swedish native-speaker consultants, Viking Ericsson and Alexandar Medak, for providing data on Swedish. We would also like to express our deepest gratitude to Masayoshi Shibatani for his insightful comments on the presentation of this study, which we gave at our research group meeting on October 28, 2023. Any remaining shortcomings are, of course, the authors' alone. The present study has been supported by grants from the Japan Society for the Promotion of Science (KAKENHI Project IDs: 22H00659 and 23K21931, 'Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers', PI: Kazuhiro Kawachi) and the National Institute for Japanese Language and Linguistics ('Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspectives', PI: Prashant Pardeshi) as well as the Keio Gijuku Fukuzawa Memorial Fund for the Advancement of Education and Research, the Keio University Academic Development Funds, and the Ushioda Memorial Fund.

## REFERENCES

- Åkerblom, Sandy. 2012. Genus – tilldelat eller inherent? In Gunlög Josefsson (ed.) *Perspektiv på genus*, Nordlund 32, 65–100. Lund: Nordiska språk, Språk- och litteraturcentrum, Lunds universitet.
- Bohnacker, Ute. 2003. Nominal phrases. In Gunlög Josefsson, Christer Platzak, and Gisela Håkansson. (eds.) *The acquisition of Swedish grammar*, 195–260. Amsterdam: John Benjamins.
- Craig, Colette G. 1986. Jacaltec noun classifiers: A study in grammaticalization. *Lingua* 70.4, 241–284.
- Grinevald, Colette. 2000. A morpho-syntactic typology of classifiers. In Gunter Senft. (ed.) *Systems of nominal classification*, 50–92. Cambridge University Press.
- Grinevald, Colette. 2002. Making sense of nominal classification systems: Noun classifiers and the grammaticalization variable. In Ilse Wischer and Gabriele Diewald. (eds.) *New reflections on grammaticalization*, 259–275. Amsterdam: Benjamins.
- Grinevald, Colette. 2024. The ‘determinatives’ of Ancient Egyptian hieroglyphs are ‘root classifiers’. How egyptology and linguistics met to show it. In Niv Allon and Racheli Shalomi-Hen. (eds.) *Signs of life: Ancient Egyptian script, language, and writing studies in honour of Orly Goldwasser*, 63–74. Turnhout, Belgium: Brepols.  
<https://doi.org/10.1484/M.MRE-EB.5.137005>
- Grinevald, Colette, and Frank Seifart. 2004. Noun classes in African and Amazonian languages: Towards a comparison. *Linguistic Typology* 8, 243–285.
- Haspelmath, Martin. 1999. Explaining article-possessor complementarity: Economic motivation in noun phrase syntax. *Language* 75, 227–243.
- Holmes, Philip, and Ian Hinchliffe. 2013 [1994]. *Swedish: A comprehensive grammar, third edition*. New York, NY: Routledge.
- Josefsson, Gunlög. 2002. *Tjänis, knäppis och brallis; om suffixet -is i modern svenska*. Ms. Institutionen för nordiska språk, Lunds Universitet.
- Josefsson, Gunlög. 2005. *Ord*. Lund: Studentlitteratur.
- Josefsson, Gunlög. 2006. Semantic and grammatical genders in Swedish: Independent but interacting dimensions. *Lingua* 116, 1346–1368.
- Källström, Roger. 1995. *Om svenskans genusystem*. Göteborg: Göteborgs Universitet.
- Källström, Roger. 1996. On gender assignment in Swedish. In Kjartan G. Ottósson, Ruth V. Fjeld, and Arne Torp (eds.) *Proceedings of the Ninth International Conference of Nordic and General Linguistics, University of Oslo, January 10–11, 1995*, 151–167. Oslo: Novus Forlag.
- Kilarski, Marcin, and Grzegorz Krynicki. 2005. Not arbitrary, not regular: The magic of gender assignment. In Dirk Geeraerts, Johan van der Auwera, and Nicole Delbecque. (eds.) *Perspectives on Variation: Sociolinguistic, Historical, Comparative*, 235–250. Berlin: De Gruyter Mouton.
- Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.
- Luraghi, Silvia. 2021. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45.2, 435–464.

- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)
- Shibatani, Masayoshi, and Hiromi Shigeno. 2013. Amami nominalizations. *International Journal of Okinawan Studies* 7, 107–139.
- Shibatani, Masayoshi, and Marília Ferreira. This issue. Dynamic functional typology of Tukanoan classifying nominalizations.
- Skrzypek, Dominika. 2021. Animacy, anaphora and language change in Swedish. *GEMA Online Journal of Language Studies* 21.3: 1–15.  
<http://dx.doi.org/10.17576/gema-2021-2103-01>
- Teleman, Ulf. 1987. *Grammatik på villovägar*. Stockholm: Almqvist & Wiksell.
- Teleman, Ulf, Staffan Hellberg, and Erik Andersson. 1999. *Svenska akademiens grammatik (SAG) I-4*. Stockholm: Svenska Akademien.
- Trosterud, Trond. 2006. Gender assignment in Old Norse. *Lingua* 116., 1441–1463.
- Van Epps, Briana, Gerd Carling, and Yair Sapir. 2021. Gender assignment in six North Scandinavian languages: Patterns of variation and change. *Journal of Germanic Linguistics* 33.3, 264–315.

# Distribution of gender marking in Slovene

Prashant Pardeshi<sup>1</sup> and Andrej Bekeš<sup>2</sup>

## ABSTRACT

In this paper, we will offer a preliminary report on the distribution of gender marking on: (a) numerals, (b) demonstratives, (c) genitives, (d) adjectives, and (e) the so-called relative clauses, in Slovene (South Slavic, Indo-European) spoken in Slovenia and adjacent areas of neighboring countries, in their two usages: (i) NP-use, where these structures are used as the head of a noun phrase, and (ii) modification-use, where these structures are used as modifiers of a noun, adopting the nominalization framework proposed in Shibatani (2017, 2018, 2019, 2021, 2023). Following Shibatani, we will analyze gender marking as a nominalization device, which not only nominalizes the structure to which it is attached, but also classifies it into one of three groups, masculine, feminine, and neuter. Finally, we will show that the distribution of gender marking in Slovene supports Shibatani's hierarchy.

## 1. INTRODUCTION

In this paper, we will offer a preliminary report on the distribution of gender marking on: (a) numerals, (b) demonstratives, (c) genitives, (d) adjectives, and (e) the so-called relative clauses, in Slovene (South Slavic, Indo-European) spoken in Slovenia and adjacent areas of neighboring countries, in their two usages: (i) NP-use, where these structures are used as the head of a noun phrase, and (ii) modification-use, where these structures are used as modifiers of a noun, adopting the nominalization framework proposed in Shibatani (2017, 2018, 2019, 2021, 2023). Following Shibatani, we will analyze gender marking as a nominalization device, which not only nominalizes the structure to which it is attached, but also classifies it into one of three groups, masculine, feminine, and neuter. Finally, we will show that the distribution of gender marking in Slovene supports Shibatani's hierarchy. The data presented in the paper is based on the native intuition of the second author and further supplemented with attested data in Standard Slovene Dictionary (SSKJ) (in Slovene) available at: <http://bos.zrc-sazu.si/sskj.html>

Nouns or nominals in Slovene have three characteristic features, (i) gender, (ii) number, and (iii) case. Nouns/nominals in Slovene are divided into three genders, masculine (M), feminine (F), and neuter (N). Further, nouns/nominals in Slovene inflect for three numbers, singular, dual, and plural and inflect for case, nominative, genitive, dative, accusative, locative, and instrumental. These features serve as an indicator for assessing whether or not a structure is nominal in Slovene. Furthermore, it should be noted that pronouns and adjectives also undergo change in their form according to gender and number of their denotation and also inflect for cases like nouns. Unlike some widely spoken Western European languages,

---

<sup>1</sup> National Institute for Japanese Language and Linguistics, Tokyo, Japan. Email: prashant@ninjal.ac.jp

<sup>2</sup> The University of Ljubljana, Slovenia



Slovene has a very few classifiers that can arguably be classified as ‘sortal’ classifiers (e.g. *tri glave govedi* (three heads of cattle)), *en par čevljev* ‘a pair of (the) shoes,’ like Japanese or Chinese, but has many ‘mensural’ classifiers (aka measure words), which are found in various languages, such as: *kos* ‘piece’, *en kos kruha* ‘a piece of bread’; *kozar(ec)* ‘glass’, *en kozarec vode* ‘a glass of water’; *skodelica* ‘cup’, *ena skodelica čaja* ‘a cup of tea’; *steklenica* ‘bottle’, *ena steklenica vina* ‘a bottle of wine’; *žlica* ‘spoonful’, *ena skleda juhe* ‘a bowl of soup’ etc.

- b. kupil                      bom                      {enega/eno}  
 buy.PSTPTCP.MSG    be.1SG.FUT    one.MSG.ACC/one.FSG.ACC  
 'I will buy {one (MSG)/one (FSG)}.'

### [1b] Ordinal numerals

Ordinal numerals in Slovene decline as per the gender, number and case of their denotation. The following table gives the ordinal numbers in the nominative case form.

Table 2. Cardinal numbers in Slovene

| Number           | Masculine | Feminine | Neuter |
|------------------|-----------|----------|--------|
| 1st              | prvi      | prva     | prvo   |
| 2nd              | drugi     | druga    | drugo  |
| 3rd <sup>r</sup> | tretji    | tretja   | tretje |
| 4th              | četrti    | četrtā   | četrtō |
| 5th              | peti      | peta     | peto   |
| 6th              | šesti     | šesta    | šesto  |
| 7th              | sedmi     | sedma    | sedmo  |
| 8th              | osmi      | osma     | osmo   |
| 9th              | deveti    | deveta   | deveto |
| 10th             | deseti    | deseta   | deseto |

Masculine (e.g., pineapple):

- (2) všeč      mi      **je**      **bil**                      **drugi**      z      leve  
 pleasing    1SG.DAT   be.3PRS   PSTPTCP.3MSG   second.MSG   from   left  
 'I liked the second one (MSG) from the left.'

Feminine (e.g., banana):

- (3) všeč      mi      **je**      **bila**                      **druga**      z      leve  
 pleasing    1SG.DAT   be.3PRS   PSTPTCP.3FSG   second.FSG   from   left  
 'I liked the second one (FSG) from the left.'

Neuter (e.g., apple)

- (4) všeč      mi      **je**      **bito**                      **drugo**      z      leve  
 pleasing    1SG.DAT   be.3PRS   PSTPTCP.3NSG   second.NSG   from   left  
 'I liked the second one (NSG) from the left.'

## [2] Demonstratives

Demonstrative pronouns in Slovene are as summarized in Table 3 below.

Table 3. Demonstratives in Slovene

|                                                 | M     |                |       | F     |             |       | N     |             |       |
|-------------------------------------------------|-------|----------------|-------|-------|-------------|-------|-------|-------------|-------|
|                                                 | SG    | DU             | PL    | SG    | DU          | PL    | SG    | DU          | PL    |
| Proximal                                        | ta    | ta<br>(dva)    | ti    | ta    | ti<br>(dve) | te    | to    | ti<br>(dve) | ta    |
| Medial<br>(that)                                | tisti | tista<br>(dva) | tisti | tista | tisti       | tiste | tisto | tisti       | tista |
| Distal<br>(that,<br>over<br>there, far<br>away) | oni   | ona<br>(dva)   | oni   | ona   | oni         | one   | ono   | oni         | ona   |

- (5) a. **tisti** so dragi [NP-use]  
 DIAT.MPL be.PRS.3PL expensive.MPL  
 ‘Those (MPL) are expensive.’
- b. **tisti** ananasi so dragi [Mod-use]  
 DIST.MPL pineapples.MPL be.PRS.3PL expensive.MPL  
 ‘Those pineapples (MPL) are expensive.’
- (6) a. **tiste** so drage [NP-use]  
 DIST.FPL be.PRS.3PL expensive.FPL  
 ‘Those (FPL) are expensive.’
- b. **tiste** banane so drage [Mod-use]  
 DIST.FPL bananas.FPL be.PRS.3PL expensive.FPL  
 ‘Those bananas (FPL) are expensive.’
- (7) a. **tista** so draga [NP-use]  
 DIST.NPL be.PRS.3PL expensive.NPL  
 ‘Those (NPL) are expensive.’
- b. **tista** jabolka so draga [Mod-use]  
 DIST.NPL apples.NPL be.PRS.3PL expensive.NPL  
 ‘Those apples (NPL) are expensive.’

## [3] Genitives

In Slovene, possessive forms and genitive forms of personal pronouns as well as general nouns are different. The possessive forms of personal pronouns express the person, gender,

number and case of the possessor and the gender, number and case of the possesee. The following table illustrates the forms of possessive pronouns only for possessor in singular in the nominative case.

Table 2. Possessive pronouns

| Possessor<br>(Singular) | Possessee |         |         |         |         |         |         |         |         |
|-------------------------|-----------|---------|---------|---------|---------|---------|---------|---------|---------|
|                         | M         |         |         | F       |         |         | N       |         |         |
|                         | Sg        | Du      | Pl      | Sg      | Du      | Pl      | Sg      | Du      | Pl      |
| 1 (M/F/N)               | moj       | moja    | moji    | moja    | moji    | moje    | moje    | moji    | moja    |
| 2 (M/F/N)               | tvoj      | tvoja   | tvoji   | tvoja   | tvoji   | tvoje   | tvoje   | tvoji   | tvoja   |
| 3M/N                    | njegov    | njegova | njegovi | njegova | njegovi | njegove | njegovo | njegovi | njegova |
| 3F                      | njen      | njena   | njeni   | njena   | njeni   | njene   | njeno   | njeni   | njena   |

- (8) a. To je [NP-use]  
 this.one.here.FSG.NOM be.PRS.3SG  
**moja**  
 POSS.1.SG.NOM+POSSEE.F.SG.NOM  
 ‘This one here is **mine**.’
- b. To je [Mod-use]  
 the one here.FSG.NOM be.PRS.3SG  
**moja** **knjiga**  
 POSS.1.SG.NOM+POSSEE.F.SG.NOM book.FSG.NOM  
 ‘This is my book (FSG).’
- (9) a. To je [NP-use]  
 this.one.here.FSG.NOM be.PRS.3SG  
**tvoja**  
 POSS.2P.SG.NOM+POSSEE.F.SG.NOM  
 ‘This one here is yours.’
- b. To je [Mod-use]  
 the.one.here.FSG.NOM be.PRS.3SG  
**tvoja** **knjiga**  
 POSS.2P.SG.NOM+POSSEE.F.SG.NOM book.FSG.NOM  
 ‘This is your book.’
- c. To je [NP-use]  
 the.one.here.FSG.NOM be.PRS.3SG  
**njegova**  
 POSS.3MSG.NOM+POSSEE.FSG.NOM  
 ‘This is his.’

- d. To je [Mod-use]  
the.one.here.FSG.NOM be.PRS.3SG  
**njegova knjiga**  
POSS.3MSG.NOM+POSSEE.FSG.NOM book.FSG.NOM  
‘This is his book.’
- e. To je [NP-use]  
the.one.here.FSG.NOM be.PRS.3SG  
**njena**  
POSS.3FSG.NOM+POSSEE.FSG.NOM  
‘This is hers.’
- f. To je [Mod-use]  
the.one.here.FSG.NOM be.PRS.3SG  
**njena knjiga**  
POSS.3FSG.NOM+POSSEE.FSG.NOM book.FSG.NOM  
‘This is her book.’

#### [4] Adjectives

In standard Slovene, all adjectives decline according to the number, gender, and case of the entity they evoke. Nominative forms of a declinable adjective *lep* ‘beautiful’ are shown below.

Table 5. Adjectives

| Nominative | Singular  | Dual      | Plural    |
|------------|-----------|-----------|-----------|
| Masculine  | lep avto  | lepa avta | lepi avti |
| Feminine   | lepa hiša | lepi hiši | lepe hiše |
| Neuter     | lepo okno | lepi okni | lepa okna |

Note: *avto* (car, masculine), *hiša* (house, feminine), and *okno* (window, neuter) are in nominative case.

However, some loanwords (like ‘cool’ or ‘super’), slang and fixed phrases (po *domače* – in a homely/local way, v *živo* – live broadcast) do not decline.

#### [5] Verb-based nominalizations

##### [5a] Relative clauses

According to Greenberg (2006: 150-153) in Slovene there are three ways of forming relative clauses: (i) using indeclinable relativizing particle *ki* alone, (ii) using (a) relativizing particle *ki* plus clitic pronoun or (b) using preposition in ACC, GEN, DAT plus *kater-* in the case where preposition is required, (iii) using preposition in LOC and INSTR plus *kater-*. Indeclinable relativizing particles *ki* alone is used when the referent is in nominative case. The combination of relativizing particles *ki* plus clitic pronoun is used when the referent is in a case other than the NOM, but without a preposition. The clitic pronoun agrees in number and gender with its

referent. If the role of a referent requires a preposition in ACC, GEN, DAT, the combination of preposition in ACC, GEN, DAT plus *kater-* is used. Finally, if the role of a referent requires a preposition in LOC, INSTR, the combination of preposition in LOC, INSTR plus *kater-* is used. The table below is reproduced from Greenberg (2006: 151).

| CASE  | M                                     | N                                     | F                                   | PL                                    | DL                                      |
|-------|---------------------------------------|---------------------------------------|-------------------------------------|---------------------------------------|-----------------------------------------|
| NOM   | ki                                    | ki                                    | ki                                  | ki                                    | ki                                      |
| ACC   | ki...ga,<br>PREP<br>+ <i>katerega</i> | ki...ga,<br>PREP<br>+ <i>katerega</i> | ki...jo,<br>PREP<br>+ <i>katero</i> | ki...jih,<br>PREP<br>+ <i>katerih</i> | ki...ju,<br>PREP<br>+ <i>katerih</i>    |
| GEN   | ki...ga,<br>PREP<br>+ <i>katerega</i> | ki...ga,<br>PREP<br>+ <i>katerega</i> | ki...je,<br>PREP<br>+ <i>katere</i> | ki...jih,<br>PREP<br>+ <i>katerih</i> | ki...ju,<br>PREP<br>+ <i>katerih</i>    |
| DAT   | ki...mu,<br>PREP<br>+ <i>kateremu</i> | ki...mu,<br>PREP<br>+ <i>kateremu</i> | ki...ji,<br>PREP<br>+ <i>kateri</i> | ki...jim,<br>PREP<br>+ <i>katerim</i> | ki...jima,<br>PREP +<br><i>katerima</i> |
| LOC   | PREP +<br><i>katerem</i>              | PREP +<br><i>katerem</i>              | PREP +<br><i>kateri</i>             | PREP +<br><i>katerih</i>              | PREP +<br><i>katerih</i>                |
| INSTR | PREP +<br><i>katerim</i>              | PREP +<br><i>katerim</i>              | PREP +<br><i>katero</i>             | PREP +<br><i>katerimi</i>             | PREP +<br><i>katerima</i>               |

The above table gives canonical forms, but there are cases that deviate from it. Concrete examples illustrating these are given in Section 3.

### [5b] Complement clauses

- (10) verjamem, [da je fant pošten]  
believe.1SG that be.3SG boy honest.MSG  
'I believe [that the boy is honest].'
- (11) slišal sem [govorico, [da je Andrej kupil hišo]]  
hear.1MSG.PST rumour.ACC that be.3SG Andrej.buy.PST house.ACC  
'I heard a [rumour [that Andrej bought a house]].'
- (12) [mučenje živali] je nesprejemljivo  
torture.GER animal.PL.GEN be.3SG unacceptable  
'[Torturing of animals] is unacceptable.'

### 3. DATA

#### 3.1 Gender markers

##### [1] Numerals

As mentioned earlier, the numerals ‘one’ to ‘four’ decline for gender while the rest do not. The gendered forms of the numerals ‘one’ to ‘four’ in the nominative case are given in the table below. Boldface indicates distinct forms while plain text indicates case of gender neutralization.

| NUM | 1         |             |             | 2           |      |      | 3              |             |             | 4              |               |               |
|-----|-----------|-------------|-------------|-------------|------|------|----------------|-------------|-------------|----------------|---------------|---------------|
| GEN | M         | F           | N           | M           | F    | N    | M              | F           | N           | M              | F             | N             |
| NOM | <b>en</b> | <b>en.a</b> | <b>en.o</b> | <b>dv.a</b> | dv.e | dv.e | <b>tr.i.je</b> | <i>tr.i</i> | <i>tr.i</i> | <b>štir.je</b> | <i>štir.i</i> | <i>štir.i</i> |

##### [1a] Cardinal numerals

**Numeral ‘one’:** It has a distinct form for each gender (M/F/N). Gender distinction is maintained on the numeral ‘one’ (like in Konkani, and Southern Marathi, see the paper on Konkani and Southern Marathi, in this volume). Compare example (16) [NP-use] with (17) [Mod-use].

- (13) koliko {fantov/deklic/telet}  
how.many.NOM.NSG {boy.PL.GEN/girl.PL.GEN/calf.PL.GEN}  
je.prišlo včera? [NP-use]  
be.SG.come.PSTPTCP.NSG yesterday  
‘How many {boys/girls/calves} came yesterday?’
- (14) samo {e(de)n/ena/eno} [NP-use]  
only {one.MSG.NOM/one.FSG.NOM/one.NSG.NOM}  
je.priš {el/la/lo} včera  
be.SG.come.PSTPTCP {MSG/FSG/MSG} yesterday  
‘Only one {MSG/FSG/MSG} came yesterday.’
- (15) samo {en fant /ena deklica/ [Mod-use]  
only one.MSG.NOM boy/ one.FSG.NOM girl/  
eno tele} je.priš {el/la/lo} včera  
one.NSG.NOM calf.NSG.NOM come.PSTPTCP {M/F/N}.SG yesterday  
‘Only one {boy/girl/calf} came yesterday.’

**The numerals ‘two’, ‘three’ and ‘four’** have two forms only: masculine vs. feminine/neuter (i.e. gender neutralization occurs).

- (16) samo {**dva/dve/dve**} [NP-use]  
 only {two.MDU.NOM/two.FDU.NOM/two.NDU.NOM}  
 sta.priš {la/li/li} včera  
 be.DU.come.PSTPTCP {MDU/FDU/MDU} yesterday  
 ‘Only two {MDU/FDU/NDU} came yesterday.’
- (17) samo **dva** [fanta / moška] [Mod-use]  
 only two.DU.M.NOM boy.DU.NOM/man.DU.NOM  
**dve** [deklici / ženski/]  
 two.F.DU.NOM girl.DU,NOM/woman.DU.NOM  
**dve** [teleti]  
 two.N.DU.NOM calf.DU.NOM  
 sta.priš {la/li/li} včera  
 be.DU.come.PSTPTCP {MDU/FDU/MDU} yesterday  
 ‘Only two {[boys/men]/[girls/women]/calves} came yesterday.’
- (18) samo {**triје/tri/tri**} [NP-use]  
 only {three.MPL.NOM/three.FPL.NOM/three.NPL.NOM}  
 so.priš {li/le/la} včera  
 be.PL.come.PSTPTCP {MPL/FPL/MPL} yesterday  
 ‘Only three {MPL/FPL/MPL} came yesterday.’
- (19) samo **triје** [fantje / moški] [Mod-use]  
 only three.MPL.NOM. boy.PL.NOM/man.PL.NOM  
**tri** [deklice / ženske]  
 three.FPL.NOM girl.PL.NOM/ woman.PL.NOM  
**tri** [teleta]  
 three.NPL.NOM calf.NPL.NOM  
 so.priš {li/le/la} včera  
 be.PL.come.PSTPTCP {MPL/FPL/MPL} yesterday  
 ‘Only three {[boys/men]/[girls/women]/calves} came yesterday.’
- (20) samo {**štirje/štiri/štiri**} [NP-use]  
 only {four.MPL.NOM/four.FPL.NOM/four.NPL.NOM}  
 so.priš {li/le/la} včera  
 be.PL.come.PSTPTCP {MPL/FPL/MPL} yesterday  
 ‘Only four {MPL/FPL/MPL} came yesterday.’



- |      |                                                                 |               |              |   |              |           |
|------|-----------------------------------------------------------------|---------------|--------------|---|--------------|-----------|
| (21) | samo                                                            | <b>štirje</b> | [fantje      | / | moški]       | [Mod-use] |
|      | only                                                            | four.MPL.NOM. | boy.PL.NOM   | / | man.PL.NOM   |           |
|      |                                                                 | <b>štiri</b>  | [deklice     | / | ženske/]     |           |
|      |                                                                 | four.FPL.NOM  | girl.PL.NOM  | / | woman.PL.NOM |           |
|      |                                                                 | <b>štiri</b>  | teleta       |   |              |           |
|      |                                                                 | four.NPL.NOM  | calf.NPL.NOM |   |              |           |
|      | so.priš {li/le/la}                                              |               |              |   | včeraj       |           |
|      | be.PL.come.PSTPTCP {MPL/FPL/MPL}                                |               |              |   | yesterday    |           |
|      | 'Only four { [boys/men]/[girls/women]/calves } came yesterday.' |               |              |   |              |           |

For the numerals ‘two’ through ‘four’, as shown in (18) through (23) above, feminine and neuter forms are the same (gender neutralization is observed). As for **numerals ‘five’ and above**, gender is not marked on the numerals and the construction itself changes, see (24) and (25), below.

- (22) samo {pet/šest/sedem/osem/devet/deset} jih [NP-use]  
 only {five/six/seven/eight/nine/ten} 3PL.GEN  
 je.prišlo včera  
 be.SG.come.PSTPTCP.NSG yesterday  
 ‘Only {five/six/seven/eight/nine/ten} of them came yesterday.’
- (23) samo {pet/šest/sedem/osem/devet/deset} [Mod-use]  
 only {five/six/seven/eight/nine/ten}  
 [fantov/deklic/telet]  
 boy.PL.NOM/girl.PL.NOM/calf.NPL.NOM  
 je.prišlo včera  
 be.SG.come.PSTPTCP.NSG} yesterday  
 ‘Only {five/six/seven/eight/nine/ten} {boys/girls/calves} came yesterday.’

### [1b] Ordinal numerals

Ordinal numerals decline for gender as shown in the following examples.

- |      |                                                                                                          |           |
|------|----------------------------------------------------------------------------------------------------------|-----------|
| (24) | Pokličí <b>tretjo</b><br>call.2SG.IMP   third.FSG.ACC<br>‘Call the third one (FSG).’                     | [NP-use]  |
| (25) | Pokličí <b>tretjo</b> deklico<br>call.2SG.IMP   third.FSG.ACC   girl (FSG).ACC<br>‘Call the third girl.’ | [Mod-use] |

## [2] Demonstratives

In the case of demonstrative pronouns, gender distinction is marked overtly on them, both in NP-use as well as in the Modification use, as shown in the example below.

- (26) kater- $\{i/a/o\}$  {fant/deklica/dete}  
 which- $\{MSG/FSG/NSG\}$ .NOM boy/girl/baby  
 je.bist- $\{er/ra/ro\}$ ?  
 be.SG.smart.NOM- $\{MSG/FSG/MSG\}$   
 ‘Which {boy/girl/baby} is smart?’
- (27) **tist- $\{i/a/o\}$**  [NP-use]  
 that- $\{MSG/FSG/NSG\}$   
 ‘That one  $\{MSG/FSG/NSG\}$ .’
- (28) {**tist-i** fant/ **tist-a** deklica/ [Mod-use]  
 that-MSG boy.MSG.NOM/ that-FSG girl.FSG.NOM/  
**tist-o** dete} je.bist- $\{er/ra/ro\}$   
 that-NSG baby.NSG.NOM be.SG.smart.NOM- $\{MSG/FSG/MSG\}$   
 ‘That  $\{MSG/FSG/NSG\}$  {boy/girl/baby} is smart.’

### [3] Genitive (possessive)

In Slovene, the notion of possession in the case of personal pronouns is expressed by a form called possessive pronoun and it is different from the genitive case form. For example, in the case of first personal singular pronoun, the possessive pronoun is *moj*, which expresses the meaning of possession (*moja knjiga* ‘my book’) and the genitive case form is *mene* (*mene ni doma* ‘I am not at home’) which does not express the meaning of possession. In case of nouns, there is similar distinction: what could be considered as possessive form (**bor.ovo** [pine-POSS.ADJ.NSG.NOM] *lubje* [NSG.NOM] (pine’s bark) and genitive form (*cena* [FSG.NOM] **bor.a** [MSG.GEN] ‘the price of a pine tree’). In the case of possessive form there is gender agreement between the possessee and the possessor, while in the case of genitive form there is no gender agreement between the possessee and the possessor. In this report we will discuss only the possessive forms of personal pronouns.

With regard to possessive forms of the personal pronouns, gender distinction is overtly marked, both in NP-use as well as in Modification use, as shown in the example below.

- (29) čigav- $\{\emptyset/a/o\}$  {sin/hči/dete}  
 who.POSS- $\{M/F/N\}$ .NOM {son/daughter/baby}  
 je.bolj.bist- $\{er/ra/ro\}$ ?  
 be.SG.more.smart.NOM- $\{MSG/FSG/MSG\}$   
 ‘Whose {son/daughter/baby} is smarter?’

- (30) a. **moj/tvoj-{\emptyset/a/e}/** [NP-use]  
 {1/2} PSG.POSS {M/F}-MSG/FSG/NSG}.NOM  
 Mine (MSG/FSG/NSG)/Your (MSG/FSG/NSG)  
 je.bolj.bist-{\er/ra/ro}  
 be.SG.more.smart.NOM-{\MSG/FSG/MSG}  
 ‘{Mine (MSG/FSG/NSG)/Yours (MSG/FSG/NSG)} is smarter.’
- b. **njegov/njen-{\emptyset/a/o}** [NP-use]  
 3SG.POSS.M/F-{\MSG/FSG/NSG}.NOM  
 je.bolj.bist-{\er/ra/ro}  
 be.SG.more.smart.NOM-{\MSG/FSG/MSG}  
 ‘{His (MSG/FSG/NSG)/Her (MSG/FSG/NSG)} is smarter.’
- c. **Andrejev/Andrejin-{\emptyset/a/o}** [NP-use]  
 Andrej/Andreja-{\MSG/FSG/NSG}.NOM  
 je.bolj.bist-{\er/ra/ro}  
 be.SG.more.smart.NOM-{\MSG/FSG/MSG}  
 ‘Andrej’s/Andreja’s is smarter.’
- d. **moj/tvoj-{\emptyset/a/e}/** [Mod-use]  
 {1/2} PSG.POSS {M/F}-MSG/FSG/NSG}.NOM  
**{sin/hći/dete}** je.bolj.bist-{\er/ra/ro}  
 {son/daughter/baby} be.SG.more.smart.NOM-{\MSG/FSG/MSG}  
 ‘{My (MSG/FSG/NSG)/Your (MSG/FSG/NSG)} {son/daughter/baby} is smarter.’
- e. **njegov/njen-{\emptyset/a/o}** [Mod-use]  
 3SG.POSS.M/F-{\MSG/FSG/NSG}.NOM  
**{sin/hći/dete}** je.bolj.bist-{\er/ra/ro}  
 {son/daughter/baby} be.SG.more.smart.NOM-{\MSG/FSG/MSG}  
 ‘{His (MSG/FSG/NSG)/Her (MSG/FSG/NSG)} {son/daughter/baby} is smarter.’
- f. **Andrejev/Andrejin-{\emptyset/a/o}** [Mod-use]  
 Andrej/Andreja-{\MSG/FSG/NSG}.NOM  
**{sin/hći/dete}** je.bolj.bist-{\er/ra/ro}  
 {son/daughter/baby} be.SG.more.smart.NOM-{\MSG/FSG/MSG}  
 ‘Andrej’s/Andreja’s {son/daughter/baby} is smarter.’
- (31) a. **čigav-{\emptyset/a/o}** {avto/knjiga/kolo} je  
 who.POSS-{\M/F/N}.NOM {car/book/bicycle} be.SG  
 to?  
 DEM.PROX.NSG  
 ‘Whose {car/book/bicycle} is this?’

- b. **moj/tvoj-{\O/a/e}** [NP-use]  
 {1/2} PSG.POSS {M/F}-MSG/FSG/NSG}.NOM  
 ‘Mine (MSG/FSG/NSG)/Yours (MSG/FSG/NSG).’
- c. **njegov/njen-{\O/a/o}** [NP-use]  
 3SG.POSS.M/F- {MSG/FSG/NSG}.NOM  
 ‘His (MSG/FSG/NSG)/Her (MSG/FSG/NSG).’
- d. **Andrejev/Andrejin-{\O/a/o}** [NP-use]  
 Andrej/Andreja- {MSG/FSG/NSG}.NOM  
 ‘Andrej’s/Andreja’s.’
- e. **moj/tvoj-{\O/a/e}/** [Mod-use]  
 {1/2} PSG.POSS {M/F}-MSG/FSG/NSG}.NOM  
 {avto/knjiga/kolo} je.drag-{\O/a/o}  
 {car/book/bicycle} be.SG.expensive- {MSG/FSG/NSG}  
 ‘{Mine (MSG/FSG/NSG)/Yours (MSG/FSG/NSG)}  
 {car/book/bicycle} is expensive.’
- f. **njegov/njen-{\O/a/o}** [Mod-use]  
 3SG.POSS.M/F- {MSG/FSG/NSG}.NOM  
 {avto/knjiga/kolo} je.drag-{\O/a/o}  
 {car/book/bicycle} be.SG.expensive- {MSG/FSG/NSG}  
 ‘{His (MSG/FSG/NSG)/Her (MSG/FSG/NSG)} {car/book/bicycle} is expensive.’
- g. **Andrejev/Andrejin-{\O/a/o}** [Mod-use]  
 Andrej/Andreja- {MSG/FSG/NSG}.NOM  
 {avto/knjiga/kolo} je.drag-{\O/a/o}  
 {car/book/bicycle} be.SG.expensive- {MSG/FSG/NSG}  
 ‘Andrej’s/Andreja’s {car/book/bicycle} is expensive.’
- (32) a. čigav-i so.ti ključī?  
 who.POSS- {MPL}.NOM be.PL.DEM.PROX.MPL key.MPL.NOM  
 ‘Whose keys are these?’
- b. (od) [svetega Petra] [NP-use]  
 (of) saint.ADJ.MSG.GEN Peter.MSG.GEN  
 ‘(Of) St. Peter’s.’
- c. (od) [svete Klare] [NP-use]  
 (of) saint.ADJ.FSG.GEN Clara.FSG.GEN  
 ‘(Of) St. Clara’s.’



Concrete examples are given below. Example (34) and (35) are the cases which deviate from the canonical forms listed in table from Greenberg (2006: 151) which is given in Section 2.

Using indeclinable relativizing particle *ki* alone:

- (34) [{fant/deklica/dete} [ki sedi na stolu]]  
 boy/girl/baby who is.sitting LOC chair  
 je moj otrok  
 be.PRS.SG 1.POSS child  
 ‘[The {boy/girl/baby} [who is sitting on the chair]] is my child.’

In the above sentence, the relativizer ‘*ki*’ does not carry gender of the denotation of the head noun.

Using prepositions in ACC, GEN, or DAT plus *kater-*

- (35) To je knjiga, [za katero sem slišal  
 This be.PRS.3SG book.FSG.NOM for which.ACC 1SG hear.PST  
 Včeraj]  
 yesterday  
 ‘This is the book that I heard about yesterday.’
- (36) To je knjiga, [ki sem za njo  
 this be.PRS.3SG book.FSG.NOM that be.1SG for 3FSG.ACC  
 slišal včeraj]  
 hear.PTCPL yesterday  
 ‘This is the book that I heard about yesterday.’
- (37) to je odbor, [katerega naloge so  
 this be.PRS.3SG committee.MSG who.GEN.MSG task.PL be.PRS.3PL  
 znane]  
 known  
 ‘This is the committee of which tasks are known.’

Using prepositions in LOC or INSTR plus *kater-*

- (38) Mesto, [v katerem se je  
 town.NSG in which.NSG.LOC REFL be.PRS.3SG  
 rodil Janez] je blizu Ljubljane  
 born.PTCPL.MSG Janez be.PRS.3SG near Ljubljana.GEN  
 ‘The town where Janez was born is near Ljubljana.’

- (39) Volilna\_enota, [v **kateri** kandidira Janez] je  
 election ward.FSG in which.FSG.LOC is.running Janez be.PRS.3SG  
 velika].....  
 big.FSG  
 ‘The election ward where Janez is running is big/populous.’
- (40) to je roman, [o **katerem**  
 this.MSG be.PRS.3SG novel.MSG about which.MSG.LOC  
 govorimo]  
 talk.1PL.PRS  
 ‘This is the novel about which we are talking’.
- (41) to je nož, [s **katerim** je  
 this be.PRS.3SG knife.MSG with which.MSG.INSTR be.PRS.3SG  
 prerezal jabolko]  
 cut.PST.3SG apple  
 ‘This is the knife with which he cut the apple.’

The declinable *kater-* type is preferred in non-core/peripheral relationships such as possession, location, time, instrument, etc.

When the referent is unclear (blurred), it is treated as neuter and a gender invariant relative pronoun *kar* is used.

- (42) Mnogo je takega, [**kar** moramo  
 many be.PRS.3SG of.such.kind.NSG which.ACC.SG 1PL.must  
 odklanjati]  
 reject.INF  
 ‘There are many things that we must reject.’

Regarding the origin of indeclinable *ki*, there are various schools of thought and one of them posits *ki* as the result of contraction of declinable *kateri* (*ki* < *kateri*, see Sonnenhauser 2023, Section 3, pp.155-166). If the diachronic change from *ki* < *kateri* is correct, it would mean the loss of gender marking of the denotation on the relative pronoun.

### [5b] Complement clauses

Complement clauses will not be discussed in the report. Examples of complement clauses mentioned earlier are copied below for the ease of reference.

- (43) verjamem, [da je fant pošten]  
 believe.1SG that be.3SG boy honest.MSG  
 ‘I believe [that the boy is honest].’ (repeated from (10))

- (44) slišal sem [govorico, [da je Andrej kupil hišo]]  
 hear.1MSG.PST rumour.ACC that Andrej.buy.PST house.ACC  
 ‘I heard a [rumour [that Andrej bought a house]].’ (repeated from (11))
- (45) [mučenje živali] je nesprejemljivo  
 torture.GER animal.PL.GEN be.3SG unacceptable  
 ‘[Torturing of animals] is unacceptable.’ (repeated from (12))

#### 4. ANALYSIS AND DISCUSSION

The distribution of gender marking in Slovene discussed in the foregoing section can be summarized in the form of Shibatani’s (2023) hierarchy as follows:

##### Gender marking in Slovene

Figure 1. Gender marking pattern in Slovene

NP use: LIM/NEUT NUM > OBLDEM > OBLGEN > OBLADJ > OBL/NEUTV-based NMLZ

Mod use: LIM/NEUT NUM > OBLDEM > OBLGEN > OBLADJ > OBL/NEUTV-based NMLZ

(LIM/NEUT = limited to a few numerals ‘four’ and below, neutralized for numerals ‘five’ and above, OBL = obligatory; OBL/NEUT = obligatory in relative (declinable) adjective *kateri*-type and participle-type, neutralized in indeclinable operator *ki*-type).

Note: The slot GEN on the above hierarchy is filled by the possessive forms of pronouns and nouns.

Shibatani’s hierarchy (above) can account for the distribution of the gender marking in Slovene, with the assumption that Shibatani’s GEN category comprises of possessive forms and genitive forms expressing possession in Slovene. All the five structural categories (referred to as adnominal in linguistic literature) on this hierarchy are marked for gender, a property that lends support to our claim that these structures are nominalizations. These structures perform referential function, which is a hallmark of their nominal nature.

A note on the gender marking on numerals is in order. The distribution of gender on numerals is restricted: only lower numerals (in case of Slovene, 1 to 4) overtly show gender distinction both in NP-use and Mod-use, like Sanskrit. For higher numerals (5 and above) gender distinction is lost. In Romanian and Bulgarian, gender distinction is limited to the numerals ‘one’ and ‘two’ (Romanian: 1: *un* [M] vs. *o* [F], 2: *doi* [M] vs. *două* (F); Bulgarian: 1: *edin* (M), *edna* (F), *edno* (N), 2: *dva* (M) vs. *dve* (F/N)). In Spanish, Konkani, Southern Marathi, etc. gender marking is retained only on the numeral ‘one’ (see papers on Konkani and Southern Marathi in this volume). In some IE languages like English, Hindi-Urdu, Assamese, Persian, etc., numerals do not show gender distinction at all. These facts point to the diachronic process of loss of gender in Indo-European languages.



As mentioned before, regarding the origin of indeclinable *ki*, there are various schools of thought and one of them posits *ki* as the result of contraction of declinable *kateri* (*ki* < *kateri*, see Sonnenhauser 2023, Section 3, pp.155-166). In view of this, a case for loss of gender on the relative pronoun (*kateri* > *ki*) can be made for Slovene. More detailed work on this change, is necessary, however, not only for Slovene but also for other Slavic languages where a similar change is attested (cf. Sonnenhauser 2023, Section 2.2, pp.154–155). Slovene data points to the possibility that the starting point of the loss of gender can be both ends of Shibatani's hierarchy: (NUM) from where it would spread rightward and V-based nominalization from where it would spread leftward.

Slovene is a language where gender marking is attested on all categories on Shibatani's hierarchy. However, in Assamese, Bengali (Indo-Aryan, see Bez 2022), Persian, Balochi (Indo-Iranian) have lost gender on all the categories.

Figure 2. Modern Standard Assamese (Northeast Assam) (Bez 2022; See Shibatani (2023), Appendix A for the supporting data.)

NP-use: NUM > DEM > \*OPT GEN > \*OPT ADJ > \*OPT V-based nominalization

Mod-use: NUM > OPT DEM > OPT GEN > \*ADJ > \*V-based nominalization

(\* = no marking; OPT = optional marking; \*OPT = marking applies for a specific/definite referent; unmarked=obligatory marking)

In between these two extremes, lie languages like, Marathi, Konkani (Indo-Aryan), where the loss of gender is limited to numerals, Hindi-Urdu (Indo-Aryan) where it has spread from NUM to DEM. Data from more gendered languages, such as Gujarati, Sindhi, Romanian, Bulgarian, etc., is needed to confirm whether Shibatani's hierarchy could be used as a diagnostic tool to predict and account for the loss of gender marking in the languages of the globe.

## 5. CONCLUSION

In this paper, we offered a preliminary report on the distribution of gender marking on: (a) numerals, (b) demonstratives, (c) genitives (with the assumption that GEN category comprises of possessive forms and genitive forms expressing possession in Slovene), (d) adjectives and (e) the so-called relative clauses in Slovene in their two usages: (i) NP-use, where these structures are used as the heads of noun phrases, and (ii) modification-use, where these structures are used as modifiers of a noun, adopting the nominalization framework proposed in Shibatani (2017, 2018, 2019, 2021, 2023). Following Shibatani, we analyzed gender marking as a nominalization device, which not only nominalizes the structure to which it is attached, but also classifies it into one of three groups, viz. masculine, feminine and neuter. We demonstrated that the distribution of gender marking in Slovene supports Shibatani's hierarchy. Further work is needed to assess whether Shibatani's hierarchy can predict and account for the attrition of gender marking in the languages of the globe.

## ABBREVIATIONS

|         |                 |
|---------|-----------------|
| 1       | first person    |
| 2       | second person   |
| 3       | third person    |
| ACC     | accusative      |
| ADJ     | adjective       |
| DAT     | dative          |
| DEM     | demonstrative   |
| DIST    | distal          |
| DU      | dual            |
| F       | feminine        |
| FUT     | future          |
| GEN     | genitive        |
| GER     | gerund          |
| INF     | infinitive      |
| INSTR   | instrumental    |
| LOC     | locative        |
| M       | masculine       |
| N       | neuter          |
| NOM     | nominative      |
| NUM     | numeral         |
| PL      | plural          |
| POSS    | possessor       |
| POSSEE  | possessee       |
| PROX    | proximate       |
| PRS     | present         |
| PST     | past            |
| PTCP    | participle      |
| PSTPTCP | past participle |
| REFL    | reflexive       |
| SG      | singular        |

## ACKNOWLEDGMENTS

An earlier version of this paper was presented at the Linguistic Circle Meeting of the University of Ljubljana on 25 February 2025. We would like to thank the participants for their comments and suggestions. The present study has been supported by grants from: (i) National Institute for Japanese Language and Linguistics (‘Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective’, PI: Prashant Pardeshi) and (ii) Japan Society for the Promotion of Science (KAKENHI Project IDs: 22H00659 and 23K21931, ‘Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers’, PI: Kazuhiro Kawachi).

## REFERENCES

- Bez, Gitanjali. 2022. Assamese and Bengali classifiers: A comparative study in nominalization perspective. Talk given at the Central Institute of Indian Languages, Mysore, India, on August 6, 2022.
- Pardeshi, Prashant, and Masayoshi Shibatani. 2023. Marathi relative and complement clauses in nominalization perspective. In Rajesh Kumar and Om Prakash. (eds.) *Language studies in India: Cognition, structure, variation*, 321–352. Singapore: Springer Nature. [https://doi.org/10.1007/978-981-19-5276-0\\_19](https://doi.org/10.1007/978-981-19-5276-0_19)
- Shibatani, Masayoshi. 2017. Nominalization. In Masayoshi Shibatani, Shigeru Miyagawa, and Hisashi Noda. (eds.) *Handbook of Japanese syntax*, 271–332. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9781614516613-009>
- Shibatani, Masayoshi. 2018. Nominalization in Crosslinguistic Perspective. In Prashant Pardeshi and Taro Kageyama. (eds.) *Handbook of Japanese contrastive linguistics*, 345–410. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9781614514077-013>
- Shibatani, Masayoshi. 2019. What is nominalization? Towards the theoretical foundations of nominalization. In Roberto Zariquiey, Masayoshi Shibatani, and David W. Fleck. (eds.) *Nominalization in languages of the Americas*, 15–168. Amsterdam & Philadelphia: John Benjamins. <https://doi.org/10.1075/tsl.124.02shi>
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52. [https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)
- Sonnenhauser, Barbara. 2023. Relative clauses in Slovene: Diachronic puzzles, synchronic patterns. In *Wiener Slavistisches Jahrbuch*, Vol. 1, 150–187. Published by: Harrassowitz Verlag; Austrian Academy of Sciences Press. Stable. <https://www.jstor.org/stable/10.13173/wienschlavjahr.1.2013.0150>
- Uppoor, Niranjana. 2022. Gender marking in Kannada and other Dravidian languages: A nominalization perspective. Talk given at the University of Tokyo on September 9, 2022.

# Gender marking and nominalization in Bari (Eastern Nilotic)

Shuichiro Nakao<sup>1</sup>

## ABSTRACT

This report describes the nominal morphosyntax of Bari, an Eastern Nilotic language spoken in South Sudan, with reference to Shibatani's (2021, 2023) hierarchy for classifier/gender marking. It shows that although Bari marks gender on demonstratives, genitive noun phrases, most types of adjectives, and verb-based nominalizations, it lacks gender marking on numerals (as well as on a few underived adjectives). This contradicts Shibatani's implicational hierarchy along the structural dimension, which regards numerals as the structure most likely to receive classifier/gender marking (Shibatani 2023: 34). On the other hand, the Bari data are consistent with Shibatani's hierarchy along the functional dimension, which emphasizes the primacy of the noun-phrase use of grammatical nominalizations over their modification use.

## 1. INTRODUCTION

Bari (ISO 639: bfa) is an Eastern Nilotic language spoken in the Republic of South Sudan with an estimated 977,100 speakers worldwide (Eberhard et al. 2025). This report presents data on the nominal morphosyntax of Bari, with a focus on the realization of gender marking in the following types of grammatical nominalizations.

- [1] Numerals ([1a] Cardinal numerals, [1b] Ordinal numerals)
- [2] Demonstratives (demonstrative pronouns/adnominal demonstratives)
- [3] Genitive noun phrases
- [4] Adjectives (with five sub-classes)
- [5] Verb-based nominalizations ([5a] Relative clauses, [5b] Complement clauses)

### Basic information on Bari

Bari distinguishes two grammatical genders—feminine and masculine (Spagnolo 1933, Nyombe 2004)—where the feminine seems to be the default. For example, an unbiased question like (1) chooses the feminine demonstrative.<sup>2</sup>

- (1)     $\eta\acute{n}\grave{a}$              $\grave{a}$              $n\acute{o}$ ?  
         MED.F           COP   what  
         ‘What is this/that/it [near you]?’ (cf.  $\eta\acute{l}\acute{o}$  ‘MED.M’)<sup>3</sup>

---

<sup>1</sup> Graduate School of Humanities, the University of Osaka, 1-1 Shinsenba, Minoh, Osaka 562-0035, Japan.  
Email: s.nakao.hmt@osaka-u.ac.jp

<sup>2</sup> This has nothing to do with the abstract noun  $\eta\acute{o}$  ‘thing’, which is masculine.

<sup>3</sup> Hereafter, Bari is transcribed based on IPA.

The nouns themselves (except a few personal names) lack overt gender marking.<sup>4</sup> Animate nouns are gender-neutral unless specified in certain grammatical contexts (e.g., when modified by a demonstrative, as in (2), where the proximal demonstratives are represented by F *ná* and M *lɔ́*). Inanimate nouns have inherent gender (with some degree of predictability, see Spagnolo 1933: 18–31).<sup>5</sup>

- (2) a.    *ɲótú*            ‘person’  
          *ná ɲótù*       ‘this.F person (woman)’  
          *lɔ́ ɲótù*       ‘this.M person (man)’
- b.    *súkùrì*       ‘chicken’  
          *ná súkùrì*      ‘this.F chicken (hen)’  
          *lɔ́ súkùrì*      ‘this.M chicken (cock)’
- (3) a.    *múgún*       ‘body’ (F)  
          *ná múgùn*      ‘this.F body’
- b.    *ʝúr*            ‘country’ (M)  
          *lɔ́ ʝúr*          ‘this.M country’

Bari has several noun-deriving affixes (instrument nouns *-êṭ/-êṭ* (PL *-és-ì/-és-ì*), verbal nouns *-a/-ʌ/-o*,<sup>6</sup> abstract noun *tú-/tɔ́-*, object noun *kV-*), all of which derive feminine nouns.

Table 1. Affixal nominalization

| Affix           | Root                                | Derived noun                              |
|-----------------|-------------------------------------|-------------------------------------------|
| <i>-êṭ/-êṭ</i>  | <i>dék-</i> ‘talk’ (verb)           | <i>dék-êṭ</i> ‘conversation’ (F)          |
| <i>-êṭ/-êṭ</i>  | <i>mòk-</i> ‘hold, grab’ (verb)     | <i>mòk-êṭ</i> ‘hand (of a mug, etc.)’ (F) |
| <i>-a/-ʌ/-o</i> | <i>pùtúk-</i> ‘knead’ (verb)        | <i>pùtúk-ʌ</i> ‘kneading’ (F)             |
| <i>-a/-ʌ/-o</i> | <i>dót-(ô)</i> ‘sleep’ (verb)       | <i>dót-ô</i> ‘sleeping’ (F)               |
| <i>tú-/tɔ́-</i> | <i>kádě</i> ‘different’ (adjective) | <i>tó-kádě</i> ‘difference’ (F)           |
| <i>tú-/tɔ́-</i> | <i>-rôn</i> ‘bad’ (adjective)       | <i>tɔ́-rôn</i> ‘problem’ (F)              |
| <i>tú-/tɔ́-</i> | <i>gélêṇ</i> ‘one’ (numeral)        | <i>tɔ́-gélêṇ</i> ‘unity’ (F)              |
| <i>kV-</i>      | <i>ɲá-</i> ‘eat’ (verb)             | <i>kí-ɲó</i> ‘food’ (F)                   |
| <i>kV-</i>      | <i>túr-(ʌn)</i> ‘blossom’ (verb)    | <i>kʌ-túrʌn</i> ‘flower’ (F)              |

The agent nouns formed by the circumfix *kà-...-nít* (PL *kà-...-k*, see below) are gender-neutral just like other animate nouns. There is also one case of passive-like deverbal

<sup>4</sup> Exceptionally, some personal names do have gender marking, e.g., *nà-pàjá* (Napaya, female name), *lù-pàjá* (Lupaya, male name).

<sup>5</sup> Bari is a two-level tonal language (Yokwe 1987). A word-final high tone affects the tonal shape of the following word, as seen in these examples (e.g., *ɲótú* → *ɲótù*, *ʝúr* → *ʝúr*).

<sup>6</sup> This suffix is homophonous with the default intransitive/passive verb form (see Nakao 2024).

nominalization—*nár-áttí* ‘beloved one’ (gender-neutral, from *nàr* ‘love’)—but this type of derivation is not productive.

Table 2. Agent noun

| Affix      | Root                   | Derived noun            |
|------------|------------------------|-------------------------|
| kà-...-nít | kìt- ‘work’ (verb)     | kà-kít-à-nít ‘worker’   |
| kà-...-nít | mér- ‘be drunk’ (verb) | kà-mér-à-nít ‘drunkard’ |

- (4) kà-tódìn-à-nít ‘teacher’ (from tò-dín- ‘teach’)  
ná ká-tódìn-à-nít ‘this.F teacher’  
ló ká-tódìn-à-nít ‘this.M teacher’

As for grammatical number, Bari distinguishes between singular and plural. Most nouns have a number marking either on the singular form, the plural form or both (with an exception of *ḡḡ* ‘thing(s).M’, which is number-neutral). Some nouns show a tonal alternation or stem alternation. There is no case of alternation of gender according to number as observed in many languages with a gender system (famously in Afroasiatic).

Table 3. Nominal number marking

| SG             | PL           |                    |
|----------------|--------------|--------------------|
| ǰúr            | ǰúr-àṇ       | ‘country’ (M)      |
| dék-ét         | dék-és-ì     | ‘conversation’ (F) |
| pátál-ì        | pátá         | ‘rope’ (F)         |
| mángà-tât      | mángà        | ‘mango’ (M)        |
| kà-tódìn-à-nít | kà-tódìn-à-k | ‘teacher’ (F/M)    |
| ḡútú           | ḡútû         | ‘person’ (F/M)     |
| múgún          | ḡárík        | ‘body’ (F)         |

The feminine/masculine gender system is traceable to Proto-Eastern Nilotic (Heine & Vossen 1983), while most other branches of Nilo-Saharan lack grammatical gender (Kadu and Koman are notable exceptions, but their affiliation to Nilo-Saharan is disputed).

Bari and the languages closely related to it (Mandari, Kakwa, Nyangbara, Pojulu, Kuku, Nyepu) are typologically divergent from the rest of the Eastern Nilotic languages (i.e., the Teso-Turkana and Lotuho-Maa branches; see Vossen 1982). For example, Bari has generalized AVO word order without argument case marking, while the non-Bari branches have VAO word order with a ‘marked nominative’ case system. Some languages of the non-Bari branches have also developed a third gender (e.g., ‘neutral’ in Turkana and ‘place nouns’ in Maa).

## Data sources

This study is primarily based on the author’s first-hand data provided by Dr. David Gore, a native speaker of the Bari dialect of Rejaf-Juba area, collected in Osaka, Japan, over the period 2020–2024. The data has been transcribed according to IPA-based phonemic

representation. Major literature on Bari includes Spagnolo (1933) and Yokwe (1987) (see also Nakao 2024, 2025).

## 2. OVERVIEW OF GENDER MARKING AND GRAMMATICAL NOMINALIZATIONS IN BARI

### 2.1 Gender marking in Bari

Bari gender marking is realized in various word classes, while they are neutralized in some grammatical environments. In summary, it is realized on demonstratives (including the demonstrative predicates/copulas), genitive particles and genitive personal pronouns, some classes of adjectives, relative markers and focus markers, but not on nouns, (independent) personal pronouns, interrogatives, numerals, complementizers, verbs or (non-demonstrative) copulas.

Nouns can also be classified either human or non-human (cf. *ŋâ* ‘who’ vs. *ɲɔ̃* ‘what’), but this feature is not morphologically marked. So far, no “nominal classifier” has been attested in Bari.

### 2.2 Grammatical nominalizations in Bari

This section gives the overview of the morphological realizations of the gender marking.

#### [1] Numerals

While many Eastern Nilotic languages mark gender on lower numerals (e.g., 1 through 4), Bari (almost) completely lacks gender marking on both cardinal and ordinal numerals. The ordinal numerals are formed by prefixing *tò-/tɔ̃-* (with [±ATR] vowel harmony). Exceptionally, however, the ordinal numeral can have a sort of gender marking in the NP usage described in Section 3.

Table 4. Numerals

| Number | Cardinal numeral      | Ordinal numeral      |
|--------|-----------------------|----------------------|
| 1      | gèlén                 | tò-gèlén             |
| 2      | mùrék                 | tò-mùrék             |
| 3      | mùsálâ                | tò-mùsálâ            |
| 4      | ìŋwán                 | tò-ìŋwán             |
| 5      | mùkánàt               | tò-mùkánàt           |
| 6      | bùkjêr                | tò-bùkjêr            |
| 7      | bùrjá                 | tò-bùrjá             |
| 8      | bùdák                 | tò-bùdák             |
| 9      | bùŋân                 | tò-bùŋân             |
| 10     | pwák                  | tò-pwák              |
| 11     | pwák wàt gèlén        | pwák tò-gèlén        |
| 20     | mérjâ mùrék           | mérjâ mùrék          |
| 21     | mérjâ mùrék wàt gèlén | mérjâ mùrék tò-gèlén |
| 100    | mîa                   | n.a.                 |

The numeral-related lexemes ‘the first born’, which literally means ‘of the head’, and *ná/ló-gélèŋ* ‘the sole.F/M’ (cf. ‘one’) has gender marking developed from the genitive particle, but morpho-syntactically they cannot be categorized as numerals.

- (5) a.    (ŋúrò)        ná/ló-kwê  
           child        (GEN.)F/M-head  
           ‘the first born.F/M (child)’
- b.    ŋúrò        ná/ló-gélèŋ  
           child        (GEN.)F/M-sole  
           ‘the sole.F/M child’

## [2] Demonstratives

Bari has three-way distinctions in demonstratives—proximal, medial and distal—which all exhibit gender and number markings.

Table 5. Nominal demonstratives in Bari

|   | Proximal/definite |      | Medial |      | Distal   |      |
|---|-------------------|------|--------|------|----------|------|
|   | SG                | PL   | SG     | PL   | SG       | PL   |
| F | ná                | kónè | ŋínà   | kínè | ŋínù, nú | kúnù |
| M | ló                | kúlò | ŋílò   | kílò | ŋílù, lú | kúlù |

When a demonstrative modifies a noun, it can be either preposed (in which case a polysyllabic demonstrative will have high tones) or postposed without any semantic difference, except for the fact that a postposed proximal demonstrative can also be interpreted as a definite marker.

- (6) a.    ŋíló        ŋútù  
           MED.SG.M person  
           ‘this/that person (near you)’
- b.    ŋútú        ŋìlò  
           person        MED.SG.M  
           ‘this/that person (near you)’
- (7) a.    ló        ŋútù  
           PROX.SG.M person  
           ‘this.M person’
- b.    ŋútú        lò  
           person        PROX/DEF.SG.M  
           ‘this/the.M person’



Bari demonstratives have predicate forms, which function as existential copulas, obligatorily followed by a place adverb or realized with the presentative prefix *má-*. In these constructions, the predicate agrees with the subject in number and gender.

Table 6. Demonstrative predicates in Bari

|   | Proximal/definite |       | Medial |      | Distal   |      |
|---|-------------------|-------|--------|------|----------|------|
|   | SG                | PL    | SG     | PL   | SG       | PL   |
| F | nà                | kùnnè | ɲìnà   | kìnè | ɲìnû, nû | kùnû |
| M | lò                | kùlò  | ɲìlò   | kìlò | ɲìlû, lû | kùlû |

- (8) a. nân lò nî.  
 1SG PRED.PROX.SG.M here  
 ‘I am.M here.’
- b. mà má nû mîdî.  
 my.mother PRED.DIST.SG.F at.home  
 ‘My mother is.F at home.’

- (9) nân má-lò.  
 1SG PSTV-PRED.PROX.SG.M  
 ‘Here I am.M.’

The demonstrative predicates are also grammaticalized as progressive markers (with deictic connotation). Apart from this, verb phrases never have gender marking.<sup>7</sup>

- (10) a. nân ɲìnà tù kànô.  
 1SG PRED.MED.SG.F IPFV:go out  
 ‘(As you see,) I am.F going out.’
- b. ɲé nû kòndjâ ɲô?  
 3SG PRED.DIST.SG.F IPFV:do what  
 ‘What is.F she doing (over there)?’

### [3] Genitive noun phrases

Bari marks genitive noun phrases with genitive particles agreeing with the possessum (SG.M *lɔ́*, SG.F *ná*, PL *tĩ*). The gender distinction is neutralized in the plural.

<sup>7</sup> Unlike the rest of Eastern Nilotic, Bari lacks subject-marking on verbs except for in the imperative, where a plural referent is marked by =*tà*, e.g., *wàrà-ní!* (walk-IMP) ‘Walk(.SG)!’, *wàrà-ní=tà!* (walk-IMP=PL) ‘Walk(.PL)!’

- (11) a. júr                      lò                      bàrì  
land(M)                      GEN.SG.M      Bari  
‘the land of the Bari’
- b. kútúk                      nà                      bàrì  
mouth(F)                      GEN.SG.F      Bari  
‘the Bari language’
- c. júr-àn                      tí                      àfríkà  
land(M)-PL                      GEN.PL              Africa  
‘the lands of Africa’
- d. kútús-èn                      tí                      àfríkà  
mouth(F)-PL                      GEN.PL              Africa  
‘the African languages’

There are some lexicalized usages of the genitive particle aside from the examples given in [1]. Here we can add at least two fossilized cases: *lúlâlêt* ‘male’ (from *lâlêt* ‘husband’; PL *lúnsàk*) and *lóntóròk* ‘gizzard’ (from *tórók* ‘small stone (M)’).

- (12) a. ɲórò                      ná                      náràkwán  
child                      GEN.SG.F      woman  
‘female.F child’<sup>8</sup>
- b. dǐòŋ                      ló                      múdŋ  
dog                      GEN.SG.M      forest  
‘wild.M dog’

The personal pronouns have genitive forms with gender/number marking, in addition to the independent forms (used for subject/object) and dative forms.

Table 7. Personal pronouns

|     | Independent forms    | Genitive forms |         |            |       | Dative forms |
|-----|----------------------|----------------|---------|------------|-------|--------------|
|     |                      | SG.F           | SG.M    | PL.F       | PL.M  |              |
| 1SG | nân                  | níò            | líò     | kwê        |       | kájà         |
| 2SG | dó                   | (ɲ)ínòt        | (ɲ)ílòt | kúnàk      | kúlàk | kónùt, kúnàk |
| 3SG | ɲé, lòpéɲ            | nápìt          | lópìt   | kápìt      |       | kápìt        |
| 1PL | jî                   | níkàɲ          | líkàɲ   | kàɲ, kájàɲ |       | kájàɲ        |
| 2PL | tá                   | násù           | lósù    | kásù       |       | kásù         |
| 3PL | sé, lòpèɲât, lòpèkât | nàsè           | lósè    | kàsè       |       | kàsè         |

<sup>8</sup> The plural form of (12a) is suppletive: *ɲwájík wájâ* (child.PL female.PL) ‘female(.PL) children’. Cf. *dǐóŋ-àn tí múdŋ* (dog-PL GEN.PL forest) ‘wild dogs’.

#### [4] Adjectives

Bari has five major classes of words that can meet a broader definition of adjective. In this report, we roughly categorize all words that can modify a noun without the genitive particle or the relativizer (see [5] below) as adjectives.<sup>9</sup>

The first class of adjectives (ADJ1) consists of two determiners, which have gender and number marking that looks similar to that of demonstratives (otherwise, we can include demonstratives in ADJ1). Both precede the modified noun, but ‘certain’ can also follow it (with tonal modification).

Table 8. ADJ1

| SG.F | SG.M | PL.F  | PL.M  |                           |
|------|------|-------|-------|---------------------------|
| nán  | lón  | kúnèn | kúlòn | ‘which’                   |
| néné | lélé | kúnjé | kúljé | ‘certain, one, some, any’ |

- (13) a. lélé                      lòr  
           certain.SG.M      day(M)  
           ‘one/certain.M day’
- b. lòr                      lélé  
           day(M)              certain.SG.M  
           ‘one/certain.M day’

The second class of adjectives (ADJ2; ‘relative adjectives’, Spagnolo 1933: 61–63) exemplified in Table 9 always have gender marking by prefix and number marking by a stem alternation (which is often tonal). They require the copula à in the predicate use (15).

Table 9. ADJ2

| SG.F   | SG.M   | PL.F   | PL.M   |        |
|--------|--------|--------|--------|--------|
| ná-rôn | lô-rôn | ná-rôk | lô-rôk | ‘bad’  |
| ná-bút | lô-bút | ná-bût | lô-bût | ‘good’ |
| ná-tór | lô-tór | ná-tôr | lô-tôr | ‘red’  |

- (14) a. ɲórò              lô-rôn  
           child            M-bad.SG  
           ‘bad.M child’
- b. kák              ná-tór  
           soil(F)        F-red.SG  
           ‘red.F soil’

<sup>9</sup> The first three (and a part of the fourth) classes of adjectives can be discussed as cases of verb-based nominalization, but we tentatively present them in [3]. Meanwhile, ordinal numerals, presented in [1], can be discussed as a separate class of adjectives.

(15) a. nân à ná-ḃút.  
1SG COP F-well.SG  
'I am fine.F.'

b. jî à ló-ḃút.  
1PL COP M-well.PL  
'We are fine.M.'

A number of these adjectives have corresponding inchoative verbal forms (suffixed with *-an/-an/-on*).

(16) a. lòpéŋ à ló-ḃút?  
3SG COP M-good.SG  
'Is he well.M?' (adjective)

b. lòpéŋ à ḃús-án?  
3SG PRET good-INCH  
'Did he get well.M?' (verb)

(17) a. ŋútú ló-pír  
person M-fat  
'a fat.M person' (adjective)

b. ŋútú ló pír-án  
person REL.M fat-INCH  
'a person who.M became fat' (verb)

The third class of adjectives (ADJ3; 'defective-verbs adjectives', Spagnolo 1933: 61–63) has gender marking in non-predicate (i.e., NP and Modification) uses and has no number marking.<sup>10</sup> It lacks the copula à in the predicate use, which would imply that this category is closer to the verbs.

(18) a. sâi pàpé.<sup>11</sup>  
tea(PL.F) hot  
'The tea is hot.'

<sup>10</sup> Although it is beyond the scope of this article, this lack of gender marking in the predicate usage is comparable to the German adjectives (Shibatani 2021).

<sup>11</sup> These adjectives tend to have a fossilized element *pa-/pa-* (e.g., *pàfə* 'far', *pàràráŋ* 'salty', *pàtár* 'bitter'), which Spagnolo (1933) compares with the genitive particle *pa* in Lwo (Western Nilotic) languages.

- b.    sâi            ná-pápè  
       tea(PL.F)    F-hot  
       ‘(the) hot.F tea’
- c.    ná-pápè  
       F-hot  
       ‘the hot.F one’

The fourth class of adjectives (ADJ4; ‘true adjectives’, Spagnolo 1933: 61–63) lacks gender marking, except in the NP use (see Section 3). In the predicate usage, these adjectives require the copula à. Some of them have obligatory number marking (e.g., SG *dúmà* PL *dúmà-lák* ‘big’, SG *lùdúkà-tjá* PL *lùdúkà* ‘new’, SG *kwáríní-t* PL *kwáríní-kà* ‘rich’), while the others do not (e.g., *mógôr* ‘hungry’, *rígwò* ‘true’).

- (19) a.    ñô            dùmà  
           thing(M)    big  
           ‘the big thing, something big’
- b.    ná            à        dùmà.  
       PROX.SG.F    COP    big  
       ‘This.F (one) is big.’

ADJ4 includes the deverbal ‘purpose’ adjectives suffixed with *-êl/-êt* (PL *-és-i/-és-i*).

- (20) a.    pîrît            dér-êt  
           place(F)    cook-PURP  
           ‘kitchen (place for cooking)’
- b.    títo            dér-és-i  
       instrument(PL.F) cook-PURP-PL  
       ‘utensils (instruments for cooking)’

The fifth class of adjectives (ADJ5) comprises two quantifiers *jàré* ‘much/many’ and *kùdík* ‘a few/little’ and shares features with ADJ3 and ADJ4.<sup>12</sup> They have two forms, with or without the gender marking for both NP and Modification uses. The gender-marked forms (with no number marking) imply contrast (*ná/ló-jórè*, *ná/ló-kúdík* ‘(of) larger/smaller amount.F/M’). When used predicatively, they lack gender marking and the copula.

<sup>12</sup> Otherwise, we can include *lîŋ* ‘all’ and numerals in ADJ5, which lack the copula in the predicate use (e.g., *lòpèkât lîŋ* (3PL all) ‘all of them’/‘They are all.’, *lòpèkât mûsálâ*. (3PL three) ‘they three’/‘They are three.’). However, they lack the gender-marked ‘contrastive’ form (but cf. *ná/ló-gélèŋ* ‘only, sole’).

- (21) a. kùḍḍík  
little  
'a small amount'
- b. ná-kúḍḍík  
F-little  
'the one with a smaller amount'
- (22) a. píòŋ kùḍḍík  
water(PL.F) little  
'a little water'
- b. píòŋ ná-kúḍḍík  
water(PL.F) F-little  
'the smaller.F (bottle of) water'
- (23) a. lòpèŋâ jòré.  
3PL many  
'They are numerous.'
- b. \*lòpèŋâ à ná-jórè.  
3PL COP F-many

## [5] Verb-based nominalizations

This section describes the realization of gender marking on verb-based nominalization strategies.

### [5a] Relative clauses

Bari has two types of relativizers (short and long forms), both with gender marking (F *ná*, *nágwôn* and M *ló*, *lógwôn*; the long form is etymologically related to *gwôn* 'be, stay'), which require different constructions. Since these constructions need careful descriptions and fall largely beyond the scope of this report, we will omit the details here (see Nakao 2024). In a nutshell, the long forms have wider uses and do not affect the clause structure, but the short forms neutralize tense-aspect-modality distinctions and are always followed by a verb.

- (24) a. gwándà lógwôn ɲótú à ɲá lò  
cassava(PL.M) REL.M person PRET eat:DOM DEF.SG.M  
'the cassava that.M the man ate'
- b. gwándà ló ɲá ɲótù ló  
cassava(PL.M) REL.M eat:DOM person DEF.SG.M  
'the cassava that.M the man ate/is eating/will eat'

There are a number of stative verbs that are usually used in the non-past form (cf. inchoative: *lík-án* ‘become cold’). In non-predicate uses they require relativization.

- (25) kúnè            lí~lík.  
 PROX.PL.F NPST~be.cold  
 ‘It.PL.F (e.g., water) is cold.’

- (26) píòŋ            ná            lí~lík  
 water(PL.F) REL.F            NPST~be.cold  
 ‘cold.F water’

The feminine long relativizer also functions as the time clause marker (*nágwôn* ‘when’), probably referring to the noun *dìŋít* ‘time (F)’. The feminine relativizers also appear in the manner clause (with the preposition *gwòsò* ‘as’, perhaps referring to *kùljâ* ‘word, matter (F)’).

- (27) nágwôn            nân a            ʔjá            lópèŋ í            màkàtàb            nápít,            ...  
 REL.F/when            1SG PRET visit:DOM            3SG LOC office(F)            SG.F:GEN.3SG  
 ‘When I visited him at his office, ...’

- (28) a.    gwòsò            nágwôn            ɲótô            kú~kùljâ, ...  
           as            REL.F            person.PL            NPST~say
- b.    gwòsò            ná            kùljári            ɲótô, ...  
           as            REL.F            say:APPL.INST            person.PL  
           ‘As people say, ...’

Bari has focus markers that have developed from the relativizers (agent focus: F *nâ*, M *lâ*; other focus: F *nàgwôn*, M *lògwôn*). They agree with the focalized noun in gender and are followed by a ‘focus’ form of the verb.

- (29) jî            lò            dèrjá            súkùrì.  
 1PL FOC.M            FOC:cook            chicken  
 ‘We<sub>[focus]</sub> cooked the chicken.’

- (30) súkùrì            nàgwôn            nân à            dér.  
 chicken(F)            FOC.F            1SG PRET cook:DOM  
 ‘It is *chicken*<sub>[focus]</sub> that I cooked.’

## [5b] Complement clauses

Complement clauses are marked with *àdɪ* ‘COMP’, which is not a nominalizer by itself. A noun-modifying complement clause requires a relativizer (32). Given its realization in (33), where the main verb (if it is ‘said’) can be deleted, it might represent a case of grammaticalization of a verbal phrase.

- (31) bàbá jèjéjù àdí nân dê pó~pô.  
 my.father FREQ:think COMP 1SG FUT NPST~come  
 ‘My father thinks that I am coming.’
- (32) dó àgú à jíj jámêť nágwôn àdí nân à nòlâ?  
 2SG Q PRET hear:DOM word(F) F.REL COMP 1SG PRET be.happy  
 ‘Have you heard the expression *nan a nyola* (‘I am happy’)?’
- (32) bàbá (à kùljâ) àdí jé pó~pô.  
 my.father PRET say COMP 3SG NPST~come  
 ‘My father said that he is coming.’

### 3. DATA

This section demonstrates data on the realizations of gender marking of various word classes with regards to their NP and Modification (Mod) uses.

The results are summarized in Figure 1 in accordance with Shibatani’s (2021: 524, 2023: 35–37) hierarchy. As shown, the general lack of gender marking for numerals and ADJ5 does not support his idea that the left-side categories will tend to have gender marking, but the optional gender marking on ordinal numerals and obligatory gender marking on ADJ4 only in NP use do support his idea that gender marking will tend to appear on NP uses when Modification use lack it (and not *vice versa*), deduced from his hypothesis that the classifier/gender marking serves primarily for nominalization.

Figure 1. Classifier/Gender/Noun class marking pattern in Bari

|          |                                                                      |
|----------|----------------------------------------------------------------------|
| NP use:  | *CARD.NUM > DEM > GEN > ADJ1/2/3/4/(*)5 > V-based NMLZ<br>OPTORD.NUM |
| Mod use: | *CARD.NUM > DEM > GEN > ADJ1/2/3/*4/(*)5 > V-based NMLZ<br>*ORD.NUM  |

(\* = no marking or impossible; OPT = optional; unmarked = obligatory marking)

#### [1] Numerals

Contrary to Shibatani’s implicational hierarchy along the structural dimension, Bari does not mark gender on cardinal numerals, even though gender marking occurs with other types of structures on the hierarchy. On the other hand, ordinal numerals can have optional gender marking (with a relativizer) only in the NP use, without any semantic difference.



## [1a] Cardinal numerals

NP use

- (34) nân dɛ̀kɛ̀ gɛ̀lɛ̀ŋ/mùsálâ.  
1SG IPFV:want one/three  
'I want one/three.'

Mod use

- (35) a. nân à mèddjâ ɲútú gɛ̀lɛ̀ŋ.  
1SG PRET see person one  
'I saw a person (man/woman).'
- b. nân à mèddjâ ɲútú mùsálâ.  
1SG PRET see person.PL three  
'I saw three persons (men/women).'

## [1b] Ordinal numerals

NP use

- (36) nân ɲà~ɲár tò-gɛ̀lɛ̀ŋ.  
1SG NPST~love ORD-one  
'I love the first one.'
- (37) nân ɲà~ɲár nágwôn tò-gɛ̀lɛ̀ŋ.  
1SG NPST~love REL.F ORD-one  
'I love the first one.F.'

Mod use

- (38) ɲínà à dɪ̀ɲít (\*nágwôn) tò-gɛ̀lɛ̀ŋ.  
MED.SG.F COP time(F) REL.F ORD-one  
'It is the first time.'

## [2] Demonstratives

Demonstratives always require gender (and, as a portmanteau morpheme, number) marking.

NP use

- (39) nân dɛ̀kɛ̀ kúnɛ̀.  
1SG IPFV:want PROX.PL.F  
'I want these (e.g., flowers).'

Mod use

- (40) nân ðèkè kúné kátúràn.  
1SG IPFV:want PROX.PL.F flower.PL.F  
'I want these.F flowers.'

### [3] Genitive noun phrases

Genitive noun phrases always require gender (and/or number) marking (the gender distinction is neutralized in the plural number as mentioned in 2.2). The gender marking refers to the property of the possessum.

NP use

- (41) nân jà~jár ñìnòt.  
1SG NPST~love SG.F:GEN.2SG  
'I love yours.F (e.g., food)'
- (42) dō à jál nà bàbá!  
2SG PRET eat:DOM GEN.SG.F my.father  
'You ate my father's.F!'

Mod use

- (43) nân jà~jár kínò ñìnòt.  
1SG NPST~love food.F SG.F:GEN.2SG  
'I love your.F food.'
- (44) dō à jál kínò ná bàbá!  
2SG PRET eat:DOM food.F GEN.SG.F my.father  
'You ate my father's.F food!'

### [4] Adjectives

ADJ1/2/3 and the contrastive use of ADJ5 always require gender (and, if any, number) marking, but the non-NP uses of ADJ4 and the normal use of ADJ5 always lack it. ADJ4 requires gender marking (with the relativizer) only in the NP use. As mentioned in 2.2, ADJ3 and contrastive ADJ5 lack gender marking in predicate use, but the data are not presented here. Note that ADJ2 requires a relativizer for its NP use (in parallel with ADJ4).

ADJ1

NP use

- (45) dō à jésú nân?  
2SG PRET eat which.SG.F  
'Which.F (e.g., food) did you eat?'

- (46) nân đêkè lélé.  
 1SG IPFV:want certain  
 ‘I want one/any.M (of them).’

Mod use

- (47) dó à jésú nán kípò?  
 2SG PRET eat which.SG.F food(F)  
 ‘Which.F food did you eat?’

- (48) nân đêkè gálàm lélé.  
 1SG IPFV:want pen(M) certain.M  
 ‘I want a.M pen./I want any.M of the pens.’

ADJ2

NP use

- (49) a. nân jà~jár nágwôn à ná-ké.  
 1SG NPST~love REL.F COP F-beautiful.SG  
 ‘I love the beautiful.F one.F.’  
 b. \*nân jà~jár ná-ké.  
 1SG NPST~love F-beautiful.SG

Mod use

- (50) nân jà~jár kâk ná-ké.  
 1SG NPST~love land(F) F-beautiful.SG  
 ‘I love the beautiful.F land.’

ADJ3

NP use

- (51) nân đêkè ná-pápè.  
 1SG IPFV:want F-hot  
 ‘I want the hot one.F (e.g., tea).’

Mod use

- (52) nân đêkè sâi ná-pápè.  
 1SG IPFV:want tea(PL.F) F-hot  
 ‘I want hot.F tea.’

ADJ4

NP use

- (53) a. \*nân đêkè dúmà.  
 1SG IPFV:want big  
 ‘I want the big one.’

- b. nân đêkê nágwôn à dómà.  
 1SG IPFV:want REL.F COP big  
 ‘I want the big one.F.’

Mod use

- (54) nân đêkê lókóré dùmà.  
 1SG IPFV:want meat(F) big  
 ‘I want the big meat.’

ADJ5

NP use

- (55) nân đêkê kùdfík.  
 1SG IPFV:want little  
 ‘I want a little.’

- (56) nân đêkê ná-kúdfík.  
 1SG IPFV:want F-little  
 ‘I want the one.F with the smaller amount.’

Mod use

- (57) nân đêkê píòŋ kùdfík.  
 1SG IPFV:want water(PL.F) little  
 ‘I want a little water.’

- (58) nân đêkê píòŋ ná-kúdfík.  
 1SG IPFV:want water(PL.F) F-little  
 ‘I want the small(er).F (bottle of) water.’

## [5] Verb-based nominalizations

Since complement clauses are not NPs in Bari, we only give examples of relative clauses in this section. Relative clauses (with both long and short relativizers) always require gender marking. However, example (61) is ungrammatical according to my informant. This may imply that the NP use of a relative clause is limited to non-human NPs, but further investigation is needed.

NP use

- (59) nân đêkê ná lí~lík.  
 1SG IPFV:want REL.F NPST~be.cold  
 ‘I want (the) cold one.F.’

- (60) nân đêkê nágwôn dós à dér nà.  
 1SG IPFV:want REL.F 2SG PRET cook:DOM DEF.SG.F  
 ‘I want the one.F that you cooked.’

- (61) \*nân à mét lógwôn à pô ló.  
 1SG PRET see:DOM REL.M PRET come DEF.SG.M  
 (intended: ‘I saw the one who came.’)

Mod use

- (62) nân dèkè píòŋ ná lí~lík.  
 1SG IPFV:want water(PL.F) REL.F NPST~be.cold  
 ‘I want (the) cold.F tea.’

- (63) nân dèkè kíjón nágwôn dó à dër nà.  
 1SG IPFV:want food(F) REL.F 2SG PRET cook:DOM DEF.SG.F  
 ‘I want the food that.F you cooked.’

- (64) nân à mét ɲútù lógwôn à pô ló.  
 1SG PRET see:DOM person REL.M PRET come DEF.SG.M  
 ‘I saw the man who.F came.’

#### 4. ANALYSIS AND DISCUSSION

As shown in Section 3, although Bari marks gender on demonstratives, genitive noun phrases, most types of adjectives, and verb-based nominalizations, it lacks gender marking on numerals (as well as a few underived adjectives). This contradicts Shibatani’s implicational hierarchy along the structural dimension, which regards numerals as the structure most likely to receive classifier/gender marking (Shibatani 2023: 34). On the other hand, the Bari data support Shibatani’s hypothesis along the functional dimension, since it explains the optional gender marking on ordinal numerals and the obligatory gender marking on ADJ4 only for the NP use. Diachronically speaking, it is noteworthy that the non-Bari branches of Eastern Nilotic attest to gender marking on (lower) numerals (Heine & Vossen 1983; see, for example, Dimmendaal 1983 for Turkana, Tucker & Mpaayei 1955 for Maasai, Moodie & Billington 2020 for Lopit), although it remains unclear whether the Bari case represents attrition at this point.<sup>13</sup>

Currently, Shibatani’s hierarchy does not take into consideration the (relative) lack of gender marking on (copula-less) predicates in languages such as Bari (as explained in 2.2; predicate demonstratives being notable exceptions), and that aspect was not the central concern of this report. However, this point can also be deductively explained by this theory, which considers the primary role of gender marking to be nominalization.

Another point on which Bari paves way for the development of this theory is that grammatical number may also have a role in nominalization. In Bari, many determiners (demonstratives and ADJ1) code both gender and number in portmanteau morphemes, while some word classes lack either gender (some ADJ4) or number (relativizer), and these properties cannot be discussed separately.

<sup>13</sup> Note, however, that the non-Bari branches (except for Lopit) also have gender marking on nouns unlike Bari, which cannot be reconstructed for Proto-Eastern Nilotic (Heine & Vossen 1983).

Gender marking in Bari deserves more in-depth investigation both synchronically and diachronically. Nyombe (2004) provides some competing data on Bari gender marking based on his native intuitions, which contrast with my data in some way or another at this point. Although Bari gender marking is mostly coherent in both the NP and Modification uses, this situation might be a recent development. It should be recalled that Heine (1980: 19) clearly states that one of the functions of grammatical gender is nominalization, citing instances from the Eastern Nilotic language Sampur (Lotuho-Maa sub-branch) *-inó* ‘your’ vs. *I-inó* ‘yours.M’, *torróno* ‘bad’ vs. *I-torróno* ‘the bad one.M’ (*I-* ‘M’). A similar phenomenon is attested in the genitive system of Nyangatom, an Eastern Nilotic of the Teso-Turkana sub-branch, where gender marking is optional in Modification use but obligatory in NP use.<sup>14</sup>

Mod use

- (65) à-bèrú      kân  
F-woman   GEN.1SG  
‘my wife’

- (66) à-kân      à-bèrú  
F-GEN.1SG F-woman  
‘my.F wife’

NP use

- (67) à-bèrú      ná-kân  
F-woman   F-GEN.1SG  
‘The woman is mine.F.’<sup>15</sup>

The emergence of gender in Eastern Nilotic itself is still a largely uninvestigated problem. Although there was a hypothesis that it emerged from contacts with East Cushitic languages, an idea that has some historical and linguistic grounds, Heine & Vossen (1983) express doubt on the issue, partially due to the fact that the gender markers are not Afroasiatic (Proto-Eastern Nilotic *\*na* ‘F’ might be grammaticalized from *\*naa* ‘girl’ and *\*lo* ‘M’ a reanalysis of human noun prefix).<sup>16</sup> Also, attention should be given to the fact that Ubangian (Niger-Congo) languages such as Zande spoken near the Bari-speaking areas also have a (seemingly) newly developed gender system (Zande *na-* ‘F’, *ba-* ‘M’; Gore 1926).

## 5. CONCLUSION

This report provided basic Bari nominal morphosyntax with focus on the gender marking in various word classes and two grammatical environments (NP and Modification uses).

<sup>14</sup> The data were collected by me in Jinka, Ethiopia, in March 2025, from Mr. L.K. from Kankaten.

<sup>15</sup> Nyangatom lacks a copula and this predicate use is considered a NP.

<sup>16</sup> While Bari nouns lack overt gender marking, *lòṇàsêr* ‘brother’ and *kìṇàsêr* ‘sister’ have a prefix-like element, which might be fossilized gender markers. Interestingly, there are few Bari nouns that have word-initial /n/, while *nárákwán* ‘woman’ is a notable exception.

The result shows that Bari generally marks gender on non-noun nominals for both uses if the morpheme has any gender-marking. However, there are two notable exceptions. First, Bari lacks gender marking on cardinal numerals, which contradicts Shibatani's (2021, 2023) implicational hierarchy along the structural dimension. Second, the optional gender marking on ordinal numerals and the obligatory gender marking on ADJ4 (when used with a relativizer) are consistent with Shibatani's hierarchy along the functional dimension,

## ABBREVIATIONS

|              |                                     |
|--------------|-------------------------------------|
| *            | ungrammatical/reconstructed form    |
| ~            | reduplication                       |
| 1/2/3        | person                              |
| ADJ1/2/3/4/5 | adjective (class1/2/3/4/5)          |
| APPL.INST    | applicative (instrumental)          |
| COMP         | complementizer                      |
| COP          | copula                              |
| DEF          | definite                            |
| DEM          | demonstrative                       |
| DIST         | distal                              |
| DOM          | differential object marking         |
| F            | feminine                            |
| FOC          | focus                               |
| FREQ         | frequentative                       |
| FUT          | future                              |
| GEN          | genitive                            |
| IMP          | imperative                          |
| INCH         | inchoative                          |
| IPFV         | imperfective                        |
| LOC          | locative                            |
| M            | masculine                           |
| MED          | medial                              |
| Mod          | modification                        |
| NMLZ         | nominalization                      |
| NP           | noun phrase                         |
| NPST         | non-past                            |
| NUM          | numeral                             |
| ORD          | ordinal                             |
| PL           | plural                              |
| PRED         | predicate (forms of demonstratives) |
| PRET         | preterite                           |
| PROX         | proximal                            |
| PSTV         | presentative                        |
| PURP         | purpose                             |
| REL          | relativizer                         |

SG                      singular

## ACKNOWLEDGMENTS

I appreciate Dr. Gore David for providing Bari data. This study is supported by JSPS KAKENHI Grant-in-Aid (Grant Numbers: 19K13160, 22H00659/23K21931, 24K16054, 24KK0007) and the National Institute for Japanese Language and Linguistics project ‘Empirical study of the typology of nominalization: From theoretical, fieldwork, historical and dialectal perspectives’ (PI: Prashant Pardeshi).

## REFERENCES

- Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.) 2025. *Ethnologue: Languages of the world*. Twenty-eighth edition (Online version). Dallas, Texas: SIL International. <http://www.ethnologue.com>.
- Dimmendaal, Gerrit J. 1983. *The Turkana language*. Dordrecht: Foris Publication.
- Gore, Edward Clive. 1926. *A Zande grammar*. London: Sheldon Press.
- Heine, Bernd. 1980. *The non-Bantu languages of Kenya*. Berlin: Dietrich Reimer.
- Heine, Bernd & Rainer Vossen. 1983. On the origin of gender in Eastern Nilotic. In Rainer Vossen & Marianne Bechhaus-Gerst (eds.) *Proceedings of the International Symposium on Languages and History of the Nilotic Peoples, Cologne, January 4–6, 1982, Vol. 2*, 245–268. Berlin: Reimer.
- Moodie, Jonathan & Rosey Billington. 2020. *A grammar of Lopit: An Eastern Nilotic language of South Sudan*. Leiden: Brill.
- Nakao, Shuichiro. 2024. Between antipassive and differential object marking in Bari: A markedness paradox and the antipassive-to-active shift. *Working Papers in African Linguistics* 2, 53–82.
- Nakao, Shuichiro. 2025. Bari-go (Higashi-Nairu-goha) gaisetsu (in Japanese: An introduction to Bari, Eastern Nilotic). *Journal of Swahili and African Studies* 36, 133–153.
- Nyombe, Bureng G.V. 2004. Null heads and DP-internal concord in Bari. *Occasional Papers in Language and Linguistics* 2, 17–44.  
<https://uonjournals.uonbi.ac.ke/ojs/index.php/unjll/article/view/1655>
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Suita: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf-char/en)
- Spagnolo, Lorenzo M. 1933. *Bari grammar*. Verona: Missioni Africane.
- Tucker, Archibald N. & J. Tompo Ole Mpaayei. 1955. *A Maasai grammar*. London: Longman, Green & Co.
- Vossen, Rainer. 1982. *The Eastern Nilotes: Linguistic and historical reconstructions*. Berlin: Dietrich Reimer.



Yokwe, Eluzai Moga. 1987. The Tonal grammar of Bari. Doctoral dissertation, University of Illinois at Urbana-Champaign.

# Noun class system and grammatical nominalisations in the Kimakunduchi dialect of Swahili

Makoto Furumoto<sup>1</sup>

## ABSTRACT

The present study describes grammatical nominalisations in the Kimakunduchi dialect of Swahili. Kimakunduchi, like other Swahili varieties and Bantu languages, has a noun class system. Most Kimakunduchi nouns and grammatical nominalisations are accompanied by an overt marker indicating their noun class, and this marker appears in both noun-phrase use and modification use. This observation suggests that Kimakunduchi can be analysed as a mature language at an advanced stage in Shibatani's (2021: 524, 2023: 35) hierarchy, which illustrates the typological and developmental patterns of classifier and gender/noun class marking.

## 1. INTRODUCTION

Swahili is a Bantu language with a wide range of varieties. Kimakunduchi, also known as Kikae or Kihadimu,<sup>2</sup> is one of the coastal Swahili varieties that spread much earlier than inland varieties. Kimakunduchi is spoken in the south-eastern part of Unguja Island in Zanzibar, Tanzania. Kimakunduchi may appear similar to Standard Swahili, a prestigious variety of Swahili. However, when drawing a family tree that illustrates the genetic relationships among coastal Swahili varieties, it is likely that Kimakunduchi is not a derivative of the Kiunguja variety, which served as the basis for Standard Swahili (Nurse 1982, Nurse & Hinnebusch 1993), and detailed descriptions have revealed that Kimakunduchi has developed distinctive features and a unique system (e.g., Furumoto 2015, 2024a, Furumoto & Gibson 2022).

According to Shibatani (2019, 2021), forms that function as modifiers of nouns are understood as grammatical nominalisations, which, like nouns, denote entities. This study aims to describe grammatical nominalisations in Kimakunduchi, with a focus on their noun-phrase (NP) use and modification (Mod) use. More specifically, it primarily deals with numerals (both ordinal and cardinal), demonstratives, genitive noun phrases, adjectives, and relative clauses. Unless otherwise noted, the data presented here were collected through the author's on-site fieldwork.

Kimakunduchi, like many other Swahili varieties and Bantu languages, has a noun class system. Every noun belongs to one of these classes, and most nouns are accompanied by a prefix that indicates their class. Similarly, most grammatical nominalisations vary in form according to the noun class, which is mostly realised by prefixation. These formal features are described in detail in Section 2.

---

<sup>1</sup> Teikyo University of Science/Tokyo University of Foreign Studies.

<sup>2</sup> In Swahili and related Bantu languages, language names are accompanied by the prefix *ki-* or its cognate. The language name Kimakunduchi can be decomposed into this prefix *ki-* and *makunduchi*, the latter of which refers to the place where Kimakunduchi is spoken.

In Kimakunduchi, when grammatical nominalisations co-occur with nouns to form a single noun phrase, they consistently follow the noun. Demonstratives can occur pre-nominally, but in such cases the demonstrative and the noun are assumed to be in an appositive relation, thus forming separate noun phrases (for more details, see Section 3).

In Kimakunduchi, it is rare for a noun to be followed by more than two grammatical nominalisations. The observations of possible combinations of two grammatical nominalisations indicate that there is a fixed order of grammatical nominalisations, as summarised in (1).

- (1) GEN.PRO – NUM – ADJ – DEM – GEN (except GEN.PRO) – V-based NMLZ (RC)

## 2. OVERVIEW OF NOUN CLASSIFICATION AND GRAMMATICAL NOMINALISATIONS IN KIMAKUNDUCHI

### 2.1. Noun classification and its marking in Kimakunduchi

In Bantu languages, nouns tend to exhibit forms which can be decomposed into a prefix and a stem,<sup>3</sup> and in Bantu linguistics, there is a convention to classify nouns based on their prefixes. This approach to noun classification has been developed in comparative linguistic studies aimed at reconstructing Proto-Bantu (see Maho 1999, Katamba 2003, Van de Velde 2019). In this conventional classification system, each noun class is assigned a unique number. Singular and plural forms of nouns usually share stems and have different prefixes, which means that they are sorted into separate classes with distinct numbers. Singular nouns are placed in odd-numbered classes, and their plural counterparts belong to the following even-numbered class, with some exceptions.

Kimakunduchi nouns can be classified not only according to their prefixes but also according to the forms associated with nouns (e.g., adjectives, genitives, demonstratives, agreement/pronominal markers on verbs), which is a more typical way of identifying grammatical gender and noun class (e.g., Hockett 1958, Greenberg 1978, Heine 1982, Dixon 2010). In Kimakunduchi, the classification of nouns based on their own forms (i.e., noun prefixes) and that based on their associated forms mostly coincide, but there are some discrepancies. Table 1<sup>4</sup> shows examples of nouns from each class, and their corresponding forms for the adjective ‘small’, the genitive pronoun ‘my/mine’, the proximal demonstratives ‘this/these’, and relative clause markers (RMs), which can accompany verbs.<sup>5</sup> The numbers in

<sup>3</sup> More strictly, in many Bantu languages, but not in Kimakunduchi, nouns are also accompanied by a prefix, called an ‘augment’. This ‘augment’ prefix appears before the noun prefix which is related to noun classification.

<sup>4</sup> For transcription of examples, the orthography of Standard Swahili is used with the following additions: aspiration and nasal syllabicity are marked with IPA symbols, the first characters of sentences and proper nouns are written in lower case, periods are not added at the end of sentences, commas are added to represent prosodic break, Ø is used for the zero prefix, morpheme boundaries are signalled with hyphens (affixes), equals (clitics), or plus signs (compounds), and asterisks mark unacceptability without specifying whether this is due to ungrammaticality or other factors (e.g., pragmatic infelicity). Numbers in the gloss primarily express noun class information. While the first and second persons are represented by numbers 1 and 2 in the same way as classes 1 and 2, they, unlike the class numbers, are shown together with SG or PL. The transcriptions, glosses, and translations of examples from previous studies have been modified to facilitate analysis.

<sup>5</sup> Not all possible allomorphs of the prefixes are covered in Table 1.

parentheses indicate the classification based on the associated forms of nouns. They are only displayed when the two-classification criteria conflict.

Like nouns, most grammatical nominalisations can also be segmented into a prefix and a stem. To clearly demonstrate this morphological feature, a hyphen is inserted between the prefix and the stem. This hyphenation does not necessarily mean that prefixation is synchronically active for word formation (see also Section 4). In addition to the forms shown in Table 1, there are other forms of adjectives, genitives, demonstratives, and RMs. Most of the nouns corresponding to these forms come from verbs (i.e., infinitives) or from other nouns (i.e., locative nouns). These other forms, also assigned unique class numbers, are introduced where necessary.

Table 1. Examples of Kimakunduchi nouns and their associated forms

| class | Nouns                                | ADJ<br>'small'     | GEN PRO<br>'my/mine' | DEM<br>'this'  | RM     |
|-------|--------------------------------------|--------------------|----------------------|----------------|--------|
| 1     | m-t <sup>h</sup> u 'person' (sg.)    | m-dogo             | y-angu               | yu-no          | m-, e- |
| 2     | wa-t <sup>h</sup> u 'persons' (pl.)  | wa-dogo            | w-angu               | wa-no          | o-     |
| 9(1)  | n-dege 'bird(s)' (sg./pl.)           | m-dogo             | y-angu               | yu-no          | m-, e- |
| 10(2) | askari 'police officer(s)' (sg./pl.) | wa-dogo            | w-angu               | wa-no          | o-     |
| 3     | m-kono 'hand' (sg.)                  | m-dogo             | w-angu               | u-no           | o-     |
| 4     | mi-kono 'hands' (pl.)                | mi-dogo            | y-angu               | i-no           | yo-    |
| 11(3) | u-jiti 'tree' (sg.)                  | m-dogo             | w-angu               | u-no           | o-     |
| 4     | mi-jiti 'trees' (pl.)                | mi-dogo            | y-angu               | i-no           | yo-    |
| 11(3) | u-funguo 'key' (sg.)                 | m-dogo             | w-angu               | u-no           | o-     |
| 10~4  | funguo~mi-funguo 'keys' (pl.)        | n-dogo~<br>mi-dogo | z-angu~<br>y-angu    | zi-no<br>~i-no | zo~yo- |
| 11(3) | u-kuni 'firewood' (sg.)              | m-dogo             | w-angu               | u-no           | o-     |
| 10    | k <sup>h</sup> uni 'firewood' (pl.)  | n-dogo             | z-angu               | zi-no          | zo-    |
| 5     | tunda 'fruit' (sg.)                  | dogo               | ly-angu              | li-no          | lyo-   |
| 6     | ma-tunda 'fruits' (pl.)              | ma-dogo            | y-angu               | ya-no          | yo-    |
| 7     | ki-t <sup>h</sup> u 'thing' (sg.)    | ki-dogo            | ch-angu              | ki-no          | cho-   |
| 8     | vi-t <sup>h</sup> u 'things' (pl.)   | vi-dogo            | vy-angu              | vi-no          | vyo-   |
| 9     | n-guo 'clothes' (sg./pl.)            | n-dogo             | y-angu               | i-no           | yo-    |
| 10    | embe 'mango(es)' (sg./pl.)           | n-dogo             | z-angu               | zi-no          | zo-    |

The referents of nouns in classes 1/2, including those classified based on their associated forms, are consistently animate, but animate nouns do not necessarily belong to classes 1/2; there are also nouns in other classes that denote animate entities (e.g., *m-kunga/mi-kunga* 'eel(s)' [cl.3/4], *bata/ma-bata* 'duck(s)' [cl.5/6], *ki-tu/vi-tu* 'black cat(s)' [cl.7/8]). For other classes, the correlation between noun classification and semantic feature is more limited. For example, nouns that refer to trees or plants are included in classes 3/4 (e.g., *m-nazi/mi-nazi*

‘coconut tree(s)’); there are some nouns that denote fruits in classes 5/6 (e.g., *fenesi/ma-fenesi* ‘jack fruit(s)’); names of languages (e.g., *ki-kae* ‘the Kae (Makunduchi) language’) are included in class 7.

Classes 5/6 include augmentative nouns referring to large entities. In class 5, a very limited number of nouns have the prefix *ji-* (e.g., *ji-no/me-no* ‘tooth/teeth’), and augmentative nouns also take this prefix (e.g., *ji-t<sup>h</sup>u/ma-ji-t<sup>h</sup>u* ‘giant(s)’). Diminutive nouns for small entities belong to classes 7/8 (e.g., *ki-dege/vi-dege* ‘small bird(s)’).

There are nouns in both class 5 and classes 9/10 that lack noun prefixes, which means that their classes cannot be identified by their own morphological features. Furthermore, the noun prefixes of classes 1 and 3, which are classes of singular forms, are identical, whereas their corresponding plural forms of classes 2 and 4 have different prefixes. As this suggests, it is sometimes necessary to consider both the singular and plural forms, or the forms associated with nouns, when classifying nouns, rather than referring to only one of them. In the same vein, even for associated forms, it is not necessarily possible to detect their noun classes on the basis of their own formal features (e.g., genitives of classes 1 and 9, associated forms of classes 3 and 11). These observations suggest that nouns and their associated forms are not always accompanied by markers that are sufficiently distinctive for noun classification.

Classes 9/10 differ from other classes in that the noun forms are identical between singular and plural. This means that it is impossible to identify noun prefixes by comparing singular and plural forms. While there are nouns in these classes that have no prefix, a number of nouns in these classes are analysed as having a prefix characterised by a nasal. This analysis can be supported by noun-to-noun derivations, realised by changing the noun prefix, e.g., *n-dege* ‘bird(s)’ to *ki-dege* ‘small bird’.

For nouns with the class 11 prefix *u-* in the singular, the corresponding plural forms belong to either class 4 or class 10, and some class 11 singular nouns have both class 4 and class 10 plural forms. Although class 11 forms have a different noun prefix from class 3 forms, their associated forms are the same. The variation in plural forms corresponding to class 11, along with the shared associated forms of classes 3 and 11, can be seen as indicative of an ongoing process in which class 11/10, originally independent of class 3/4, is merging into class 3/4 (Furumoto 2015).

## 2.2. Grammatical nominalisations in Kimakunduchi

The word forms of grammatical nominalisations in Kimakunduchi are explained in more detail below.

### [1] Numerals

#### [1a] Cardinal numerals

Cardinal numerals expressing ‘one’, ‘two’, ‘three’, ‘four’, and ‘eight’ can be analysed as consisting of a prefix and a stem. The form of the prefix varies according to the noun class. In most cases, these prefixes are identical to those found on nouns, with the exception of class 11, where the numeral ‘one’ is accompanied by the prefix *m-*, as in class 3. The class 5 form (*moja* ‘one’) lacks an overt prefix. Classes 9 and 10 share numeral forms; some numerals begin with

a nasal that can be analysed as a prefix (*m-bili* ‘two’, *n-ne* ‘four’), others are characterised by the initial aspiration (*t<sup>h</sup>atu* ‘three’, *t<sup>h</sup>ano* ‘five’), and some have no explicit prefix (*moja* ‘one’, *nane* ‘eight’). The numerals ‘six’, ‘seven’, and any numerals larger than ‘eight’ have no prefix.

Prefixes accompanying numerals show the same variation as those of adjectives; numerals cannot be distinguished from adjectives by the prefix they take. Odd-numbered noun classes are associated with singular marking and do not include numerals denoting ‘two’ or greater, whereas even-numbered classes, related to plural marking, lack a numeral for ‘one’.<sup>6</sup> Only the exception can be seen in class 6, includes the singular form, *ma-moja*, which can be used to count *maziko* ‘funeral’.

Table 2. Cardinal numerals in Kimakunduchi

| class | ‘one’   | ‘two’   | ‘three’            | ‘four’ | ‘five’             | ‘six’ | ‘seven’ | ‘eight’ |
|-------|---------|---------|--------------------|--------|--------------------|-------|---------|---------|
| 1     | m-moja  | N/A     | N/A                | N/A    | N/A                | N/A   | N/A     | N/A     |
| 2     | N/A     | wa-wili | wa-tatu            | wa-ne  | wa-tano            | sita  | saba    | wa-nane |
| 3, 11 | m-moja  | N/A     | N/A                | N/A    | N/A                | N/A   | N/A     | N/A     |
| 4     | N/A     | mi-wili | mi-tatu            | mi-ne  | mi-tano            | sita  | saba    | mi-nane |
| 5     | moja    | N/A     | N/A                | N/A    | N/A                | N/A   | N/A     | N/A     |
| 6     | ma-moja | ma-wili | ma-tatu            | ma-ne  | ma-tano            | sita  | saba    | ma-nane |
| 7     | ki-moja | N/A     | N/A                | N/A    | N/A                | N/A   | N/A     | N/A     |
| 8     | N/A     | vi-wili | vi-tatu            | vi-ne  | vi-tano            | sita  | saba    | vi-nane |
| 9     | moja    | N/A     | N/A                | N/A    | N/A                | N/A   | N/A     | N/A     |
| 10    | N/A     | m-bili  | t <sup>h</sup> atu | n-ne   | t <sup>h</sup> ano | sita  | saba    | nane    |

### [1b] Ordinal numerals

Ordinal numerals from ‘first’ to ‘eighth’ are presented in (2). The ordinal numerals greater than ‘eight’ share the form with the cardinal numerals. In Kimakunduchi, ordinal numerals appear with a genitive marker, as in (3). The forms of these genitive markers, which vary according to the noun class, is summarised in Table 4.

- (2) kwaza ‘first’  
 pili ‘second’  
 tatu ‘third’  
 ne ‘fourth’  
 tano ‘fifth’  
 sita ‘sixth’  
 saba ‘seventh’  
 nane ‘eighth’

<sup>6</sup> Strictly speaking, *-moja* can be used with prefixes from even-numbered noun classes. In such cases, it primarily expresses unity or identity, which is another use of *-moja* rather than denoting ‘one’.

- (3) siku ya=pili  
 9.day GEN.9=second  
 ‘The second day’

## [2] Demonstratives

Kimakunduchi has proximal, medial, and distal demonstratives. Demonstrative forms vary according to the noun class and location of the referent. The forms of the Kimakunduchi demonstrative are summarised in Table 3. This table presents not only the forms for classes 1 to 10 but also for classes 16, 17, and 18. Demonstratives of classes 1–11 refer to (in)animate objects, whereas those of classes 16, 17, and 18 (locative classes) point to places. For more details on the differences in referents for demonstratives in the locative classes, see Furumoto (2024a).

Table 3. Demonstratives in Kimakunduchi (Furumoto 2024a, see also Racine-Issa 2002)

|       | Proximal     |            |              | Medial |            |              | Distal |
|-------|--------------|------------|--------------|--------|------------|--------------|--------|
| class | Basic        | Contracted | Reduplicated | Basic  | Contracted | Reduplicated | Basic  |
| 1     | yuno         | =yu        | yuyu(yu)     | uyo    | =yo        | yuyo(yo)     | yulya  |
| 2     | wano         | =wa        | wawa(wa)     | wao    | =o         | wao(o)       | walya  |
| 3, 11 | uno          | =u         | uu(u)        | uo     | =o         | uo(o)        | ulya   |
| 4     | ino          | =i         | ii(i)        | iyo    | =yo        | iyo(yo)      | ilya   |
| 5     | lino         | =li        | lili(li)     | ilyo   | =lyo       | lilyo(lyo)   | lilya  |
| 6     | yano         | =ya        | yaya(ya)     | yayo   | =yo        | yayo(yo)     | yalya  |
| 7     | kino         | =ki        | kiki(ki)     | icho   | =cho       | kicho(cho)   | kilya  |
| 8     | vino         | =vi        | vivi(vi)     | ivyo   | =vyo       | vivyo(vyo)   | vilya  |
| 9     | ino          | =i         | ii(i)        | iyo    | =yo        | iyo(yo)      | ilya   |
| 10    | zino         | =zi        | zizi(zi)     | izo    | =zo        | zizo(zo)     | zilya  |
| 16    | vano         | =va        | vava(va)     | avo    | =vo        | vavo(vo)     | valya  |
| 17    | kuno         | =ku        | kuku(ku)     | uko    | =ko        | kuko(ko)     | kulya  |
| 18    | muno<br>~mno | =mu        | mumu(mu)     | umo    | =mo        | mumo(mo)     | milya  |

The proximal and medial demonstratives have disyllabic basic forms, monosyllabic contracted forms, and trisyllabic reduplicated forms, whereas the distals only have basic forms. The proximal and medial contracted forms almost correspond to the first and second syllables, respectively, of the uncontracted basic forms. As they, along with formal reduction, have lost their independence as words, they are described as clitics (Furumoto 2024b). In the proximal reduplicated forms, the first syllable of the basic forms is repeated, whereas the medial reduplicated forms appear to have been created by replacing the first syllable of the medial basic forms with that of the proximals and duplicating the second syllable of the medial basic forms. The reduplicated forms can be disyllabic, as the parentheses suggest.

Additionally, Kimakunduchi has compound demonstrative, which can only be used for reference to entities in the speech situation. The proximal basic forms of classes 1–11 can be combined with +*va* (from class 16 *vano*) as in (4a) and +*ku* (from class 17 *kuno*) as in (4b).

The medial basic forms of these classes can be compounded with +*vo* (from class 16 *avo*) as in (4c) and +*ko* (from class 17 *uko*) as in (4d).

- (4) a. yuno+va  
DEM.PROX.1+DEM.PROX.16  
'This one (person) here'
- b. yuno+ku  
DEM.PROX.1+DEM.PROX.17  
'This one (person) here'
- c. uyo+vo  
DEM.MED.1+DEM.MED.16  
'That one (person) there'
- d. uyo+ko  
DEM.MED.1+DEM.MED.17  
'That one (person) there'

### [3] Genitive noun phrases

Table 4 shows genitive makers and genitive pronouns in Kimakunduchi. As can be seen in this table, their form also varies according to the noun class. For class 15, the class of the infinitive form of verbs, only the forms found in previous works (e.g., Racine-Issa 2002) and in the author's fieldwork are shown.

Table 4. Genitive makers and genitive pronouns in Kimakunduchi

| class | GEN           | 1SG               | 2SG             | 3SG             | 1PL             | 2PL             | 3PL           |
|-------|---------------|-------------------|-----------------|-----------------|-----------------|-----------------|---------------|
| 1     | y-a=<br>~w-a= | y-angu<br>~w-angu | y-ako~<br>w-ako | y-ake~<br>w-ake | y-etu~<br>w-etu | y-enu~<br>w-enu | y-ao~<br>w-ao |
| 2     | w-a=          | w-angu            | w-ako           | w-ake           | w-etu           | w-enu           | w-ao          |
| 3, 11 | w-a=          | w-angu            | w-ako           | w-ake           | w-etu           | w-enu           | w-ao          |
| 4     | y-a=          | y-angu            | y-ako           | y-ake           | y-etu           | y-enu           | y-ao          |
| 5     | ly-a=         | ly-angu           | ly-ako          | ly-ake          | ly-etu          | ly-enu          | ly-ao         |
| 6     | y-a=          | y-angu            | y-ako           | y-ake           | y-etu           | y-enu           | y-ao          |
| 7     | ch-a=         | ch-angu           | ch-ako          | ch-ake          | ch-etu          | ch-enu          | ch-ao         |
| 8     | vy-a=         | vya-ngu           | vy-ako          | vy-ake          | vy-etu          | vy-enu          | vy-ao         |
| 9     | y-a=          | y-angu            | y-ako           | y-ake           | y-etu           | y-enu           | y-ao          |
| 10    | z-a=          | z-angu            | z-ako           | z-ake           | z-etu           | z-enu           | z-ao          |
| 15    | kw-a=         |                   |                 | kw-ake          |                 |                 |               |
| 16    | v-a=          | v-angu            | v-ako           | v-ake           | v-etu           | v-enu           | v-ao          |
| 17    | kw-a=         | kw-angu           | kw-ako          | kw-ake          | kw-etu          | kw-enu          | kw-ao         |
| 18    | mw-a=         | mw-angu           | mw-ako          | mw-ake          | mw-etu          | mw-enu          | mw-ao         |



When the genitive marker is used to express a possessor-possessioned relation, it occurs just before the noun phrase expressing the possessor and follows the noun phrase referring to the possessioned item, as can be seen in (5).

- (5) bata    lya=sigombe  
       5.duck GEN.5=Sigombe (PN)  
       ‘Sigombe’s duck’

#### [4] Adjectives

Adjectives can also be decomposed into a prefix and a stem. The form of the prefix varies according to the noun class, and its variation is the same as that seen for the prefix of cardinal numerals; adjective prefixes are formally identical to noun prefixes, with the exception of class 11. The forms of the adjective meaning ‘small’ can be seen in Table 1.

In class 5, most adjectives do not have a part at the beginning that can be analysed as a prefix; however, adjectives whose stems begin with a vowel have the prefix *j-* (e.g., *j-ekundu* ‘red’). Classes 9/10 share the same form of adjectives. Among adjectives in these classes, some have a nasal sound in the initial position that can be analysed as a prefix (e.g., *n-dogo* ‘small’ and *m-baya* ‘bad’), others begin with an aspirated plosive (e.g., *k<sup>h</sup>ubwa* ‘big’ and *p<sup>h</sup>ya* ‘new’), and still others lack any explicit prefix (e.g., *zuri* ‘good’, *refu* ‘long’, and *fupi* ‘short’).

#### [5] Verb-based nominalisations

##### [5a] Relative clauses and the infinitive form of verbs

In Kimakunduchi, as in many other Bantu languages, verbal inflection is realised by prefixation. The structure of inflected verbs can be schematised as shown in (6).

- (6) Template of the inflected verb forms  
       Negative-Subject-Negative-Tense/Aspect/Mood (TAM)-Relative-Object-Stem

To form a relative clause, Kimakunduchi employs a verbal prefix which occurs just after the TAM prefix slot. This can be seen in (7), where the relative clause is formed through the presence of the relative clause marker (RM) *o-*. The form of the RM also varies according to the noun class, as summarised in Table 1. The RM ends with the vowel *o*, except for the class 1 form; hereafter, it is referred to as the *o*-marker.

- (7) wanak<sup>h</sup>ele    wa-na-o-tenda            usheza  
       2.children SM2-IPFV-RM2-do    play  
       ‘The children who are playing’ (Whiteley 1959: 64)

For the formation of first/second person singular and class 1 subject relative clauses, the prefix *m-*, which occurs in the same slot as the subject prefix and may be cognate with the class 1 noun prefix, can also be used (8a). In this context, it is also possible to use the *o*-maker occurring after the TAM slot (8b).

- (8) a. mie            m-na-tenda            kazi            ino  
 PRO.1SG RM.SM1-IPFV-do 9.work DEM.PROX.9  
 ‘I who am doing this work’ (Whiteley 1959: 64)
- b. mie            n-na-e-tenda            kazi            ino  
 PRO.1SG SM1SG-IPFV-RM1-do 9.work DEM.PROX.9  
 ‘I who am doing this work’ (Furumoto & Gibson 2022)

These relative clause prefixes can be combined with the copula verb *-wa* to form an independent RM. The TAM slot of the independent RM is typically occupied by a perfective marker.

- (9) a. duka            li-Ø-lyo-wa            ha-li-ja-fungwa  
 5.shop SM5-PFV-RM5-COP NEG-SM5-PRF-close.PASS  
 ‘The shop which has not (yet) been closed’ (Furumoto & Gibson 2022)
- b. mw-a-wa            ha-na-tenda            kazi  
 RM.SM1-PFV-COP NEG.SM1-IPFV-do work  
 ‘The one (person) who doesn’t work’ (Racine-Issa 2002: 155)

The infinitive form can also be considered a verb-based nominalisation. This form consists of the prefix *ku-* (*~kw-*) and a verb stem (e.g., *ku-soma* ‘to study’). The infinitive form has similar morphosyntactic properties to nouns; for example, it can function as the head of a noun phrase (10) and also take a locative suffix (11). In the numbering convention for the noun class system of Bantu languages, the infinitive form is assigned class number 15, as indicated in the gloss of the genitive pronoun in (10). Note that in (10), the class 15 genitive pronoun can be replaced with that of class 9.

- (10) kw-enenda kw-ake/y-ake  
 INF-go 15-his/9-his  
 ‘His way of moving on’ (Lit. ‘His going’)
- (11) tw-ende=ni            ku-ganga-ni  
 SM1PL-go.SUBJ=AL.PL INF-heal-LOC  
 ‘Let’s go together to the place of healing.’

### [5b] Complement clauses

In Kimakunduchi, complement clauses can be formed without any overt maker (12a). However, the marker *kuwa* can optionally appear in some contexts (12b). This complement marker can be decomposed in the infinitive prefix *ku-* and the copula verb stem *-wa*, and can also be used to form a so-called gapless relative clause (13).

- (12) a.  $\eta$ -Ø-sikii                      ka-cha-kuja                      kesho  
          SM1SG-PFV-hear   SM1-FUT-come   tomorrow  
          ‘I heard that he would come tomorrow.’
- b.  $\eta$ -Ø-sikii                      ku-wa                      ka-cha-kuja                      kesho  
          SM1SG-PFV-hear   INF-COP   SM1-FUT-come   tomorrow  
          ‘I heard that he would come tomorrow.’
- (13) hadithi    ku-wa                      sungura   ka-Ø-shindwa                      mashindano  
      9.story    INF-COP   1.hare   SM1-PFV-win.PASS   race  
      ‘A story in which a hare was lost in a race.’ (Furumoto 2019: 300)

### 3. DATA

Grammatical nominalisations in Kimakunduchi change their forms according to the noun class, as described in Section 2. Shibatani (2021: 524) has proposed a cross-linguistically universal pattern whereby the range of grammatical nominalisations to which classifying nominalisation marking (realised, for example, by classifiers and gender markers) is applied gradually expands (see also Shibatani 2023: 35). By using the hierarchy illustrating this proposed pattern, noun class marking in Kimakunduchi can be summarised as shown in Figure 1.

Figure 1. Noun class marking pattern in Kimakunduchi

NP use: NUM > DEM > GEN > ADJ > V-based NMLZ (RC)/(\*COMP)

Mod use: NUM > DEM > GEN > ADJ > V-based NMLZ (RC)/(\*COMP)

(\* = no marking or impossible; unmarked = obligatory marking)

As suggested by Figure 1, most grammatical nominalisations in Kimakunduchi show noun class marking in both NP use and Mod use. This is exemplified below.

#### [1] Numerals

##### [1a] Cardinal numerals

Cardinal numerals expressing ‘one’, ‘two’, ‘three’, ‘four’, ‘five’ and ‘eight’ change their forms according to the noun class. Such a numeral can be seen in (14) and (15), where the cardinal numeral *wa-nne* ‘four’ is accompanied by the class 2 prefix.

NP use

- (14) A: ku-na                      wana-we                      wa-ngavi  
          SM2SG-have   2.children-your   2-how.many  
          ‘How many (your) children do you have?’

B: wa-nne  
 2-four  
 ‘Four (children)’.

Mod use

- (15) wanawake wa-nne na wanaume wa-nne wa-ka-oana  
 2.women 2-four COM 2.men 2-four SM2-CONS-marry.REC  
 ‘Four women and four men were married to each other.’

Other cardinal numerals lack any overt noun class marker and appear in the same form in both NP use and Mod use, as the cardinal numeral *ishirini* ‘twenty’ demonstrates in (15) and (16).

NP use

- (16) uyo mw-a-ja-vyaa ishirini njo=ka-Ø-tende  
 DEM.MED.1 RM.SM1-PFV-come-bear twenty EMPH=SM1-PFV-do  
 ‘That one, the one who bore twenty (children), did it.’

Mod use

- (17) a-ka-vyaa wana ishirini  
 SM1-CONSC-bear 2.children twenty  
 ‘She bore twenty children.’

### [1b] Ordinal numerals

In (18), ordinal numerals accompanied by the class 1 genitive marker appear. Only the first example, *ya=pili*, is an instance of Mod use, while the other three are examples of NP use.

NP use and Mod use

- (18) mwanak<sup>he</sup>le ya=pili na=e a-k-enda ja viyvo  
 1.child GEN.1=second COM=PRO1 SM1-CONSC-go like DEM.MED.8  
 na ya=tatu a-k-enda ja viyvo  
 COM GEN.1=three SM1-CONSC-go like DEM.MED.8  
 na ya=ne a-k-enda ja viyvo  
 COM GEN.1=four SM1-CONSC-go like DEM.MED.8  
 ‘The secondborn son also went (to ask for permission to marry and failed just like the firstborn son did), and the thirdborn also did so, and the forthborn also did so.’  
 (Furumoto 2024a: 127)

### [2] Demonstratives

In (19) and (20), the class 1 medial demonstrative *uyo* indicates NP and Mod use of demonstratives, respectively. Demonstratives appear in the same form in both NP and Mod use, as can be seen in the following examples.

NP use

- (19) uyo                      a-k-enenda                      mpaka      viambo-ni  
 DEM.MED.1   SM1-CONSC-go   up.to      villages-LOC  
 ‘That one went to villages.’

Mod use

- (20) wa-na-m-chaka                      mwanak<sup>h</sup>ele      uyo  
 SM2-IPFV-OM1-want   1.child                      DEM.MED.1  
 ‘They want that child.’ (Furumoto 2024a: 118)

Demonstratives can occur before a noun or another grammatical nominalisation. I assume that in such constructions, a demonstrative and its following constituent are in an appositive relation; they refer to an entity independently in the same manner as when they form a noun phrase on their own, without any other constituent (see also Furumoto 2024a). Under this view, the following examples, where a demonstrative is followed by a noun or another grammatical nominalisation, are considered cases of demonstratives in NP use.

NP use

- (21) kwanini      uyo                      mwanak<sup>h</sup>ele      ka-na-lia  
 why                      DEM.MED.1   1.child                      SM1-IPFV-cry  
 ‘Why is that one, the child crying?’ (Racine-Issa 2002: 275) (DEM + noun)

- (22) ka-Ø-baki                      yuno                      m-moja  
 SM1-PFV-remain   DEM.PROX   1-one  
 ‘This guy, one person, stayed.’ (DEM + NUM)

- (23) lola=ga                      uyo+ko                                              mw-eupe  
 look.IMP=INTS   DEM.MED.1+DEM.MED.17   1-white  
 ‘Look at that guy over there, the white one’ (DEM + ADJ)

- (24) uyo                      mw-a-ja-vyaa                                              ishirini      njo=ka-Ø-tende  
 DEM.MED.1   RM.SM1-PFV-come-bear   twenty      EMPH=SM1-PFV-do  
 ‘That one, the one who bore twenty (children), did it.’ (DEM + RC)

### [3] Genitive noun phrases

In (25), NP and Mod uses of the class 9 genitive pronouns can be found. The genitive pronoun *yangu* ‘my’ follows the noun *simu* ‘phone’, while *yako* ‘your’ is used without it.

NP use and Mod use

- (25) simu      y-angu      i-Ø-bomoko                                              hea y-ako      i-na-tenda      kazi  
 9.phone   9-my      SM9-PFV-break.STAT                      but 9-your      SM9-IPFV-do      work  
 ‘My phone is broken but yours is working.’

NP use

- B: ya=haji  
GEN.1=Haji (PN)  
'Haji's.'

(27) uyo+ko                      ng'ombe    ya=shabani  
DEM.MED.1+DEM.MED.17   1.cow       GEN.1=Shabani (PN)  
'That is Shabani's cow.'

In (28), *khubwa* ‘big’ and *ndogo* ‘small’ show that adjectives have Mod use and NP use, respectively.

(28) ny-evu            ha-chaka  
SM1SG-PST   SM1SG.CONSC-want  
a-ka-ja            a-lete            embe            k<sup>h</sup>ubwa  
SM1-COND-come SM1-bring.SBJV   9.mango    big.9  
hea ka-Ø-lete            n-dogo  
but SM1-PFV-bring   9-small  
'I wanted him to bring a big mango when he came, but he brought a small one.'

287

- (29) a. gumua tumbo ly-angu  
touch.IMP 5.belly 5-my  
'Touch my belly.'
- b. \*gumua tumbo ly-angu kubwa  
touch.IMP 5.belly 5-my big.5  
Intended: 'Touch my big belly.'
- (30) chukua peni y-angu n-dogo  
take.IMP 9.pen 9-my 9-small  
'Take my small pen.'

#### [5] Verb-based nominalisations (Relative clauses)

Examples of relative clauses characterised by the *o-* marker are presented in (31), while examples including relative clauses formed with the *m-* marker can be seen in (32). Both types of relative clauses have NP use and Mod use. For complement clauses formed with *kuwa*, see (12) and (13).

NP use and Mod use

- (31) a. tu-na-kulya a-Ø-zo-leta pandu  
SM1PL-IPFV-eat SM1-PFV-RM10-bring 1.Pandu (PN)  
'We are eating ones (mangoes) that Pandu has brought.' (Furumoto 2016: 170)
- b. tu-na-kulya embe a-Ø-zo-leta pandu  
SM1PL-IPFV-eat 10.mangoes SM1-PFV-RM10-bring 1.Pandu (PN)  
'We are eating mangoes that Pandu has brought.' (Furumoto 2016: 169)

NP use and Mod use

- (32) a. m-na-kwimba nyimbo nyi-Ø-m-kut<sup>h</sup>u  
RM.SM1-IPFV-sing songs SM1-PFV-OM1-meet  
'For someone who was singing, I met.' (Furumoto 2016: 170)
- b. m<sup>th</sup>u m-na-kwimba nyimbo nyi-Ø-m-kut<sup>h</sup>u  
1.person RM.SM1-IPFV-sing songs SM1-PFV-OM1-meet  
'For someone who was singing, I met.' (Furumoto 2016: 170)

## 4. ANALYSIS AND DISCUSSION

In Kimakunduchi, noun class marking appears on most grammatical nominalisations, with some exceptions (e.g., most cardinal numerals, complement clauses). This observation indicates that Kimakunduchi almost fits into Shibatani's (2021: 524, 2023: 35) hierarchy as an advanced and mature pattern language.

Assuming a gradual development of overt forms related to noun classification (i.e., classifiers, gender/noun class markers) (see also Shibatani 2023, cf. Greenberg 1978, Van de

Velde 2019), Kimakunduchi might be located at a later stage of the process, where overt markers have started to erode and disappear. I consider that at this stage, the noun class system is still retained, and nouns and grammatical nominalisations can still change their forms according to the noun class, but there is no longer a strict one-to-one correspondence between the noun class marking (i.e., form) and the noun class system (i.e., classification). In this view, the absence of overt noun class makers on Kimakunduchi numerals except for those expressing ‘one’, ‘two’, ‘three’, ‘four’, ‘five’ and ‘eight’, and the presence of syncretism for noun prefixes and grammatical nominalisations are interpreted as indicative of beginning of the collapse of the correlation between the form and classification.

In Kimakunduchi, grammatical nominalisations (except for V-based NMLZ) that can be analysed as having a noun class marker, most of which are prefixes, never appear without such a part. This observation allows us to hypothesise that these grammatical nominalisations are stored in the lexicon with their noun class markers. While Shibatani (2021) argues that classifiers and gender/noun class markers function as nominalisers, in Kimakunduchi, nominalisation through the addition of these markers may not be synchronically active. When erosion of noun class markers has occurred, no overt nominaliser is required for nominalisation. This view is compatible with the fact that in Kimakunduchi there are grammatical nominalisations that lack an overt noun class marker. More generally, Kimakunduchi suggests that the elaboration of synchronic morphological features, with a focus on the relation between form and meaning/function, may provide new insights into diachronic development (and decline) of markers related to noun classification.

Furthermore, the observations in Kimakunduchi regarding the syntactic relation between a noun and a co-occurring grammatical nominalisation, as well as that between multiple co-occurring grammatical nominalisations, should also be noted. As described in Section 3, in Kimakunduchi, prenominal demonstratives are assumed to be in an appositive relation with the following noun or grammatical nominalisation, whereas adjectives, which appear only post-nominally, always function restrictively. The difference between an appositive noun phrase and a noun phrase consisting of a noun and a grammatical nominalisation in modification use on is not necessarily analysed by Shibatani’s nominalisation theory (Shibatani 2019, see also Shibatani 2021, 2023), which only stipulates that each noun and grammatical nominalisation denotes a set of concepts that are metonymically evoked, with its precise reference determined by the discourse context in accordance with Grice’s Cooperative Principle. If nouns and grammatical nominalisations function in this way, they should be able to either restrict the denotation of the co-occurring constituent (as restrictive relatives do) or form an appositive construction with it (as non-restrictive relatives do), regardless of their order and position. However, this is not actually the case in Kimakunduchi. The observations on Kimakunduchi suggest that when two or more grammatical nominalisations co-occur, there may be language-specific rules governing how they interact with one another and how their relationship may change depending on their order.

## 5. CONCLUSION

This article has described the noun classification system of Kimakunduchi and outlined its grammatical nominalisations. In Kimakunduchi, almost all grammatical nominalisations



(except most cardinal numerals and complement clauses) appear with a noun class marker in both NP use and Mod use. These features can be interpreted as suggesting that Kimakunduchi is at a later stage in the gradual development of gender/noun class marking, which still awaits further elaboration in theoretical considerations. Furthermore, observations on demonstratives and adjectives in Kimakunduchi have raised the need for a detailed analysis of both language-specific constraints and cross-linguistic variation regarding the semantic correlation of two or more co-occurring nouns or grammatical nominalisations.

## ABBREVIATIONS

|      |                            |
|------|----------------------------|
| ADJ  | adjective                  |
| AL   | allocutive                 |
| COM  | comitative                 |
| COMP | complement clause          |
| COND | conditional                |
| CONS | consecutive                |
| COP  | copula                     |
| DEM  | demonstrative              |
| EMPH | emphatic                   |
| FUT  | future                     |
| GEN  | genitive                   |
| IMP  | imperative                 |
| INF  | infinitive                 |
| INTS | intensifier                |
| IPFV | imperfective               |
| LOC  | locative                   |
| MED  | medial                     |
| NEG  | negative                   |
| NMLZ | nominalizer/nominalization |
| NUM  | numeral                    |
| OM   | object marker              |
| PASS | passive                    |
| PFV  | perfective                 |
| PL   | plural                     |
| PN   | proper noun                |
| PRF  | perfect                    |
| PRO  | pronoun                    |
| PROX | proximal                   |
| RC   | relative clause            |
| REC  | reciprocal                 |
| RM   | relative clause marker     |
| SBJV | subjunctive                |
| SG   | singular                   |
| SM   | subject marker             |

|      |         |
|------|---------|
| STAT | stative |
| V    | verb    |

## ACKNOWLEDGMENTS

This study was made possible through the generous support of the Grant-in-Aid for JSPS Fellowship, JSPS KAKENHI (13J03150, 16J03295, 22H00659/23K21931, 23H00622/23K25319, 24K16047), the NINJAL project “Empirical Study of the Typology of Nominalization—from Theoretical, Fieldwork, Historical and Dialectal Perspective” (PI: Kazuhiro Kawachi), and the ILCAA joint research project “Diachronic Perspectives on Language Description and Typology in Bantu (jrp000292)” (PI: Makoto Furumoto), for which I am deeply appreciative. I would also like to extend my thanks to the three Kimakunduchi native speakers for providing natural language data for this study.

## REFERENCES

- Dixon, R. M. W. 2010. *Basic linguistic theory, Vol. 1*. New York: Oxford University Press.
- Furumoto, Makoto. 2015. 奇妙な複数形と文法的性としての名詞クラスについて：スワヒリ語の二つの地域変種における事例 (Non-standard plural forms and noun class as grammatical gender in Swahili: The case of Kiunguja and Kikae). *Kyoto University linguistic research* 34, 25–40.
- Furumoto, Makoto. 2016. スワヒリ語カエ方言の「関係節」－準体言としての記述－ (‘Relative clauses’ in the Kikae dialect of Swahili: A description as grammatical nominalisation). *Journal of Kijutsuken (Descriptive Linguistics Study Group)* 8, 147–172.
- Furumoto, Makoto. 2019. Makunduchi (G43c). In Daisuke Shinagawa and Yuko Abe (eds.) *Descriptive materials of morphosyntactic microvariation in Bantu, Vol. 1*, 288–307. Tokyo: Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.
- Furumoto, Makoto. 2024a. Reorganisation of the demonstrative system in Kimakunduchi. In Steve Nicolle (ed.) *Discourse functions of demonstratives: Contributions from the field*, 115–129. Dallas, TX: SIL International.
- Furumoto, Makoto. 2024b. The grammaticalisation of Kimakunduchi demonstratives: Insights into the emergence of post-stem object markers in Bantu. In James Essegbey, Brent Henderson, Fiona McLaughlin and Michael Diercks (eds.) *Pushing the boundaries: Selected papers from the 51–52 Annual Conference on African Linguistics (Contemporary African Linguistics 10)*, 205–226. Berlin: Language Science Press.
- Furumoto, Makoto, and Hannah Gibson. 2022. Variation in Kimakunduchi and Standard Swahili: Insights from verbal morphosyntax. *Linguistique et langues africaines* 8 (1). <https://doi.org/10.4000/lla.1994>
- Greenberg, Joseph H. 1978. How does a language acquire gender markers? In Joseph H. Greenberg, Charles A. Ferguson and Edith A. Morcsik (eds.) *Universals of human language Vol. 3: Word structure*, 47–82. Stanford: Stanford University Press.

- Heine, Bernd. 1982. African noun class systems. In Hansjakob Seiler and Christian Lehmann (eds.) *Apprehension: Das sprachliche erfassen von gegenständen, I: Bereich und ordnung der phänomene*, 189–216. Tübingen: Gunter Narr.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.
- Katamba, Francis. 2003. Bantu nominal morphology. In Derek Nurse and Gérard Philippson (eds.) *The Bantu languages*, 103–120. London: Routledge.
- Maho, Jouni. 1999. *A Comparative study of Bantu noun classes*. Gothenburg: Acta Universitatis Gothoburgensis.
- Nurse, Derek. 1982. A tentative classification of the primary dialects of Swahili. *Sprache und Geschichte in Afrika* 4, 165–206.
- Nurse, Derek, and Thomas J. Hinnebusch. 1993. *Swahili and Sabaki: A linguistic history*. Berkeley: University of California Press.
- Racine-Issa, Odile. 2002. *Description du kikae: Parler swahili du sud de Zanzibar: Suivie de cinq contes*. Leuven: Peeters.
- Shibatani, Masayoshi. 2019. What is nominalization? Towards the theoretical foundations of nominalization. In Roberto Zariquiey, Masayoshi Shibatani and David W. Fleck (eds.) *Nominalization in languages of the Americas*, 15–167. Amsterdam: John Benjamins.
- Shibatani, Masayoshi. 2021. 連体修飾の文法：類別詞と文法性を中心に (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Nominalization theory and linguistic analysis*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical linguistics in Japan* 12, 1–52.
- Van de Velde, Mark. 2019. Nominal morphology and syntax. In Mark Van de Velde, Koen Bostoen, Derek Nurse and Gérard Philippson (eds.) *The Bantu languages, Second edition*, 237–269. London: Routledge.
- Whiteley, Wilfred H. 1959. An introduction to the rural dialects of Zanzibar, part 1. *Swahili* 30, 41–69.

# Gender marking on grammatical nominalizations in Sidaama

Kazuhiro Kawachi<sup>1</sup>

## ABSTRACT

This study examines gender marking on grammatical nominalizations that function as the heads of noun phrases and as nominal modifiers in Sidaama, a Highland East Cushitic language of Ethiopia. Sidaama has two types of classifying grammatical nominalization markers that distinguish feminine and masculine gender. One type is an enclitic used to form a singular noun phrase, while the other consists of the nominative, genitive, and dative-locative case suffixes. The enclitic occurs with ordinal numerals, genitive noun phrases, relative clauses, and complement clauses, whereas the case suffixes attach to cardinal numerals and adjectives. Both types of markers are analyzed as being included in demonstratives. The present study explores how Shibatani's (2021, 2023) implicational hierarchy accounts for a language with two types of gender nominalization markers. The Sidaama data support his hypothesis concerning the functional dimension, according to which classifying nominalizers are more likely to occur obligatorily with a given nominalization in noun-phrase use than in modification use. However, they diverge from his prediction along the structural dimension, which posits that classifying nominalizers are more likely to occur obligatorily in the following order: numerals, demonstratives, genitive noun phrases, adjectives, and verb-based nominalizations. This paper also discusses why the Sidaama data deviate from Shibatani's prediction along the structural dimension.

## 1. INTRODUCTION

The present study examines how gender is marked on the following types of structures to derive grammatical nominalizations in Sidaama, a Highland East Cushitic language of Ethiopia (ISO 639-3: sid). (Other types of structures for example, quantifiers and interrogatives, are not discussed here.)

- [1] Numerals ([1a] Cardinal numerals, [1b] Ordinal numerals)
- [2] Demonstratives (demonstrative pronouns/adnominal demonstratives)
- [3] Genitive noun phrases
- [4] Adjectives
- [5] Verb-based nominalizations ([5a] Relative clauses, [5b] Complement clauses)

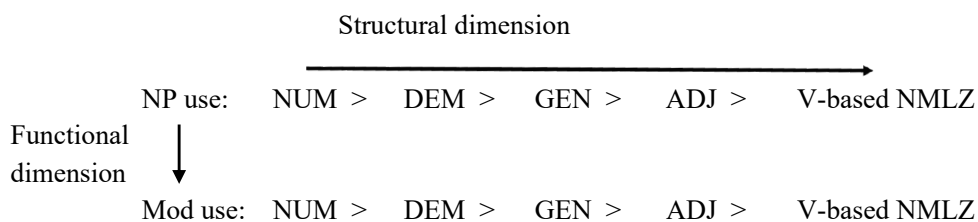
This language has two types of classifying grammatical nominalization markers: an enclitic used to form a singular noun phrase, and the nominative, the genitive, and the dative-locative case suffixes. The present study examines whether the gender marking pattern of this language conforms to Shibatani's (2021, 2023) hypothesis. Using the implicational hierarchy (Figure 1),

---

<sup>1</sup> Faculty of Business and Commerce, Keio University, 4-1-1 Hiyoshi, Koohoku-ku, Yokohama, Kanagawa, Japan 223-8521. email: kawachi@keio.jp

he hypothesizes that classifying grammatical nominalization markers such as classifiers and gender markers exhibit more marking patterns for the derivation of grammatical nominalizations higher in the hierarchy than those lower down and more for their noun-phrase use (NP use) than for their modification use. He hypothesizes that the use of classifying grammatical nominalization markers develops from numerals downward to other types of structures (along the structural dimension) and from their NP use to their modification use (along the functional dimension).

Figure 1. Hierarchy constraining classifier/gender marking (Shibatani 2023: 35)



The question that the present study addresses is how this hypothesis holds for languages with more than one type of gender nominalization marker.

Of the two types of gender nominalization markers, the noun-phrase enclitic (NPC) attaches to numerals, genitive noun phrases, relative clauses, and complement clauses, and is incorporated into demonstratives (Kawachi 2011). The other type of gender nominalization marker, the nominative, the genitive, and the dative-locative case suffixes, occurs with numerals and adjectives. The Sidaama data support Shibatani's hypothesis in that only numerals can be gender-marked with either type of marker, and gender marking appears to have originated in NP use before extending to modification use. However, unlike other structures—and like nouns—adjectives can take the case suffixes but not the noun-phrase enclitic, similar to nouns. These findings raise further issues for the analysis of gender marking in a language such as Sidaama.

### Basic information on Sidaama (Kawachi 2020a, 2023)

Sidaama exhibits a nominative-accusative case-marking system and shows both head-marking and dependent-marking properties. This language is predominantly head-final, with SOV basic word order. When a noun is modified, the noun phrase typically follows the structure {DEM–CARD.NUM–ADJ–REL.CL–GEN–N} or {DEM–ADJ–ORD.NUM–REL.CL–GEN–N}. However, when a relative clause and a genitive noun phrase cooccur within the same noun phrase, the noun-phrase enclitic =*ta* (SG.F)/=*ha* (SG.M)/=*re* (PL) must accompany the relative clause.<sup>2</sup>

Nouns and adjectives share a few properties. First, both can have up to three types of number forms, morphologically unmarked, singular-suffixed, and plural-suffixed forms.

<sup>2</sup> Nevertheless, the noun can be followed by cardinal numerals (optionally with the emphatic NPC =*n=ta* (F)/=*n=ka* (M)), demonstratives, and adjectives, and by genitive noun phrases or relative clauses with the NPC. However, ordinal numerals cannot follow the noun.

Second, both can form a noun phrase by themselves. In the case of unmarked forms and singular-suffixed forms, the noun-adjective distinction is most clear-cut for words for inanimates, whose gender is lexically based if they are nouns and depends on the gender of the referent or the lexical item if they are adjectives, is less so for those for animals, whose gender may be lexical or biological, and is blurred for those for humans, whose gender is biologically based. (The genders of plural-suffixed forms for inanimates and many animates are feminine, but those for some animates especially humans depend on the plural suffix, which distinguishes the gender of the referents according to the gender of the allomorph of the case suffix and that of the verb suffix(es).) However, there is one criterion for drawing a line between the noun-adjective continuum: unlike an adjective, whose any grammatical case form can modify a noun, a noun cannot modify another noun except that it can as a possessor when it is in the genitive case. (Nevertheless, there are a small number of words whose unmarked and singular-suffixed forms cannot modify nouns but whose plural forms can.)

Lexical nominalizations are formed through suffixation. Suffixes that nominalize verbs, those that nominalize adjectives, and those that turn nouns into abstract nouns are shown in Table 1, Table 2, and Table 3, respectively, with examples.<sup>3</sup>

Table 1. Suffixes that nominalize verbs

| Suffix       | Verb                  | Noun                                        |
|--------------|-----------------------|---------------------------------------------|
| -a           | giir- ‘burn’          | giir-a ‘flame, burning’                     |
| -o           | hank’- ‘become angry’ | hank’-o ‘anger’                             |
| -e           | fool- ‘breathe’       | fool-e ‘breath’                             |
| -anšo        | t’on- ‘insult’        | t’on-anšo ‘insult’                          |
| -atto        | huučč’- ‘pray’        | huučč’-atto ‘prayer’                        |
| -ano         | mug- ‘become sleepy’  | mug-ano ‘sleepiness’                        |
| -ille/-iille | iibb- ‘become hot’    | iibb-ille ‘heat, fever’                     |
| -imma        | geedǵ- ‘become old’   | geedǵ-imma ‘old age’                        |
| -aančo       | ofoll- ‘sit’          | ofoll-aančo ‘seat, chair, someone who sits’ |
| -aasinčo     | sut- ‘hang’           | sut-aasinčo ‘hook, swing’                   |
| -amaančo     | il- ‘give birth’      | il-amaančo ‘new-born baby’                  |

Table 2. Suffixes that nominalize adjectives

| Suffix | Adjective             | Noun                                    |
|--------|-----------------------|-----------------------------------------|
| -imma  | gowwa ‘foolish’       | goww-imma ‘foolishness’                 |
| -ille  | guraččo ‘left-handed’ | guračč-ille ‘left-handedness’           |
| -a     | miniweelo ‘homeless’  | miniweel-a ‘(instance of) homelessness’ |

<sup>3</sup> The suffix *-a* can attach to almost all verbs to form verbal nouns, which express an abstract action or state change. On the other hand, all the other suffixes that nominalize verbs are each restricted to certain verbs and express more concrete concepts. In this respect, these suffixes could be analyzed as classifying nominalization markers.

Table 3. Suffixes that nominalize nouns to derive abstract nouns

| Suffix | Noun            | Noun                                   |
|--------|-----------------|----------------------------------------|
| -imma  | k'aakk'o 'baby' | k'aakk'-imma 'babyhood, childlikeness' |
| -oo'ma | rodoo 'sibling' | rod-oo'ma 'establishing siblinghood'   |

Nouns derived with these suffixes, which are apparently non-classifying lexical nominalizers, do not appear to form any semantic class. Nevertheless, nouns derived with *-e*, *-anšo*, *-atto*, or *-imma* are all feminine and those derived with *-ille/-iille* or *-oo'ma* are all masculine, though those derived with all the other suffixes can be feminine or masculine. Moreover, the suffixes *-aančo*, *-aasinčo*, and *-amaančo* in Table 1 each contain one of the allomorphs of the singular noun suffix *-čo* and express the singularity of a concrete denotation, unlike the other nominalizing suffixes. (The suffix *-aasinčo* is the causative form of *-aančo* and *-amaančo* is its passive form.)

Sidaama has three non-classifying grammatical nominalization markers, two of which form verb-based nominalizations and the other of which can produce verb-based nominalizations.

First, the infinitive suffix *-a* nominalizes a clause. When the formed noun phrase is in the nominative or the genitive case, it is treated as feminine, as in (1) and (2).<sup>4</sup>

- (1) tumá                      mur-a  
 onion.ACCOBL cut-INF.NOM.F  
 hindiidd-i-s-s-u-nke.  
 shed.tears-EP-CAUS-3SG.F-R.PRF.3SG.F-1PL  
 'Cutting an onion made us shed tears.'
- (2) há'r-a-nke                      di = danča = te.  
 leave.and.go-INF.NOM.F-1PL.POSS      NEG=good=NPC.F.PRED  
 'It is not good that we left/are leaving.' (*lit.*, 'Our leaving is not good.')

Second, one of the complementizers in this language =*gede* (e.g., (3)) nominalizes a complement clause (Kawachi 2020b: 696, 710). This clitic, which is invariant in form, does not make a gender distinction.

<sup>4</sup> The subject of the nominalized verb appears in the nominative case when the verb is transitive (e.g., *áti tumá mur-a* 'Your (SG) cutting an onion' in (1)), and it is either in the nominative case or the genitive case when the verb is intransitive (e.g., *áti/até há'r-a* (2SG.NOM/2SG.GEN leave.and.go-INF.NOM.F) 'your (SG) going' in (2)).

- (3) hakkunni            gedensaanni,        šool-u = n = ku  
 that.M.GEN        after                    four-NOM.M=EMPH=NPC.NOM.M  
 mitte-e-nni                    sikk'-uwá                    mur-t-e  
 one.F-LV-ABLINS        stick-PL.ACCOBL        cut-3PL-CVB  
 abb-i-t-anno = gede                    haĵaj-ø-i-nsa.  
 bring-EP-3PL-IPFV.3=COMP        order-3SG.M-R.PRF.3SG.M-3PL  
 'After that, he ordered all the four to cut the sticks and bring them at the same time.'  
 (adapted from Kaarricha et al. 2021: 36)

Third, the locational enclitic = *wa* 'at/to the (...) location (of/where/that) ...' (nominative and genitive: = *wi*) is used to form a noun phrase referring to a location or a goal of motion.<sup>5</sup> This enclitic attaches to a numeral (either cardinal or ordinal), an adjective (e.g., (4)), a genitive noun phrase (e.g., (5)), or a relative clause to form a noun phrase without a noun (not to function as a noun modifier).

- (4) ...    t'a            k'arr-am-i-tt-o-kki-nni  
           now            cause.problem-PASS-R.PRF.2-2SG-M-NEG-ABLINS  
 múlla = wa        ka'-ø-e                    boo-'-n-oo-tt-o = na,                    ...  
 empty=place        start-3SG.M-CVB        brag-MID-brag-D.PRF.2-2SG-M=so  
 '... now without getting into a problem, in the place where there is no problem (*lit.*, the empty place), you (SG.M) started to talk and bragged yourself, so ...' (Story: 'K'arrantukinni lik'iidde kaffaltukki-nni reitino')
- (5) ...    mánčo            weesé            hok-k'-i-t-á-nni  
           person.NOM.F    enset.ACCOBL    process.enset-MID-EP-3SG.F-INF-ABLINS  
 no-o = té = wa                    mar-t-ino.  
 come.to.exist.D.PRF.3-LV=NPC.F.GEN=place        go-3SG.F-D.PRF.3  
 '... she (the bird) went to a woman's place where she (the woman) was processing enset.'  
 (Story: 'Waanje dagge')

## Data sources

The present author examined data collected during twenty-five field trips to the Sidaama National Regional State, Ethiopia, each lasting two-and-a-half to five weeks, conducted from 2008 to the present. The data presented in this paper comes from conversations, folk tales, true stories, published written materials, and forms elicited from the author's consultants, as well as speakers' descriptions of video clips, specifically those for the Causality across Languages project designed by Jürgen Bohnemeyer and Erika Bellingham (Kawachi et al. 2018, 2020, in press, Bellingham et al. 2000), those for the NINJAL MEDAL motion project (Matsumoto

<sup>5</sup> The clitic = *wa* could be analyzed as a classifying grammatical nominalization marker in that it attaches to elements that do not denote a location. For nominals that denote a location, the dative-locative suffix is used to express a location or a goal of motion (e.g., *ros-ú min-i-ra* (education-GEN.M house-GEN.M.MOD-DATLOC.MOD) 'at/to school'). Therefore, this language differentiates two types of nominals.



2025) designed by Yo Matsumoto (Kawachi 2025), and memory-based descriptions of Chafe's (1980) Pear Film.

## 2. OVERVIEW OF GENDER MARKING AND GRAMMATICAL NOMINALIZATIONS IN SIDAAMA

### 2.1 Gender marking in Sidaama

Sidaama makes a gender distinction between feminine and masculine.<sup>6</sup> Most nouns are assigned lexical gender. The biological gender distinction appears overtly on gender suffixes on some nouns and adjectives. There are various grammatical categories that make a gender distinction: the numeral 'one', pronouns, verb suffixes, the noun-phrase enclitic, and the case suffixes. The most relevant to the present study are the noun-phrase enclitic and the case suffixes, which are classifying grammatical nominalization markers.

#### Gender suffixes on singular nouns and adjectives

The singular-suffixed forms of some nouns and adjectives have two different gender forms, feminine and masculine forms, as shown in Table 4.<sup>7</sup> The distinction between these forms is based on the biological gender in the case of a human referent and on the lexical gender of the noun that would refer to an inanimate referent. (Many of such adjectives additionally have singular-suffixed forms with no gender distinction, as in Table 4.) Those words that can modify nouns are analyzed (e.g., the last four items in Table 4) as adjectives, though their genitive case forms (as a possessor noun phrase) can also modify nouns. (In the case of nouns (e.g., the first three items in Table 4), only their genitive case forms can modify other nouns.)

Table 4. Gender suffixes on singular nouns and adjectives

| FEM            | MASC           | gender-unspecified                      | gloss                                                        |
|----------------|----------------|-----------------------------------------|--------------------------------------------------------------|
| jaal-itte      | jaal-ičča      | jaal-iččo                               | 'friend'                                                     |
| fiit'-itte     | fiit'-ičča     | fiit'-iččo                              | 'relative'                                                   |
| murr-itte      | murr-ičča      | —                                       | 'leader (of a local organization)'                           |
| hojjaam-eette  | hojjaam-eessa  | hojjaam-eessiččo                        | 'tall'                                                       |
| dur-eette      | dur-eessa      | dur-eessiččo                            | 'rich'                                                       |
| bull-itte      | bull-ičča      | bull-iččo                               | 'gray, brown'                                                |
| airr-aditte    | airr-adičča    | airr-adiččo (for a human referent only) | 'heavy (for inanimates and animals), respected (for humans)' |
| hank'-aaleette | hank'-aaleessa | hank'-aaleessiččo                       | 'becoming angry easily'                                      |

<sup>6</sup> When the denotation or the referent is unidentifiable, masculine gender is used as a default gender.

<sup>7</sup> Some of such nouns and adjectives have morphologically number-unmarked forms (e.g., *jaala* 'friend', *fiit'a* 'relative') or verbs from which they derived (e.g., *airr-* 'become heavy, become respected', *hank'-* 'become angry'). Therefore, the gender suffixes and the gender-unspecified singular suffix could be analyzed as classifying grammatical nominalization markers.

## Numeral for ‘one’

The cardinal numeral for ‘one’ makes the gender distinction with the two gender forms, *mitte* (F) and *mitto* (M), unlike other cardinal numerals.

## Pronouns

Pronouns also make the gender distinction only for the third-person singular, as shown in Table 5.

Table 5. Pronouns

| PERS/NUM(/GENDER) |   | NOM   | ACC   | GEN   | DAT      |
|-------------------|---|-------|-------|-------|----------|
| 1SG               |   | ani   | ané   | ané   | ané-ra   |
| 2SG               |   | ati   | até   | até   | até-ra   |
| 3SG               | F | ise   | isé   | isé   | isé-ra   |
|                   | M | isi   | isó   | isí   | isí-ra   |
| 1PL               |   | ninke | ninké | ninké | ninké-ra |
| 2PL               |   | ki’ne | ki’né | ki’né | ki’né-ra |
| 3PL               |   | insa  | insá  | insá  | insá-ra  |

## Verb suffixes

The gender distinction is made for the first-person subject with the gender suffix on a finite verb and for the third-person singular subject with allomorphs of the recent perfect suffix and with those of the person/number suffix, as in Table 6.<sup>8</sup>

Table 6. Aspect suffixes and subject person/number suffixes, and subject gender suffixes on the verb

| PERS/<br>NUM<br>(/GEND) | V-suffixes                | ASPECT |       |                  | PERS/<br>NUM | GEND |    |
|-------------------------|---------------------------|--------|-------|------------------|--------------|------|----|
|                         |                           | IMPRF  | R.PRF | D.PRF            |              | F    | M  |
| 1SG                     | V-ASP-PERS/NUM-GEND       | -ee    | -u    | -oo              | -mm          | -a   | -o |
| 2SG                     | V-ASP-PERS/NUM-GEND       | -a     | -i    | -oo              | -tt          | -a   | -o |
| 3SG.M                   | V-PERS/NUM/GEND-ASP       | -anno  | -i    | -ino             | -ø           | —    |    |
| 3SG.F/<br>3PL           | V-PERS/NUM(/GEND)-<br>ASP | -anno  | -u    | -ino             | -t           | —    |    |
| 1PL                     | V-PERS/NUM-AS-PERS        | -ee    | -u    | -oo              | -n -mmo      | —    |    |
| 2PL                     | V-PERS/NUM-ASP            | -anni  | -i    | -oonni<br>(-ooi) | -tin         | —    |    |

<sup>8</sup> The suffixes for the third-person singular feminine and the third-person plural are identical. In the glosses of examples, the present study uses 3SG.F/3PL when the subject is in the plural and 3SG.F when the subject is in the singular.

The third-person subject suffix on non-finite verbs (with the infinitive suffix, the connective-converb suffix, or the instrumental-converb suffix) makes the gender distinction only for a third person (3SG.F/PL: *-t*, 3SG.M: *-ø*).

The pronominal object suffix makes the gender distinction only for the third person, as in Table 7.

Table 7. Pronominal object person/number(/gender) suffix

| PERS/NUM(/GEND) |   | Pronominal object suffix |
|-----------------|---|--------------------------|
| 1SG             |   | -’e                      |
| 2SG             |   | -he                      |
| 3SG             | F | -se                      |
|                 | M | -si                      |
| 1PL             |   | -nke                     |
| 2PL             |   | -’ne                     |
| 3PL             |   | -nsa                     |

### Noun-phrase enclitic

Sidaama has two types of noun-phrase enclitics (NPCs), (i) the (non-locational or general) NPC =*ta* (SG.F)/=*ha* (SG.M)/=*re* (PL) (in the accusative-oblique case), whose singular forms appear to originate from the proto-Afro-Asiatic demonstratives containing *t* (FEM)/*k* (MASC) (Kawachi 2011, 2020: 683–684), and (ii) the locational NPC =*wa* ‘at/to the (...) location (of/where/that) ...’ (Section 1). Neither type of NPC attaches to nouns unless they are in the genitive case. This paper focuses on the former type, whose main function is the formation of a noun phrase with certain kinds of structures through nominalization without using a noun.

The NPC takes different forms depending on the grammatical case of a noun phrase constituted with it (Table 8).<sup>9</sup> The dative-locative and the ablative-instrumental case forms of the NPC are formed with its genitive form with the dative-locative and the ablative-instrumental case suffixes, respectively. The accusative-oblique case forms of the NPC, which show no syncretism with any other case forms, are used as their representative forms, though it is not clear which case form is the most basic. The vowels in the NPC forms right after *t* (F)/*h* (M) can be analyzed as case suffixes that also make the gender distinction: *i* vs. *u* for the nominative, *e* vs. *u* for the genitive, and *e* vs. *o* for the predicative. (The feminine NPC in the genitive case =*te* is identical in form with the genitive and the dative-locative case suffixes for Unmodified feminine noun phrases *-te*.)

<sup>9</sup> The NPC also occurs in the cleft and the pseudo-cleft constructions (Kawachi 2020: 726–731).

Table 8. Different case forms of the NPC *=ta* (F)/ *=ha* (M)/ *=re* (PL)

|                     | ACCOBL      | NOM         | GEN         | DATLOC         | ABLINS          | PRED        |
|---------------------|-------------|-------------|-------------|----------------|-----------------|-------------|
| FEM                 | = <i>ta</i> | = <i>ti</i> | = <i>te</i> | = <i>te-ra</i> | = <i>te-nni</i> | = <i>te</i> |
| MASC                | = <i>ha</i> | = <i>hu</i> | = <i>hu</i> | = <i>hu-ra</i> | = <i>hu-nni</i> | = <i>ho</i> |
| “Mod”.N/PROP.N/PRON |             |             |             |                |                 | = <i>ti</i> |
| PL                  | = <i>re</i> | = <i>ri</i> | = <i>ri</i> | = <i>ri-ra</i> | = <i>ri-nni</i> | —           |

The plural NPC *=re*, which is used for plural referents, forms an argument, but not a predicate.<sup>10</sup>

The NPC *=ta* (F)/ *=ha* (M)/ *=re* (PL) has both NP and modification uses. In either use, the formed noun phrase is usually definite, though it is not necessarily definite when it is a predicate. Nevertheless, there are cases where the NPC appears out of context, even at the beginning of discourse. In such cases, the noun phrase is interpreted by default as having a generic referent, which the speaker and the hearer have shared knowledge about: a noun phrase formed with *=ta* (F) refers to ‘woman’ or ‘thing (propensity)’, whereas one formed with *=ha* (M) refers to ‘person’, ‘thing (object)’, or ‘thing (event)’.

There are differences between the singular NPC *=ta* (F)/ *=ha* (M) and the plural NPC *=re* (PL). The former seems to be more grammaticalized than the latter.<sup>11</sup> The singular NPC can occur to form a predicate noun phrase (see Appendix 2 for examples), whereas the plural NPC cannot. The singular NPC makes the gender distinction, but the plural NPC does not. The present study focuses on the singular NPC as a gender-classifying nominalization marker, though the singular NPC and the plural NPC as a whole could be analyzed as a pair of number-classifying nominalization markers.

### Case suffixes on nouns and on grammatical nominalizations

The case-marking system of Sidaama is nominative-accusative. The nominative, the genitive, and the dative-locative cases make the gender distinction with different allomorphs, as shown in Table 9. Another distinction between the masculine allomorphs of the nominative, between the allomorphs of the genitive suffixes, and between the allomorphs of the dative-locative suffix is based on the “Modification” distinction (Kawachi and Tekleselassie 2012) (capitalized as MOD and UNMOD for ‘Modified’ and ‘Unmodified’), which concerns whether or not a common noun is accompanied by a grammatical nominalization in its modification use (i.e., a noun modifier) and/or the possessive pronominal suffix.

<sup>10</sup> When a noun phrase constituted with *=re* is a predicate, it is followed by the predicative NPC *=ti* (e.g., *ané=re-e=ti* (1SG.GEN=NPC.PL.ACC-LV=NPC.PRED) ‘X are mine’).

<sup>11</sup> The referents of noun phrases formed with the singular NPC *=ta* (F)/ *=ha* (M) are neutral as to animacy – they can be either animate or inanimate, regardless of NP vs. modification use. On the other hand, while *=re* as the plural of the NPC in modification use is also neutral as to animacy, the referents of noun phrases formed with the plural NPC *=re* as used in NP use have to be animate (except for demonstratives).

The enclitic *=re* has another function as the plural of the enclitic *=riččo* ‘thing’. (Both forms are bound morphemes, which cannot stand alone.) See Appendix 1 for differences between the plural NPC *=re* and *=re* as the plural of *=riččo*.

Table 9. Allomorphs of case suffixes

| Case         | NOM   |     | GEN   |     | DAT-LOC |     |
|--------------|-------|-----|-------|-----|---------|-----|
| Modification | UNMOD | MOD | UNMOD | MOD | UNMOD   | MOD |
| Gender       |       |     |       |     |         |     |
| FEM          | -ø    |     | -te   | -ø  | -te     | -ra |
| MASC         | -u    | -i  | -u    | -i  | -ho     |     |

The accusative case is marked only with a suprafix, which is realized as a high pitch on the ultimate vowel, rather than the penultimate vowel, where a high pitch occurs in citation and nominative noun forms. The location of the suprafix depends on the (syntactic) modification distinction, namely, whether or not a common noun is accompanied by any grammatical nominalization in its modification use. When common nouns are modified, the suprafix occurs on the grammatical nominalizations in their modification use. (However, when a grammatical nominalization with the noun-phrase enclitic modifies a common noun, the noun also takes the suprafix.)

## 2.2 Grammatical nominalizations in Sidaama

Cardinal numerals with or without the NPC can stand alone to form a noun phrase without a noun and can modify a noun. Demonstratives, which can be analyzed as containing the NPC, can also be an independent noun phrase by themselves and be a modifier of a noun. Adjectives, to which the NPC cannot attach, can stand by themselves to form a noun phrase without a noun and can modify a noun. Without the NPC added to them, ordinal numerals, genitive noun phrases, and relative clauses can each be a noun modifier, but cannot form a noun phrase by themselves. Complement clauses without the NPC can neither form a noun phrase alone nor modify a noun.

Genitive noun phrases, relative clauses, and complement clauses must each take the NPC to form a noun phrase without a noun by themselves. In this case, the NPC makes the gender distinction of the formed noun phrase, though complement clauses take the feminine form =*ta* in most cases (Section 3.1-[5b]).

Like nouns (common nouns and some proper nouns), cardinal numerals (without the NPC) and adjectives both in their NP use and modification use inflect for case. Because the nominative, the genitive, and the dative-locative case suffixes have different gender allomorphs (described in Section 2.1 and summarized in ), like nouns, these types of nominalizations make the gender distinction with the allomorphs of the case suffixes. Demonstratives, which can be analyzed as containing the NPC, are also marked for gender with the allomorphs of the case suffixes included in them.

On the other hand, when not accompanied by the NPC, ordinal numerals, relative clauses, and complement clauses do not change their forms regardless of the grammatical case of the noun phrase. Genitive noun phrases are in the genitive case by definition.

### [1] Numerals

Cardinal numerals and corresponding ordinal numerals take forms morphologically related to each other, as exemplified in Table 10.

Table 10. Examples of cardinal and ordinal numerals in Sidaama

| Number | Cardinal numeral                                                                                         | Ordinal numeral                    |
|--------|----------------------------------------------------------------------------------------------------------|------------------------------------|
| 1      | mítte (F)/mítto (M)<br>(citation form: mítto, counting form: míte)                                       | mitikkí/umikkí/umí                 |
| 2      | láme                                                                                                     | lainkí                             |
| 3      | sáse                                                                                                     | saikkí                             |
| 4      | šoóle                                                                                                    | šoolkí                             |
| 5      | ónte                                                                                                     | ontikkí                            |
| 6      | lée                                                                                                      | leikkí                             |
| 7      | lamála                                                                                                   | lamalkí                            |
| 8      | sétte                                                                                                    | settikkí                           |
| 9      | hónse                                                                                                    | honsikkí                           |
| 10     | tónne                                                                                                    | tonnikkí                           |
| 20     | lemóo                                                                                                    | lemiikkí                           |
| 21     | lemií = na mítte (F)/lemií = na mítto (M)<br>(citation form: lemiina mítto, counting form: lemiina míte) | lemií = na mitikkí                 |
| 30     | saǰǰóo                                                                                                   | saǰǰiikkí                          |
| 32     | saǰǰií = na láme (30=and two)                                                                            | saǰǰií = na lainkí (30=and second) |

Numerals make the gender distinction for the nominative, the genitive, and the dative-locative cases with different case forms, but not for the accusative-oblique case. Only the cardinal numeral for ‘one’ *mitte* (F)/*mitto* (M) and those containing it make the gender distinction with these two gender forms even when they are in the accusative-oblique case. (The citation form for ‘one’ *mitto* and forms containing it (e.g., ‘11’, ‘21’) are used only when saying the numbers in isolation. The counting form for ‘one’ *míte* and forms containing it are used exclusively in counting.)

## [2] Demonstratives

The demonstratives in Sidaama make a four-way distinction between proximal (‘close to the speaker’), medial (‘close to the addressee or somewhat close to but not far from the speaker’), and two distals (‘far from the speaker’ and ‘very far from the speaker’). There are two types of demonstratives, (i) demonstrative pronouns/adnominal demonstratives (Table 11 and Table 12), which locate entities and (ii) place demonstrative pronouns (e.g., *kawiiččo/kowiiččo* ‘this place’), which mainly express locations. The present study deals only with the first type, whose singular forms make the gender distinction. The singular demonstratives, which preserve the proto-Afro-Asiatic demonstrative *t* (F) vs. *k* (M) contrast, seem to be cognate with the NPC =*ta* (F)/=*ha* (M), and the plural ones clearly contain =*re*

(PL).<sup>12</sup> The present study analyzes the singular demonstrative forms as containing part of the NPC *t* (F)/*k* (M) and the vowels after *t* (F)/*k* (M) as making the gender distinction with variants of the case suffixes: *i* vs. *u* for the nominative and *e* vs. *o* for the genitive and the accusative-oblique. Therefore, each singular demonstrative contains the NPC and one of the case suffixes, both of which make the gender distinction.

Table 11. Singular demonstratives<sup>13</sup>

| Distance category          | GEND | NOM   | GEN                         | ACCOBL                |
|----------------------------|------|-------|-----------------------------|-----------------------|
| ‘this (one)’               | F    | tíni  | tenné/te/teé/teyí           | teé/tenné/te          |
|                            | M    | kúni  | konní/ko/koí                | koé/konné/ko          |
| ‘that (one)’               | F    | hátti | hattenné/hatté              | hatteé/hattenné       |
|                            | M    | hákkú | hakkuí/hakkó/hakkonní       | hakkoé/hakkonné/hakkó |
| ‘(the one) over there’     | F    | tíí’i | tee’enné/tee’eé             | tee’enné/tee’eé       |
|                            | M    | kuú’u | koo’onní/koo’í/koo’ó/koo’uí | koo’onné/koo’oé       |
| ‘(the one) way over there’ | F    | tí’’i | te’’enné/te’’eé             | te’’enné/te’’eé       |
|                            | M    | kú’’u | ko’’onní/ko’’ó/ko’’uí       | ko’’onné/ko’’oé       |

Table 12. Plural demonstratives<sup>14</sup>

| Distance category                 | NOM     | GEN     | ACCOBL  |
|-----------------------------------|---------|---------|---------|
| ‘these (ones)’ (F/M)              | kúri    | kurí    | koré    |
| ‘those (ones)’ (F/M)              | hakkúri | hakkurí | hakkoré |
| ‘(the ones) over there’ (F/M)     | kuu’úri | kuu’urí | koo’oré |
| ‘(the ones) way over there’ (F/M) | ku’’úri | ku’’urí | ko’’oré |

### [3] Genitive noun phrases

Genitive noun phrases themselves do not make a gender distinction for possessums.

Genitive noun phrases that are not pronominal take the form of either (i) nouns in the genitive (or those modified by nominalizations in their modification use) or (ii) nominalizations in their NP use in the genitive (or those modified by nominalizations in their modification use). Genitive noun phrases of the second type lack a noun.

Genitive pronouns, which refer to possessors, have different gender forms only for the third-person singular, as shown in the column for the genitive case of Table 5.

<sup>12</sup> Sasse (1984) hypothesizes that the masculine demonstratives in East Cushitic languages developed from *\*ku* (nominative) and *\*ka* (accusative) to *\*hu* (nominative) and *\*ha* (accusative).

<sup>13</sup> The genitive forms *te*, *ko*, *hakkó*, *koo’ó*, and *ko’’ó* and the accusative-oblique forms *te*, *ko*, and *hakkó* each have only modification use, but no NP use.

<sup>14</sup> The plural demonstrative forms each have both NP use and modification use.

#### [4] Adjectives

As briefly mentioned in Section 1, the distinction between adjectives and nouns is not clear for words for humans. The present study regards those lexemes whose non-genitive as well as genitive forms can modify nouns as adjectives.

#### [5] Verb-based nominalizations (specifically, relative clauses and complement clauses)

A relative clause in this language uses no relative pronoun. A relative clause is formed either by means of the gapping strategy or pronominal retention (Kawachi 2020: 685–688). The verb of the relative clause can be in any of its aspectual forms used as a matrix verb: imperfective, distant perfect, recent perfect, continuous, or progressive aspect.

In addition to the non-classifying nominalizer as one of the complementizers in this language =*gede* (e.g., (3)), the singular NPC =*ta* (SG.F) can form a complement clause as the complementizer except that only the verb *lab-* ‘appear, seem’ takes =*ha* (SG.M) as its complementizer.

### 3. DATA

This section presents data on gender marking on grammatical nominalizations with the singular noun-phrase clitic (NPC) (Section 3.1) and with the nominative and the genitive case suffixes (Section 3.2). Figure 2 summarizes the distributions of the gender marking with both the NPC and the case suffixes for grammatical nominalizations at the same time in Shibatani’s (2021, 2023) implicational hierarchy, simplifying complexity of phenomena that are difficult to analyze (discussed shortly).

Figure 2. Gender marking on five types of grammatical nominalizations with the noun-phrase enclitic =*ta* (F)/= *ha* (M) and the nominative, the genitive, and the dative-locative case suffixes (Figure 3 and Figure 5 combined)

NP use:

|                 |                    |           |            |
|-----------------|--------------------|-----------|------------|
| caseCARD.NUM/ > | bothDEM > NPCGEN > | caseADJ > | NPCREL.CL/ |
| NPCORD.NUM      |                    |           | NPCCOMP.CL |

Mod use:

|                 |                         |           |                 |
|-----------------|-------------------------|-----------|-----------------|
| caseCARD.NUM/ > | bothDEM > (NPCOPT)GEN > | caseADJ > | (NPCOPT)REL.CL/ |
| (NPCOPT)ORD.NUM |                         |           | *COMP.CL        |

|           |                                                                                             |
|-----------|---------------------------------------------------------------------------------------------|
| NPC:      | gender is marked with NPC (the use of NPC is obligatory for the formation of a noun phrase) |
| (NPCOPT): | gender is optionally marked with NPC                                                        |
| both:     | including NPC and the nominative, genitive, or dative-locative case suffix                  |
| case:     | gender is marked with the nominative, genitive, or dative-locative case suffix              |
| *:        | impossible                                                                                  |



### 3.1 Gender marking with the noun-phrase enclitic

Figure 3 presents the use of the NPC to mark gender on grammatical nominalizations in Shibatani's implicational hierarchy.<sup>15</sup>

Figure 3. Gender marking on five types of grammatical nominalizations with the noun-phrase enclitic =*ta* (F)/ =*ha* (M)

NP use:

|                  |       |       |            |         |
|------------------|-------|-------|------------|---------|
| EMPH CARD.NUM/ > | DEM > | GEN > | *ADJ >     | REL.CL/ |
| ORD.NUM          |       |       | (ADJ-iidi) | COMP.CL |

Mod use:

|                  |       |           |            |             |
|------------------|-------|-----------|------------|-------------|
| EMPH CARD.NUM/ > | DEM > | OPT GEN > | *ADJ >     | OPT REL.CL/ |
| OPT ORD.NUM      |       |           | (ADJ-iidi) | *COMP.CL    |

- ADJ-iidi can attach to an adjective with the comparative/superlative suffix ADJ-*iid̥s*  
 EMPH: optional and emphatic (The NPC is optionally used for emphasis, but without it, gender can be marked with the nominative, genitive, or dative-locative case suffix.)  
 OPT: optional (The NPC is optionally used, but without it, gender is not marked.)  
 \*: impossible (The NPC cannot attach to the structure.)  
 unmarked: obligatorily marked for gender with NPC (the use of NPC is obligatory for the formation of a noun phrase) or analyzed as including NPC

Figure 4 shows whether the fact of Modification is indicated with the nominative and the genitive suffixes on a masculine noun that the grammatical nominalization with the NPC modifies: their allomorph for Modified nouns is *-i* and that for Unmodified nouns is *-u*. (There is not sufficient data on the dative-locative suffix.)

Figure 4. Allomorphs of the nominative and the genitive suffixes on a masculine noun modified by the grammatical nominalization

Mod use:

|                          |                        |                     |                           |                          |
|--------------------------|------------------------|---------------------|---------------------------|--------------------------|
| <sup>i</sup> CARD.NUM/ > | <sup>i(-u)</sup> DEM > | <sup>-u</sup> GEN > | *ADJ >                    | <sup>-i/-u</sup> REL.CL/ |
| <sup>-u</sup> ORD.NUM    |                        |                     | ( <sup>-u</sup> ADJ-iidi) | *COMP.CL                 |

- ADJ-iidi can attach to an adjective with the comparative/superlative suffix ADJ-*iid̥s*  
<sup>-i</sup>: The allomorph for Modified nouns must be used.  
<sup>-u</sup>: The allomorph for Unmodified nouns must be used.  
<sup>-i/-u</sup>: Either allomorph can be used.  
<sup>-i(-u)</sup>: The allomorph for Modified nouns is usually used, but it is possible to use that for Unmodified nouns.

<sup>15</sup> From this section onward in the main text, the NPC refers to the singular NPC =*ta* (F)/ =*ha* (M) and excludes the plural NPC, which is described in Appendix 1.



- (10) *sás-u = n = ku* *uw-ø-i.*  
 three-NOM.M=EMP=NPC.M.NOM fall-3SG.M-R.PRF.3SG.M  
 ‘All the three (M) (DEF) fell down.’

- (11) *sáse = n = ti* *ub-b-u.*  
 three.NOM.F=EMP=NPC.F.NOM fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘All the three (F) (DEF) fell down.’

#### Mod use

- (12) *láme = n = ti* *ooso*  
 two.NOM.F=EMP=NPC.F.NOM child.PL.NOM.F  
*ub-b-u/\*uw-ø-i.*  
 fall-3SG.F/3PL-R.PRF.3SG.F/3PL/fall-3SG.M-R.PRF.3SG.M  
 ‘Both children (gender unspecified) (DEF) fell down.’  
 (Compare: *láme ooso ub-b-u.* (The) two children (gender unspecified) (DEF/INDEF) fell down.)

- (13) *sás-u = n = ku* *mann-i/\*mann-u*  
 three-NOM.M=EMP=NPC.M.NOM people-NOM.M.MOD/people-NOM.M  
*dag-g-u.*  
 come-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘All the three people (gender unspecified) (DEF) came.’  
 (Compare: *sásu manni dag-g-u.* (The) three people (gender unspecified) (DEF/INDEF) came.)

- (14) *sás-u = n = ku/\*sáse = n = ti*  
 three-NOM.M=EMP=NPC.M.NOM/three.NOM.F=EMP=NPC.F.NOM  
*mann-i/\*mann-u* *da-ø-i.*  
 people-NOM.M.MOD/people-NOM.M come-3SG.M-R.PRF.3SG.M  
 ‘All the three men (DEF) came.’

- (15) *sás-u = n = ku*  
 three-NOM.M=EMP=NPC.M.NOM  
*manč-i/beett-i/\*manč-u/\*beett-u*  
 person-NOM.M.MOD/child-NOM.M.MOD/person-NOM.M/child-NOM.M  
*da-ø-i/\*dag-g-u.*  
 come-3SG.M-R.PRF.3SG.M/come-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘All the three men/boys came.’

- (16) *sáse = n = ti* *mančo/beetto*  
 three.NOM.F=EMP=NPC.F.NOM person.NOM.F/child.NOM.F  
*dag-g-u/\*da-ø-i.*  
 come-3SG.F/3PL-R.PRF.3SG.F/3PL/come-3SG.M-R.PRF.3SG.M  
 ‘All the three women/girls came.’

The gender of the referent of the noun phrase with a numeral other than ‘one’ is unspecified unless the numeral modifies a singular noun whose gender depends on the biological gender of the referent (e.g., (15), (16)). (Because the default gender is masculine, the use of the feminine NPC can be used to indicate the feminine gender of the referent, e.g., (18), (20) below.)

The emphatic NPC *=n=ta* (F)/*=n=ka* (M) can occur with the numeral for ‘one’ in a negative sentence to express ‘not even a single ...’ (e.g., (17)–(20)), in which case the formed noun phrase is not definite.

NP use

- (17) *mítt-u = n = ku*  
 one-NOM.M=EMP=NPC.M.NOM  
*di = da-ø-ino/\*di = dag-g-ino.*  
 NEG=come-3SG.M-D.PRF.3/NEG=come-3SG.F/3PL-D.PRF.3  
 ‘Not even a single man/masculine thing came.’ or ‘Not even a person/thing came.’

- (18) *mítte = n = ti* *di = dag-g-ino.*  
 one.NOM.F=EMP=NPC.F.NOM NEG=come-3SG.F-D.PRF.3  
 ‘Not even a single woman/feminine thing came.’

Mod use

- (19) *mítt-u = n = ku* *manč-i/\*manč-u*  
 one-NOM.M=EMP=NPC.M.NOM person-NOM.M.MOD/person-NOM.M  
*di = da-ø-ino.*  
 NEG=come-3SG.M-D.PRF.3  
 ‘Not even a single man came.’ or ‘Not even a single person came.’

- (20) *mítte = n = ti* *mančo*  
 one.NOM.F=EMP=NPC.F.NOM person.NOM.F  
*di = dag-g-ino.*  
 NEG=come-3SG.F-D.PRF.3  
 ‘Not even a single woman came.’

### [1b] Ordinal numerals

The NPC can attach to an ordinal numeral to form a noun phrase (e.g., (21), (22)). An ordinal numeral with the NPC seems to modify a noun (e.g., (23)), but the fact of Modification cannot be conveyed by the use of the allomorph of the nominative or the genitive suffix on a

Modified masculine noun *-i*. For example, in (23), the allomorph for an Unmodified masculine noun *-u* must be used, but that for a Modified masculine noun *-i* cannot be used.

NP use

- (21) saikkí = hu                      dí = no.  
 third=NPC.M.NOM    NEG=come.to.exist.D.PRF.3  
 ‘The third one (M) is missing.’ (Pear 12-(13)/13-(7))
- (22) (Talking about a marathon race)  
 GG: etiopp’íy-u                      meekkiha                      ful-ø-i?  
 Ethiopia-NOM.M              which.rank.M.NOM              exit/ascend-3SG.M-R.PRF.3SG.M  
 ‘Which rank did Ethiopia get?’  
 KA: saikkí = ha                      ful-ø-ino.  
 third=NPC.M.ACCOBL              exit/ascend-3SG.M-D.PRF.3  
 ‘Ethiopia got the third rank.’ (Conversation 2023.8-5: K.A. & G.G.)

Mod use

- (23) saikkí = hu                      gimbool-u/\*gimbool-i  
 third=NPC.M.NOM              basket-NOM.M/basket-NOM.M.MOD  
 dí = no.  
 NEG=come.to.exist.D.PRF.3  
 ‘The third basket is missing.’

## [2] Demonstratives

As stated in Section 2.2, the present study analyzes the singular demonstratives as containing the NPC and one of the case suffixes, which both make the gender distinction. (24) and (25) are examples of NP use and (26) and (27) are ones of modification use. In modification use, when the Modified noun is masculine and is in the nominative or the genitive, the fact of Modification is usually indicated by the selection of the allomorph for a Modified noun *-i* (e.g., (27a)), but the allomorph for an Unmodified noun *-u* can occur (e.g., (27b)), though its use is less common and is restricted to informal speech.

NP use

- (24) hakonné                      fe-ø-é                      fe-ø-é                      haa’r-ø-é                      ...  
 that.M.ACC              clean-3SG.M-CVB              clean-3SG.M-CVB              take-3SG.M-CVB  
 ‘He cleaned that one (M) repeatedly, and took them, ...’ (Pear 21-(1))
- (25) kúni                      danča = ho.  
 this.M.NOM              good=NPC.M.PRED  
 ‘This (M) is good.’

Mod use

- (26) hakkoné              gimboola              fe-ø-é              fe-ø-é  
       that.M.ACC        basket.ACC        clean-3SG.M-CVB    clean-3SG.M-CVB  
       haa'r-ø-é              ...  
       take-3SG.M-CVB  
       'He cleaned that basket repeatedly, and took them, ...'

- (27) a.      kúni              beett-i              danča = ho.  
           this.M.NOM    child.NOM.M.MOD    good=NPC.M.PRED  
           'This boy is good.'

- b.      kúni              beett-u              danča = ho.  
           this.M.NOM    child.NOM.M        good=NPC.M.PRED  
           'This boy is good.'

The NPC cannot additionally attach to the demonstratives, as in (28) and (29).

NP use

- (28) \*hakkonné = ha              fe-ø-é              fe-ø-é  
       that.M.ACCOBL=NPC.ACC    clean-3SG.M-CVB    clean-3SG.M-CVB  
       haa'r-ø-é              ...  
       take-3SG.M-CVB  
       *to mean*, 'He cleaned that one (M) repeatedly, and took them, ...'

Mod use

- (29) \*hakkonné = ha              gimboolá              fe-ø-é  
       that.M.ACCOBL=NPC.ACC    basket.ACCOBL    clean-3SG.M-CVB  
       fe-ø-é              haa'r-ø-é              ...  
       clean-3SG.M-CVB    take-3SG.M-CVB  
       *to mean*, 'He cleaned that basket repeatedly, and took them, ...'

### [3] Genitive noun phrases

The NPC can attach to genitive noun phrases to form a noun phrase (e.g., (30)–(33)). The NPC with genitive noun phrases seems to modify a noun (e.g., (34)–(37)), but the Modified masculine noun in the nominative or the genitive carries the allomorph of the case suffix for an Unmodified masculine noun *-u* and cannot take that for a Modified masculine noun *-i* (e.g., (35)–(37)), as in the case of the NPC with ordinal numerals in their modification use.<sup>18</sup>

<sup>18</sup> Genitive noun phrases to which the NPC *=ta/=ha* can attach include genitive demonstrative pronouns. For example, there are noun phrases like *tenné=ta* 'this one (F)'s (F)', *konní=ta* 'this one (M)'s (F)', *konní=ha* 'this one (M)'s (M)', and *tenné=ha* 'this one (F)'s (M)', *tenné=re* 'this one (F)'s (PL)', *konní=re* 'this one (M)'s (PL)', *hakkurí=ha* 'those ones' (M)'.

# NP use

- (30) mitt-ú                      gimbool-i = ha                      add-ø-e  
 one-GEN.M              basket-GEN.M.MOD=NPC.M.ACC              take-3SG.M-CVB  
 ha'r-ø-ewo.  
 leave-3SG.M-D.PRF.3  
 'He took one basket of (M) (*lit.*, one basket's) (e.g., papayas) and left.' (Pear 19-(9))

- (31) ané = hu                      danča = ho.  
 1SG.GEN=NPC.M.NOM              good=NPC.M.PRED  
 'Mine (M) (DEF) is good.'

- (32) beett-u-nni = hu                      ba'-ø-ino.  
 child-GEN.M-ABLINS=NPC.M.NOM              disappear-3SG.M-D.PRF.3  
 'The/\*a boy's (M) (DEF) disappeared.'

- (33) ané = hu                      ba'-ø-ino.  
 1SG.POSS=NPC.M.NOM              disappear-3SG.M-D.PRF.3  
 'Mine (M) (DEF) disappeared.'

# Mod use

- (34) mitt-ú                      gimbool-i = ha/gimbool-i  
 one-GEN.M              basket-GEN.M.MOD=NPC.M.ACC/basket-GEN.M.MOD  
 paappaayya                      add-ø-e                      ha'r-ø-ewo.  
 papaya.ACCOBL              take-3SG.M-CVB              leave-3SG.M-D.PRF.3  
 'He took one basket of papayas and left.' (*gimbool-i = ha = gimbool-i*)<sup>19</sup>

- (35) hakkuí                      beett-i = hu  
 that.M.GEN              child-GEN.M.MOD=NPC.M.NOM  
 rod-úu/\*rod-ii                      ba'-ø-ino.  
 sibling-NOM.M/sibling-NOM.M.MOD              disappear-3SG.M-D.PRF.3  
 'That boy's brother disappeared.'

<sup>19</sup> Another construction used for a noun phrase in the accusative-oblique case containing a cardinal numeral and a measure word uses the accusative-oblique case for all its components, as in (i).

- (i) mittó                      gimboola                      paappaayya                      add-ø-e  
 one.ACCOBL              basket.ACCOBL              papaya.ACCOBL              take-3SG.M-CVB  
 ha'r-ø-ewo.  
 leave-3SG.M-D.PRF.3  
 'He took one basket of papayas and left.'

- (36) beett-ú-nni = hu                      rod-úu/\*rod-ii  
child-GEN.M-ABLINS=NPC.M.NOM    sibling-NOM.M/sibling-NOM.M.MOD  
ba'-ø-ino.  
disappear-3SG.M-D.PRF.3  
'The boy's brother disappeared.'
- (37) ané = hu                      rod-úu/\*rod-ii  
1SG.POSS=NPC.M.NOM    sibling-NOM.M/sibling-NOM.M.MOD  
ba'-ø-ino.  
disappear-3SG.M-D.PRF.3  
'My brother disappeared.'

When the NPC attaches to a genitive noun that is Unmodified (is not accompanied by any nominalization or the possessive pronominal suffix), the suffix *-nni* (or its short form *-i*) has to intervene between the noun and the enclitic: e.g., *gereew-ú-nni = ha/gereew-ú-i = ha* (goat-GEN.M-*nni*=NPC.M.PRED) ‘sheep’s (e.g., meat)’. The function of this suffix is not clear. (It may be epenthetically inserted.)

#### [4] Adjectives

The NPC generally cannot attach to adjectives regardless of their use, as in (38) and (39).

NP use

- (38) \*busul-u=hu                      da-ø-ino.  
smart-NOM.M=NPC.M.NOM        come-3SG.M-D.PRF.3  
*to mean*, ‘The smart one (M) came.’

Mod use

- (39) \*busul-u = hu                      beett-u/beett-i  
 smart-NOM.M=NPC.M.NOM        child-NOM.M/child-NOM.M.MOD  
 da-Ø-ino.  
 come-3SG.M-D.PRF.3  
*to mean*, ‘The smart boy came.’

However, there are two subclasses of adjectives that the NPC can attach to. One is those adjectives that are invariant in form: e.g., *lowo* ‘big’, *wona* ‘previous’. Such adjectives with the NPC have both NP use and modification use (e.g., (40), (41)).



(40) mǐtt-u                  manč-i                  baďđe-é  
one-NOM.M        person-NOM.M.MOD back-LV  
no-o = hu                                                              ...     lowó = hu  
come.to.exist.D.PRF.3-LV=NP.C.M.NOM                                      big=NP.C.M.NOM  
no.  
come.to.exist.D.PRF.3  
'One man, the one (M) behind ... the big one (M) existed.' (Pear #12-(9))

(41) lowó = hu                      mánč-u/\*manč-i  
big = NPC.M.NOM              person-NOM.M/person-NOM.M.MOD  
no.  
come.to.exist.D.PRF.3  
'The big man existed.'

NP use

(42) šiim-iidí = hu                      da-ø-ino.  
small-relatively=NPC.M.NOM    come-3SG.M-D.PRF.3  
'The smaller/smallest one (M) (DEF) came.'

(43) šiim-iidí = hu                      beétt-u/\*beett-i  
small-relatively=NPC.M.NOM    child-NOM.M/child-NOM.M.MOD  
da-ø-ino.  
come-3SG.M-D.PRF.3  
'The smaller/smallest boy came.'

### [5a] Relative clauses

314

two conditions, first, when a relative clause follows a noun (e.g., (46)), and second, when there is a pause between the relative clause and the noun with both exhibiting high pitch at their ends (e.g., (47), (48)). A relative clause in modification use formed with the NPC is non-restrictive and a formed noun phrase has a definite referent. A masculine noun in the nominative or genitive case modified only by this type of relative clause takes the allomorph for Unmodified nouns (e.g., (46), (47)).<sup>20</sup>

#### NP use

- (44) wo-dǫ-ø-e                      no-o = ha  
 put-MID-3SG.M-CVB come.to.exist.D.PRF.3-LV=NPC.M.ACC  
 haw-ø-e                      ha'r-ø-ino.  
 forget-3SG.M-CVB      leave-3SG.M-D.PRF.3  
 'He forgot about the one (M) that he was wearing, and left.'

#### Mod use

- (45) wo-dǫ-ø-e                      no-o = ha                      kooffiyyá  
 put-MID-3SG.M-CVB come.to.exist.D.PRF.3-LV=NPC.M.ACC      hat.ACC  
 haw-ø-e                      ha'r-ø-ino.  
 forget-3SG.M-CVB      leave-3SG.M-D.PRF.3  
 'He forgot about the hat, which he was wearing, and left.' (Pear 19-(14))  
 (Compare: (76) 'He forgot about the/a hat that he was wearing, and left.')

- (46) beett-ú, [pause] ané                      la'-ø-a-nni  
 child-NOM.M      1SG.ACC look-3SG.M-INF-ABLINS  
 no-o = hú, [pause]                      agur-ø-e                      ka'eentinni  
 come.to.exist.3.D.PRF-LV=NPC.M.NOM      leave-3SG.M-CVB      then  
 dod-ø-a-nni                      ha'r-ø-i.  
 run-3SG.M-INF-ABLINS      go-3SG.M-R.PRF.3  
 'The boy, who was looking at me, left me, and then went running.' (NINJAL A-14, SUBJ #2)

- (47) utaám-u                      wóna                      it-ø-ino = hu  
 hedgehog-NOM.M      previously eat-3SG.M-D.PRF.3=NPC.M.NOM  
 mataat'eéss-u/\*mataat'eéss-i                      basé  
 sugar.potatoes-NOM.M/sugar.potatoes-NOM.M.MOD      space.ACCOBL  
 amad-ø-ino-si = daafira,                      ...  
 occupy-3SG.M-D.PRF.3-3SG.M=because  
 'Because the hedgehog became full with sugar potatoes, the one (M) that it (the hedgehog) previously ate, ...' (Story: 'Muddamino gotiiči saate buuda ga'mino')

<sup>20</sup> In (48), the allomorph of the nominative case suffix for Modified nouns is used in *manč-i* (person-NOM.M.MOD) because of *durreéss-u* (rich-NOM.M). If *mančo* 'person' were not accompanied by *durreéss-u* (rich-NOM.M), it would take the form *manč-u* (person-NOM.M), which carries the allomorph of the nominative case suffix for Unmodified nouns, rather than *manč-i*.

- (48) kúri lám-u = n = ku wosin-i  
 these.NOMtwo-NOM.M=EMP=NPC.M.NOM guest-NOM.M.MOD  
 woná = n = ku-nni hasaaw-ø-annó = re  
 previous=EMPH=NPC.M.GEN-ABLINS talk-3SG.M-IPFV.3=NPC.PL.ACCOBL  
 mačč'iišš-ø-á-nni no-o = hú  
 listen-3SG.M-INF-ABLINS come.to.exist.D.PRF.3-LV=NPC.M.NOM  
 durreéss-u manč-í mul-í-nsa-ra  
 rich-NOM.M person-NOM.M.MOD nearness-GEN-3PL-ALL  
 gámba y-e-é ...  
*gámba* say-CVB-LV  
 'The rich man, who was listening to the things that these two guests had been talking about since then, approached them ...' (Story: 'Durreesssu duučanta assanno')

### [5b] Complement clauses

The NPC =*ta* (SG.F) as the complementizer has only NP use in which it forms a complement clause (e.g., (49)). The NPC =*ha* (SG.M), which can be used with the verb *lab-* 'appear, seem' as the main verb also has only NP use (e.g., (50)). A complement clause with the NPC cannot modify a noun.

- (49) loos-u-i=wa ha'r-oo-tt-o=ta  
 work-GEN.M-ABLINS=place go-D.PRF.2-2SG-M=NPC.F.ACCOBL  
 di=af-oo-mm-o=nka.  
 NEG=get.to.know-D.PRF.1-1SG-M=EMP  
 'I (M) never knew that you (SG.M) went to work.' (Conversation 2023.8-5)
- (50) alb-í-ra akk'ad-ø-e  
 face-GEN.M.MOD-DATLOC.MOD plan(AMH)-1SG-CVB  
 assaaw-oo-mm-a=ri  
 think(AMH)-D.PRF.2-2SG-F=NPC.PL.NOM  
 nó-o = ha di = law-ø-anno-'e.  
 come.to.exist.D.PRF.3-LV=NPC.M.ACCOBL NEG=appear-3SG.M-IPFV.3-1SG  
 'It does not appear to me that there are things I am planning to do and thinking of doing in the future.' (Conversation 2019.3-4: A.T. & A.K.)

## 3.2 Gender marking with the nominative, the genitive, and the dative-locative case suffixes

Figure 5 shows how the nominative, the genitive, and the dative-locative case suffixes are used with the five types of grammatical nominalizations to mark gender in Shibatani's implicational hierarchy.

Figure 5. Gender marking on five types of grammatical nominalizations with the nominative, the genitive, and the dative-locative case suffixes

NP-use:

|             |       |        |       |          |
|-------------|-------|--------|-------|----------|
| CARD.NUM/ > | DEM > | *GEN > | ADJ > | *REL.CL/ |
| *ORD.NUM    |       |        |       | *COMP.CL |

Mod-use:

|             |       |        |       |          |
|-------------|-------|--------|-------|----------|
| CARD.NUM/ > | DEM > | #GEN > | ADJ > | #REL.CL/ |
| #ORD.NUM    |       |        |       | *COMP.CL |

\*: impossible

#: modifiable but does not take the case suffixes, so no gender distinction is made

unmarked: obligatorily marked for gender with one of the case suffixes (its use is obligatory when the noun phrase is in the case) or analyzed as including the case marker

## [1] Numerals

### [1a] Cardinal numerals

The referent of the noun phrase formed with a cardinal numeral can be definite or indefinite. The NP use of a cardinal numeral is exemplified in (51)–(53). When the plural form of a noun or grammatical nominalization that does not have intrinsic gender is the subject, its masculine form is used with a verb in its feminine form for gender-unspecified referents (e.g., (51)) and with a verb in its masculine form for exclusively male referents or referents for which masculine plural nouns (e.g., *k'aakk'uulle* 'babies, children', *geerra* 'elderly', *meento* 'women', *seenne* 'girls') would be used (e.g., (52)); on the other hand, its feminine form is used with a verb in its feminine form for exclusively female referents or referents for which feminine plural nouns would be used (e.g., (53)).

NP use

(51) *sás-u*                      *ub-b-u*.  
 three-NOM.M      fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 '(The) three (gender unspecified) fell down.'

(52) *sás-u*                      *uw-ø-i*.  
 three-NOM.M      fall-3SG.M-R.PRF.3SG.M  
 '(The) three (M) fell down.'

(53) *sáse*                      *ub-b-u*.  
 three.NOM.F      fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 '(The) three (F) fell down.'

A cardinal numeral can modify unmarked, singular, or plural forms of countable nouns (see Section 1, e.g., ‘sibling’, unmarked: *rodoo* (F/M), singular: *rod-iiččo* (F/M), plural: *rod-uuwa* (F(M))/*rod-uubba* (F)) (e.g., (54)–(62)).

Mod use

- (54) *sás-u*                      *rod-ii*                      *ub-b-u*.  
 three-NOM.M    sibling-NOM.M.MOD fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘(The) three siblings fell down.’
- (55) *sás-u*                      *rod-ii*                      *uw-ø-i*.  
 three-NOM.M    sibling-NOM.M.MOD fall-3SG.M-R.PRF.3SG.M  
 ‘(The) three brothers fell down.’
- (56) *sáse*                      *rodoo*                      *ub-b-u*.  
 three.NOM.F    sibling.NOM.F    fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘(The) three sisters fell down.’
- (57) *sás-u*                      *rod-iičč-i*                      *ub-b-u*.  
 three-NOM.M    sibling-SG-NOM.M.MOD    fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘(The) three siblings fell down.’
- (58) *sás-u*                      *rod-iičč-i*                      *uw-ø-i*.  
 three-NOM.M    sibling-SG-NOM.M.MOD    fall-3SG.M-R.PRF.3SG.M  
 ‘(The) three brothers fell down.’
- (59) *sáse*                      *rod-iiččo*                      *ub-b-u*.  
 three.NOM.F    sibling-SG.NOM.F    fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘(The) three sisters fell down.’
- (60) *sás-u*                      *rod-uuw-i*                      *ub-b-u*.  
 three-NOM.M    sibling-PL-NOM.M.MOD    fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘(The) three siblings fell down.’
- (61) *sás-u*                      *rod-uuw-i*                      *uw-ø-i*.  
 three-NOM.M    sibling-PL-NOM.M.MOD    fall-3SG.M-R.PRF.3SG.M  
 ‘(The) three brothers fell down.’
- (62) *sáse*                      *rod-uuwa*                      *ub-b-u*.  
 three.NOM.F    sibling-PL.NOM.F    fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
 ‘(The) three sisters fell down.’

### [1b] Ordinal numerals

Ordinal numerals cannot form a noun phrase by themselves. Although they can modify nouns, they are invariant in form irrespective of case and do not mark gender.

NP use

- (63) \*saikkí/\*mitikkí/\*umikkí/\*umí ba'-ø-ino.  
third/first/first/first/first disappear-3SG.M-D.PRF.3  
*to mean*, 'The first one disappeared.'

- (64) \*saikkí wonš-i-ø-'r-a-ra ...  
third fill-EP-3SG-MID-INF-DATLOC.MOD  
*to mean*, 'in order to fill the third one for himself, ...'

Mod use

- (65) saikkí gimboola wonš-i-ø-'r-a-ra ...  
third basket.ACC fill-EP-3SG-MID-INF-DATLOC.MOD  
'In order to fill the third basket for himself, ...'

- (66) saikkí/mitikkí/umikkí/umí gimbool-i/\*gimbool-u  
third/first/first/first basket-NOM.M.MOD/basket-NOM.M  
dí=no.  
NEG=come.to.exist.D.PRF.3  
'The third/first/first/first basket was missing.'

- (67) saikkí/mitikkí/umikkí/umí beett-i/\*beett-u  
third/first/first/first/first child-NOM.M.MOD/child-NOM.M  
ba'-ø-ino.  
disappear-3SG.M-D.PRF.3  
'The third/first/first/first boy disappeared.'

- (68) saikkí/mitikkí/umikkí/umí beetto ba'-'-ino.  
third/first/first/first/first child.NOM.F disappear-3SG.F-D.PRF.3  
'The third/first/first/first girl disappeared.'

### [2] Demonstratives

As mentioned in Section 2.2, the present study analyzes the singular demonstratives as containing the NPC and one of the case suffixes, which both make the gender distinction. See Section 3.1 for examples.

### [3] Genitive noun phrases

Genitive noun phrases cannot stand alone to form a noun phrase. They can modify a noun, but they are in the genitive case and the genitive case suffix makes the distinction of the gender of the possessor, not that of the possessum.

NP use

- (69) \*mitt-ú                      gimbool-i                      add-ø-e                      ha'r-ø-ewo.  
one-GEN.M              basket-GEN.M.MOD      take-3SG.M-CVB              leave-3SG.M-D.PRF.3  
*to mean*, 'He took one basket of something (*lit.*, one basket's) and left.'

Mod use

- (70) mitt-ú                      gimbool-i                      paappaayya                      add-ø-e  
one-GEN.M              basket-GEN.M.MOD      papaya(ACCOBL)              take-3SG.M-CVB  
ha'r-ø-ewo.  
leave-3SG.M-D.PRF.3  
'He took one basket of papayas and left.'

### [4] Adjectives

Almost all adjectives (See Section 3.1), which cannot take the NPC, can form a noun phrase by themselves and can modify nouns. When they take the nominative, the genitive, or the dative-locative case suffix, the suffix makes the gender distinction with its allomorphs.

NP use

- (71) busúl-u                      da-ø-ino.  
smart-NOM.M              come-3SG.M-D.PRF.3  
'The smart one (M) came.'

- (72) busúle                      dag-g-ino.  
smart.NOM.F              come-3SG.F-D.PRF.3  
'The smart one (F) came.'

Mod use

- (73) busúl-u                      beett-i                      da-ø-ino.  
smart-NOM.M              child-NOM.M.MOD              come-3SG.M-D.PRF.3  
'The smart boy came.'

- (74) busúle                      beetto                      dag-g-ino.  
smart.NOM.F              child.NOM.F              come-3SG.F-D.PRF.3  
'The smart girl came.'

## [5] Verb-based nominalizations (relative clauses)

### [5a] Relative clauses

Without the NPC, relative clauses cannot form a noun phrase by themselves. They can modify nouns, but no gender marking occurs, because they do not carry any case marker.

NP use

- (75) \*wo-dǝ-ø-e                      no-o                      haw-ø-e  
put-MID-3SG.M-CVB come.to.exist.D.PRF.3-LV    forget-3SG.M-CVB  
ha'r-ø-ino.  
leave-3SG.M-D.PRF.3  
*to mean, 'He forgot about the one that he was wearing, and left.'*

Mod use

- (76) wo-dǝ-ø-e                      no-o                      kooffiyya  
put-MID-3SG.M-CVB come.to.exist.D.PRF.3-LV    hat(ACCOBL)  
haw-ø-e                      ha'r-ø-ino.  
forget-3SG.M-CVB    leave-3SG.M-D.PRF.3  
*'He forgot about the/a hat that he was wearing, and left.'*

### [5b] Complement clauses

Without the NPC, complement clauses cannot form a noun phrase alone. They cannot modify nouns.

## 4. ANALYSIS AND DISCUSSION

The gender marking in Sidaama—both for each type of gender marker individually and for both types collectively—supports Shibatani's hypothesis about the distribution of classifying grammatical nominalizers along the functional dimension. All the five types of grammatical nominalizations in this language in NP use can be marked for gender with one of the gender markers. Moreover, the gender markers are more likely to obligatorily occur with grammatical nominalizations in NP use than those in modification use. There are types of grammatical nominalizations whose NP use takes the NPC obligatorily but modification use takes the NPC only optionally (ordinal numerals, genitive noun phrases, and relative clauses).<sup>21</sup> These indicate that the gender marking begins with NP use and spreads to modification use. On the other hand, the gender marking in this language does not seem to coincide with Shibatani's hypothesis about the use of classifying grammatical nominalizers along the structural dimension. There are gaps in cardinal numerals and adjectives in the hierarchy pair for the NPC and there is a gap in genitive noun phrases in the hierarchy pair for the case suffixes. It is not clear how gender marking developed along the structural dimension. It might be possible that gender marking originated from demonstratives (with the proto-Afro-Asiatic

---

<sup>21</sup> These three types of grammatical nominalizations cannot form a noun phrase alone but can modify a noun without them taking any marker themselves.



demonstrative *t* (F) vs. *k* (M) contrast) and spread to cardinal numerals and adjectives and then to other types of structures.

However, there are remaining issues to be examined. A major issue is whether the case suffixes developed from the NPC. The case marking pattern in Kambaata (Treis 2008) suggests a positive answer for this issue. The nominative case suffix forms for many feminine nouns contain *-t* and the accusative case suffix forms for them includes *-ta*, but there are feminine nouns that take nominative and accusative case suffix forms consisting only of a vowel (Treis 2008: 102–108). The accusative and the nominative case suffix forms for masculine nouns are vowels, but masculine nouns with a long final vowel can additionally carry *-hu* (nominative) and *-ha* (accusative) (Treis 2008: 126–127, 410), which Treis calls the secondary case/gender markers.

Assuming that the case suffixes developed from the NPC in Sidaama, there are several issues to address. First, there is a question of why the different types of grammatical nominalizations receive different kinds of gender marking. As compared to these three types of grammatical nominalizations and complement clauses, which are still at early stages of the development of gender marking from that with the NPC to that with the case suffixes, cardinal numerals and adjectives seem to have developed innovative gender marking with the case suffixes. Cardinal numerals take the NPC only for an emphasis on the notion of ‘the whole number of’ the referent of a definite noun phrase, whereas adjectives cannot take the NPC. Unlike the other types of grammatical nominalizations, which often form definite noun phrases, cardinal numerals and adjectives are neutral to definiteness, like most common nouns. Therefore, the NPC functions not only as a classifying grammatical nominalizer but also as a definiteness marker.

Second, there is a question of whether structures in the singular and those in the plural can be placed in the same hierarchy. The NPC *=ta* (SG.F)/*=ha* (SG.M) is used for singular referents. Noun phrases formed with a cardinal numeral for ‘two’ or above either in NP use or in modification use have plural referents. Therefore, they are, first of all, incompatible with the singular NPC. It may be that numerals higher than ‘one’ should not be placed in the hierarchy for the singular NPC, though they still cannot be accompanied by the plural NPC *=re* (PL).<sup>22</sup>

Third, it is not clear what the selection of the allomorphs of the case suffixes for Modified vs. Unmodified nouns means when grammatical nominalizations with the NPC modify a noun. The allomorph for Modified nouns must be used for cardinal numerals and is usually used for demonstratives. On the other hand, the allomorph for Unmodified nouns must be used for ordinal numerals and genitive noun phrases (and adjectives with *-jidi*) and either the allomorph

<sup>22</sup> The enclitic *=re* as the plural form of *=riččō* ‘thing’ (see Footnote 10) can be used with numerals. An example is given in (i), which is the plural counterpart of (ii). Noun phrases formed with a cardinal numeral and this enclitic are restricted to inanimate referents and cannot modify a noun.

- (i) lámu=r-i                                      uw-ø-ino/\*ub-b-ino.  
two=things-NOM.M.MOD    fall-3SG.M-R.PRF.3SG.M/fall-3SG.F/3PL-R.PRF.3SG.F/3PL  
‘(The) two things fell down.’
- (ii) mitt-u=ričč-i                                      uw-ø-ino/\*ub-b-ino.  
one-NOM.M=thing-NOM.M                      fall-3SG.M-R.PRF.3SG.M/fall-3SG.F-R.PRF.3SG.F  
‘(The) one thing fell down.’

for Modified nouns or that for Unmodified nouns can be used for relative clauses. These are grammatical nominalizations that form noun phrases usually interpreted as definite.

A related issue is whether different types of grammatical nominalizations exhibit varying degrees of tightness in their modification of nouns. When a grammatical nominalization modifies a noun without using the NPC, the fact of Modification can be indicated by the selection of specific allomorphs of the case suffixes. This applies to all grammatical nominalizations that can modify a noun, except that a noun modified by demonstratives can take the allomorphs of the case suffixes for Unmodified nouns in informal speech, though it usually takes the allomorphs of the case suffixes for Modified nouns. The use of the allomorphs of the case suffixes for Unmodified nouns is found with more types of grammatical nominalizations when the NPC attaches to them. The selection of these allomorphs seems to suggest the independent status of grammatical nominalizations with the NPC modifying a noun from the noun (e.g., (23), (27b)). There are also cases where grammatical nominalizations appear to be more independent to the extent that they are juxtaposed with other noun phrases.

## 5. CONCLUSION

The present study explored how gender is marked on grammatical nominalizations in their NP and modification uses in Sidaama. The data on the NPC and the case suffixes, both individually and collectively, support Shibatani's hypothesis regarding the synchronic distribution of classifying grammatical nominalization markers along the functional dimension, but not along the structural dimension. This, along with several other phenomena, requires further investigation. To better understand the diachronic development of gender marking in Sidaama, data from neighboring languages should be examined in detail. This remains a topic for future research.

## ABBREVIATIONS

|          |                       |
|----------|-----------------------|
| 1        | first person          |
| 2        | second person         |
| 3        | third person          |
| ABLINS   | ablative-instrumental |
| ACC      | accusative            |
| ACCOBL   | accusative-oblique    |
| ADJ      | adjective(s)          |
| ASP      | aspect                |
| CARD.NUM | cardinal numeral(s)   |
| COMP.CL  | complement clause(s)  |
| D.PRF    | distant perfect       |
| DAT      | dative                |
| DATLOC   | dative-locative       |
| DEM      | demonstrative         |
| EMP      | emphatic              |
| EP       | epenthesis            |
| F, FEM   | feminine              |

|         |                                                                               |
|---------|-------------------------------------------------------------------------------|
| GEN     | genitive noun phrase(s)                                                       |
| GEND    | gender                                                                        |
| HUM     | human                                                                         |
| IMPRF   | imperfective                                                                  |
| LV      | lengthened vowel                                                              |
| M, MASC | masculine                                                                     |
| MID     | middle                                                                        |
| Mod     | modification                                                                  |
| MOD     | “Modified”: accompanied by a modifier and/or the possessive pronominal suffix |
| NEG     | negative                                                                      |
| NMLZ    | nominalization(s)                                                             |
| NOM     | nominative                                                                    |
| NONHUM  | non-human                                                                     |
| NPC     | noun-phrase clitic                                                            |
| NUM     | number, numeral(s)                                                            |
| ORD.NUM | ordinal numeral(s)                                                            |
| PERS    | person                                                                        |
| PL      | plural                                                                        |
| PRED    | predicate                                                                     |
| PROP.N  | proper noun                                                                   |
| PRON    | pronoun                                                                       |
| R.PRF   | recent perfect                                                                |
| REL.CL  | relative clause(s)                                                            |
| SG      | singular                                                                      |
| V       | verb                                                                          |

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my Sidaama consultants, especially, Genene Gudura, Iyasu Gudura, and Legesse Gudura. I benefited from the comments on this paper provided by Yvonne Treis. My thanks also go to Kazuyuki Kiryu, Prashant Pardeshi, Masayoshi Shibata, and Satoko Shirai for their comments on earlier versions of this paper. The present study has been supported by grants from the Japan Society for the Promotion of Science (KAKENHI Project IDs: 22H00659 and 23K21931, ‘Cross-linguistic studies on grammatical nominalization: with a focus on classifiers and gender markers’, PI: Kazuhiro Kawachi) and the National Institute for Japanese Language and Linguistics (‘Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspectives’, PI: Prashant Pardeshi) as well as the Keio Gijuku Fukuzawa Memorial Fund for the Advancement of Education and Research, the Keio University Academic Development Funds, and the Ushioda Memorial Fund.

## REFERENCES

- Bellingham, Erika, Stephanie Evers, Kazuhiro Kawachi, Alice Mitchell, Sang-Hee Park, Anastasia Stepanova, and Jürgen Bohnemeyer. 2000. Exploring the representation of causality across languages: Integrating production, comprehension and conceptualization perspectives. In Elitzur Bar-Asher Siegal and Nora Boneh (eds.) *Perspectives on causation*, 75–119. Cham, Switzerland: Springer Press.
- Chafe, Wallace L. (ed.) 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Kaarricha, Abbayne Akililu, Awarra Awaajje Lagidde, Birge Ganbura Nonne, and Kifile Nennitte Saydo. 2021. *Sidaamu afoo: Rosaanote maxaafa, Sayikki kifile (The Sidaama language: Students' book, Third grade)*. Addis Ababa: Itiophiyu Feferaalitte Dimokiraasitte Ripaabilike Rosu Ministere & Hawassa: Sidaamu Dagoomu Qoqqowi Mootimma Rosu Biiro.
- Kawachi, Kazuhiro. 2011. Noun phrases without nouns in Sidaama. In Sutcliffe, Patricia (ed.) *LACUS Forum 36 (2009): Mechanisms of linguistic behavior*, 25–35. Houston, TX: Linguistic Association of Canada and the United States.
- Kawachi, Kazuhiro. 2020a. Chapter 42: Sidaama. In Vossen, Rainer, and Gerrit J. Dimmendaal (eds.) *The Oxford handbook of African languages*, 543–552. Oxford: Oxford University Press.
- Kawachi, Kazuhiro. 2020b. Chapter 16: Sidaama. In Tasaku Tsunoda (ed.) *Mermaid constructions: A compound-predicate construction with biclausal appearance*, 679–733. Berlin & Boston: De Gruyter Mouton.
- Kawachi, Kazuhiro. 2023. Chapter 17: Sidaama. In Ronny Meyer, Bedilu Wakjira, and Zelealem Leyew (eds.) *The Oxford handbook of Ethiopian languages*, 320–343. Oxford: Oxford University Press.
- Kawachi, Kazuhiro. 2025. Chapter 4: Does the degree of integration of constructions iconically reflect motion event types? The case of Sidaama. In Yo Matsumoto (ed.) *Motion Event descriptions across languages, Volume 1: Case studies of linguistic representations of motion*, 129–166. Berlin & Boston: De Gruyter Mouton.  
<https://doi.org/10.1515/9783110690866-004>
- Kawachi, Kazuhiro. In press. Sidaama speakers' atypical acceptability judgments of causative constructions used to describe causal event video stimuli. *Linguistics Vanguard* (Special issue "The grammar of causation: Language-specific perspectives").
- Kawachi, Kazuhiro, and Ababayehu Aemero Tekleselassie. 2012. Modification within a noun phrase in Sidaama (Sidamo). *Proceedings of the 34th Annual Meeting of the Berkeley Linguistics Society (2008), General Session*, 187–198. Berkeley, CA: Berkeley Linguistics Society. <https://doi.org/10.3765/bls.v34i1.3567>
- Kawachi, Kazuhiro, Erika Bellingham, and Jürgen Bohnemeyer 2018. Different types of causality and clause linkage in English, Japanese, Sidaama, and Yucatec Maya. *Papers from the 18th National Conference of the Japanese Cognitive Linguistics Association*, 47–59.
- Kawachi, Kazuhiro, Sang-Hee Park, and Erika Bellingham. 2020. Directness of causation and morphosyntactic complexity of constructions: Japanese and Korean cases. In Shoichi

- Iwasaki, Susan Strauss, Shin Fukuda, Sun-Ah Jun, Sung-Ock Sohn, and Kie Zuraw (eds.) *Japanese/Korean linguistics, Vol. 26*, 225–236. Stanford, CA: CSLI Publications.
- Matsumoto, Yo. (ed.) 2025. *Motion event descriptions across languages, Volume 1: Case studies of linguistic representations of motion*. Berlin & Boston: De Gruyter Mouton.
- Pardeshi, Prashant, and Masayoshi Shibatani. 2023. Marathi relative and complement clauses in nominalization perspective. In Rajesh Kumar and Om Prakash. (eds.) *Language studies in India: Cognition, structure, variation*, 321–352. Singapore: Springer Nature.
- Sasse, Hans-Jürgen. 1984. Case in Cushitic, Semitic, and Berber. In James Bynon. (ed.) *Current progress in Afro-Asiatic linguistics: Papers of the Third International Hamito-Semitic Congress*, 111–126. Amsterdam & Philadelphia: John Benjamins.
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf-char/en)
- Treis, Yvonne. 2008. *A grammar of Kambaata, Part 1: Phonology, nominal morphology, and non-verbal predication*. Köln: Rüdiger Köppe.

#### **APPENDIX 1: Use of the noun-phrase enclitic *=re* as the plural of *=ta* (F)/*=ha* (M) and the enclitic *=re* as the plural of *=riččo***

The enclitic *=re* as the plural NPC has both NP and modification uses. Unlike noun phrases formed with the singular NPC *=ta* (F)/*=ha* (M) in NP and modification uses and the plural NPC *=re* in modification use, whose referents can be either animate or inanimate, those formed with *=re* in NP use are used for animate referents (except that demonstratives in the plural, which contain it, can be used for either animate or inanimate referents). The referents of noun phrases formed with the plural NPC *=re* are normally definite.

The enclitic *=re* as the plural of the enclitic *=riččo* ‘thing’ has only NP use and does not have modification use. The referents of noun phrases formed with *=re* as the plural of *=riččo* are inanimate and can be either definite or indefinite. In these respects, *=riččo* and *=re* function the same way as each other and almost as inanimate common nouns except that they are bound.

Both *=riččo* and its plural *=re* are treated as masculine (e.g., *ikk-ø-ino =riččo* (become-3SG.M-D.PRF.3=thing) ‘a certain thing’, *ikk-ø-ino =re* (become-3SG.M-D.PRF.3=thing.PL) ‘certain things’), and when *=re* as the plural of *=riččo* occurs at the end of a subject noun phrase, the verb takes the third-person singular masculine suffix. On the other hand, a noun phrase formed with the plural NPC *=re* functions as plural (e.g., *ikk-i-t-ino =re* (become-EP-3PL-D.PRF.3=NPC.PL) ‘certain people/animals’), and when it is the subject of a sentence, the verb takes the third-person plural (also, the third-person singular feminine) suffix. Nevertheless, *=re* in either function has the same set of case forms in Table 8.

Both *=re* as the plural NPC and *=re* as the plural of *=riččo* are morphologically masculine and take the form of *=r-i* when the noun phrase is in the nominative or the genitive case and the form of *=r-i-ra* when the noun phrase is in the dative-locative case.

Figures 6 and 7 summarize the distributions of the use (the masculine gender marking) of the two types of *=re* for grammatical nominalizations. Figure 6 is very similar to Figure 2 (gender marking with the noun-phrase enclitic *=ta* (F)/*=ha* (M)), but differs from it in two respects. First, unlike the singular NPC *=ta* (F)/*=ha* (M), the plural NPC *=re* cannot attach to a cardinal numeral with the emphatic enclitic *=n*. Second, *=re* cannot be used with a complement clause.

Figure 6. Use of *=re* as the plural noun phrase clitic on five types of grammatical nominalizations

NP use:

|              |       |       |            |          |
|--------------|-------|-------|------------|----------|
| *CARD.NUM/ > | DEM > | GEN > | *ADJ >     | REL.CL/  |
| ORD.NUM      |       |       | (ADJ-iidi) | *COMP.CL |

Mod use:

|              |       |           |            |             |
|--------------|-------|-----------|------------|-------------|
| *CARD.NUM/ > | DEM > | OPT GEN > | *ADJ >     | OPT REL.CL/ |
| ORD.NUM      |       |           | (ADJ-iidi) | *COMP.CL    |

ADJ-iidi can attach to an adjective with the comparative/superlative suffix ADJ-*iids*

OPT: optional (The plural NPC is optionally used.)

\*: impossible (The plural NPC cannot attach to the structure.)

unmarked: possible (The plural NPC can attach to the structure.)

Figure 7. Use of *=re* as the plural of *=riččo* on five types of grammatical nominalizations

NP use:

|             |        |       |       |          |
|-------------|--------|-------|-------|----------|
| CARD.NUM/ > | *DEM > | GEN > | ADJ > | REL.CL/  |
| ORD.NUM     |        |       |       | *COMP.CL |

Mod use:

|              |        |        |        |          |
|--------------|--------|--------|--------|----------|
| *CARD.NUM/ > | *DEM > | *GEN > | *ADJ > | *REL.CL/ |
| *ORD.NUM     |        |        |        | *COMP.CL |

\*: impossible (*=re* as the plural of *=riččo* cannot attach to the structure.)

unmarked: possible (*=re* as the plural of *=riččo* can attach to the structure.)

The following descriptions deal with the two types of *=re* in parallel.



(82) saikki = r-i                                          t'arapp'eeyye  
third=NPC.PL-NOM.M.MOD      table.PL  
hiikk'-an-t-ino.  
become.broken-PASS-3SG.F/3PL-D.PERF.3  
'The third (set of) tables (definite) became broken.'

The enclitic *=re* as the plural of *=riččo* (in NP use) can attach to ordinal numerals.

NP use

(83) saikkí = r-i                      ba'-ø-ino.  
third=NP.C.PL-NOM.M.MOD    disappear-3SG.M-D.PRF.3  
'(The) third things (definite/indefinite) disappeared.'

(84) saikki = r-i  
third=NPC.PL-NOM.M.MOD  
hiikk'-am-ø-ino/\*hiikk'-an-t-ino.  
become.broken-PASS-3SG.M-D.PERF.3/become.broken-PASS-3SG.F/3PL-D.PERF.3  
'(The) third things (definite/indefinite) became broken.'

## [2] Demonstratives

Unlike *=re* as the plural of *=riččo*, the plural NPC *=re* is included in the plural demonstrative, which has both NP use and modification use.

NP use

(85) hakkú = r-i  
 that.M.NOM=NPCL.NOM.M.MOD  
 dag-g-ino/da-ø-ino.  
 come-3SG.F/3PL-D.PRF.3/come-3SG.M-D.PRF.3  
 ‘Those came.’ (*dag-q-ino*: gender unknown, *da-ø-ino*: all men)

Mod use

(86) hakkú = r-i                                mann-i/\*mann-u  
those=NP.C.PL-NOM.M.MOD      people-NOM.M.MOD/people-NOM.M  
dag-g-ino/da-ø-ino.  
come-3SG.F/3PL-D.PRF.3/come-3SG.M-D.PRF.3  
'Those people came.' (*dag-q-ino*: gender unknown, *da-ø-ino*: all men)

(87) hakkú=r-i                                          ooso  
those=NP.C.PL-NOM.M.MOD      children(NOM.F)  
dag-g-ino/\*da-ø-ino.  
come-3SG.F/3PL-D.PRF.3/come-3SG.M-D.PRF.3  
'Those children came.'



### [3] Genitive noun phrases

The plural NPC =*re* both in NP use and in modification use can attach to genitive noun phrases.

## NP use

- (88) ané=r-i danč-uúlle=ho.  
 1SG.GEN=NPC.PL-NOM.M.MOD good-PL=NPC.M.PRED  
 ‘Mine (animate/inanimate, definite) are good.’
- (89) hakkui manč-i = r-i/beett-i = r-i  
 that.M.GEN person-GEN.M.MOD=NPC.PL-NOM.M.MOD/  
 child-GEN.M.MOD=NPC.PL-NOM.M.MOD  
 ba’-’-ino/ba’-ø-ino.  
 disappear-3SG.F/3PL-D.PRF.3/disappear-3SG.M-D.PRF.3  
 ‘That man’s/boy’s (animate/inanimate, definite) disappeared.’ (*ba’-’-ino*: gender  
 unknown, *ba’-ø-ino*: all masculine)

## Mod use

- (90) hakkui                    manč-i = r-i  
 that.M.GEN            person-GEN.M.MOD=NPC.PL-NOM.M.MOD  
 k'aakk'-uull-u/\*k'aakk'-uull-i  
 baby-PL-NOM.M/baby-PL-NOM.M.MOD  
 ba'-'-ino/ba'-ø-ino.  
 disappear-3SG.F/3PL-D.PRF.3/disappear-3SG.M-D.PRF.3  
 'That man's babies disappeared.' (*ba'-'-ino*: gender unknown, *ba'-ø-ino*: all masculine)
- (91) hakkui                    beett-i = r-i  
 that.M.GEN            child-GEN.M.MOD=NPC.PL-NOM.M.MOD  
 rod-uuww-u/\*rod-uuww-i  
 sibling-PL-NOM.M/sibling-PL-NOM.M.MOD  
 ba'-'-ino/ba'-ø-ino.  
 disappear-3SG.F/3PL-D.PRF.3/disappear-3SG.M-D.PRF.3  
 'That boy's siblings (definite) disappeared.' (*ba'-'-ino*: gender unknown, *ba'-ø-ino*: all masculine)

The enclitic *=re* as the plural of *=riččo* (in NP use) can also attach to genitive noun phrases.

NP use

- (92) hakkui                      manč-i = r-i  
that.M.GEN              person-GEN.M.MOD=things-NOM.M.MOD  
ba'-ø-ino/\*ba'-'-ino.  
disappear-3SG.M-D.PRF.3/disappear-3SG.F/3PL-D.PRF.3  
'That boy's things (definite/indefinite) disappeared.'

- (93) isi = r-i  
3SG.M.GEN=things-NOM.M.MOD  
ba'-ø-ino/\*ba'-'-ino.  
disappear-3SG.M-D.PRF.3/disappear-3SG.F/3PL-D.PRF.3  
'His things (definite/indefinite) disappeared.'

#### [4] Adjectives

Like the singular NPC = *ta* (F)/ = *ha* (M), the plural NPC = *re* cannot attach to adjectives, though it can attach to an adjective with the comparative/superlative suffix ADJ-*iidi* 'more/most/relatively ...' either in NP use or in modification use.

NP use

- (94) jaw-iidi = r-i                                      ba'-'-ino.  
large-relatively=NPC.PL-NOM.M.MOD disappear-3SG.F/3PL-D.PRF.3  
'The larger/largest ones (animate/inanimate, definite) disappeared.'

Mod use

- (95) jaw-iidi = r-i  
large-relatively=NPC.PL-NOM.M.MOD  
k'aakk'uull-u/\*k'aakk'uull-i  
babies-NOM.M/babies-NOM.M.MOD  
ba'-'-ino/ba'-ø-ino.  
disappear-3SG.F/3PL-D.PRF.3/disappear-3SG.M-D.PRF.3  
'The larger/largest of the babies disappeared.' (*ba'-'-ino*: gender unknown, *ba'-ø-ino*: all masculine)

- (96) \*jaŋjab-b-u = r-i                                      k'aakk'uull-u/k'aakk'uull-i  
large-PL=NPC.PL-NOM.M.MOD babies-NOM.M/babies-NOM.M.MOD  
ba'-'-ino/ba'-ø-ino.  
disappear-3SG.F/3PL-D.PRF.3/disappear-3SG.M-D.PRF.3

The enclitic =*re* as the plural of =*riččo* (in NP use) can attach to adjectives including those with the comparative/superlative suffix ADJ-*iidi*.

NP use

- (97) *ǰaǰab-b-u = r-i* *ba'-ø-ino.*  
 large-PL-NOM.M=things-NOM.M.MOD disappear-3SG.M-D.PRF.3  
 ‘(The) large ones (inanimate, definite/indefinite) disappeared.’

- (98) *ǰaw-iidi = r-i* *ba'-ø-ino.*  
 large-relatively=things-NOM.M.MOD disappear-3SG.M-D.PRF.3  
 ‘(The) larger/largest ones (inanimate, definite/indefinite) disappeared.’

- (99) *ínsa bušá = re loos-s-é*  
 3PL.NOM bad.ACC=things(ACC) work-3PL-CVB  
*oso'I-an-t-u.*  
 laugh-PASS-3PL-R.PRF.3PL  
 ‘They did (the) bad things and were laughed at.’

## [5] Relative clauses

The plural NPC = *re* both in NP use and in modification use can attach to relative clauses. Like the use of the singular NPC = *ta* (F)/= *ha* (M), that of the plural NPC = *re* is optional in modification use (e.g., (101)–(103)), though it is obligatory in NP use (e.g., (100)). (It remains to be investigated whether a clause formed with the plural NPC = *re* in modification use is a restrictive or non-restrictive relative clause.)

NP use

- (100) *ub-b-ino = r-i*  
 fall-3SG.F/3PL-D.PRF.3 = NPC.PL-NOM.M.MOD  
*hiikk'-an-t-ino/\*hiikk'-am-ø-ino.*  
 break-PASS-3SG.F/3PL-D.PRF.3/break-PASS-3SG.M-D.PRF.3  
 ‘The ones who/that fell (animate/inanimate, definite) became broken.’

Mod use

- (101) *ub-b-ino = r-i/\*uw-ø-ino = r-i*  
 fall-3SG.F/3PL-D.PRF.3 = NPC.PL-NOM.M.MOD/  
 fall-3SG.M-D.PRF.3 = NPC.PL-NOM.M.MOD  
*mann-u hiikk'-an-t-ino/\*hiikk'-am-ø-ino.*  
 people-NOM.M break-PASS-3SG.F/3PL-D.PRF.3/break-PASS-3SG.M-D.PRF.3  
 ‘The people who fell became broken.’

- (102) *ani be'ro t'aad-oo-mm-o = r-i*  
 1SG.NOM yesterday meet-D.PRF.1SG-1SG-M=NPC.PL-NOM.M.MOD  
*ooso-nsa danč-uulle = te.*  
 children-3PL.POSS(NOM) good-PL=NPC.F.PRED  
 ‘Their children, who I (M) met yesterday, are good.’

- (103) [billaal-i-sí yanna-rá  
 scattering-NOM.M.MOD-3SG.M.POSS time-DATLOC.MOD  
 la'-á-ø-i  
 look-INF-3SG.F/3PL-ABLINS  
 no-o = r-í]  
 come.to.exist.D.PRF.3-LV=NPC.PL-NOM.M.MOD  
 [oóso] [godo'l-i-t-á-a  
 children(NOM) play-EP-3SG.F/3PL-INF-LV  
 no-o = r-í]  
 come.to.exist.D.PRF.3-LV=NPC.PL-NOM.M.MOD  
 [teennisé-te kowaasé godo'l-i-t-á-a  
 tennis-GEN.F ball(ACCOBL) play-EP-3SG.F/3PL-INF-LV  
 no-o = r-í]  
 come.to.exist.D.PRF.3-LV=NPC.PL-NOM.M.MOD  
 dag-g-é ka'eentinní, ...  
 come-3SG.F/3PL-CNN then  
 'The children, who were looking when the stone scattered the avocados (*lit.*, his (the  
 stone's) scattering), and who were playing, who were playing with a tennis ball, came,  
 and then ...' (Pear 6-(6))

NP use

## APPENDIX 2: Gender marking with the noun-phrase enclitic =te (F)/=ho (M) in the predicate

Figure 8. Gender marking on five types of grammatical nominalizations with the predicative noun-phrase enclitic =*te* (F)/ =*ho* (M)

NP use:

|             |       |       |       |          |
|-------------|-------|-------|-------|----------|
| CARD.NUM/ > | DEM > | GEN > | ADJ > | REL.CL/  |
| ORD.NUM     |       |       |       | *COMP.CL |

Mod use:

|              |        |        |        |          |
|--------------|--------|--------|--------|----------|
| *CARD.NUM/ > | *DEM > | *GEN > | *ADJ > | *REL.CL/ |
| *ORD.NUM     |        |        |        | *COMP.CL |

∗: impossible (= *re* as the plural of =*riččo* cannot attach to the grammatical nominalization.)

unmarked: possible (= *re* as the plural of =*riččo* can attach to the grammatical nominalization.)

There is one difference between the singular NPC in its predicative function and that in argument-forming function: unlike the latter, the former can attach to a noun. As in Table 8, when it attaches to a predicate nominal, it takes the form of =*te* (F)/ =*ho* (M) when it constitutes a predicate with a grammatical nominalization, except that it is =*ti* regardless of the gender of the subject noun phrase when the predicate nominal is a Modified common noun, a proper noun, a pronoun, or a demonstrative (see [2] below) in the accusative-oblique case (at least segmentally).

## [1] Numerals

### [1a] Cardinal numerals

NP use

|                                  |                     |                  |
|----------------------------------|---------------------|------------------|
| (105) dag-g-ino                  | mann-i              | sáse = ho.       |
| come-3SG.F/3PL-D.PRF.3           | person.PL-NOM.M.MOD | three=NPC.M.PRED |
| ‘The people who came are three.’ |                     |                  |

|                               |                     |                  |
|-------------------------------|---------------------|------------------|
| (106) da-ø-ino                | mann-i              | sáse = ho.       |
| come-3SG.M-D.PRF.3            | person.PL-NOM.M.MOD | three=NPC.M.PRED |
| ‘The men who came are three.’ |                     |                  |

### [1b] Ordinal numerals

NP use

|                           |                  |                   |
|---------------------------|------------------|-------------------|
| (107) hakku               | manč-i           | laínki = ho.      |
| that.M.NOM                | person-NOM.M.MOD | second=NPC.M.PRED |
| ‘That man is the second.’ |                  |                   |

## [2] Demonstratives

NP use

- (108) ané            rod-ii                    hakkónne-e = ti.  
1SG.GEN sibling-NOM.M.MOD that.M-LV=NP.C.PRED  
'My brother is that one.'

- (109) ané            rod-uuw-i                    hakkóre-e = ti.  
1SG.GEN sibling-PL-NOM.M.MOD that.PL-LV=NP.C.PRED  
'My siblings are those ones.'

## [3] Genitive noun phrases

NP use

- (110) ise                    dangur-í = te.  
3SG.F.NOM Dangura-GEN.M.PROP=NP.C.F.PRED  
'She is Dangura's (e.g., daughter).'

## [4] Adjectives

NP use

- (111) ise                    búša = te.  
3SG.F.NOM bad=NP.C.F.PRED  
'She is bad.'

## [5] Relative clauses

NP use

- (112) íse                    áni            bé'ro            la'-oó-mm-o = te.  
3SG.F.NOM 1SG.NOM yesterday see-D.PRF.1-1SG-M=NP.C.F.PRED  
'She is the one (F) who I (M) saw yesterday.'

# Classifier marking in Assamese

Gitanjali Bez

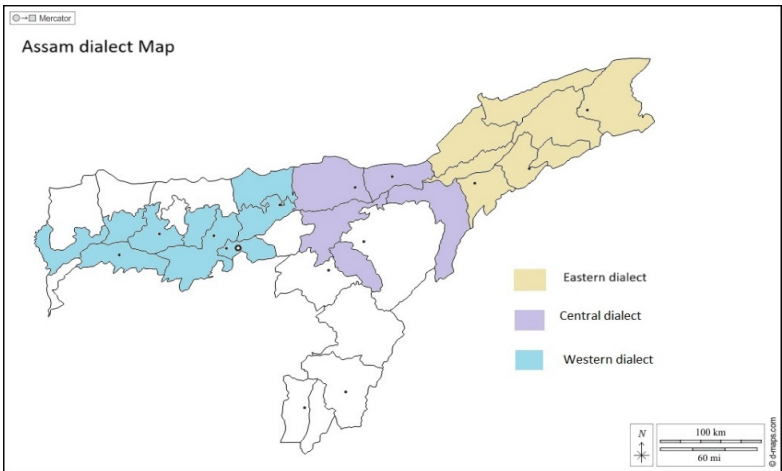
## ABSTRACT

The present study describes the use of classifiers along the structural and functional dimensions of the hierarchy (i.e., NUM > DEM > GEN > ADJ > V-based; NP use > modification use) proposed by Shibatani (2017, 2023). The study demonstrates that usage of classifiers in Assamese aligns with Shibatani’s hierarchy, strictly adhering to it.

## 1. INTRODUCTION

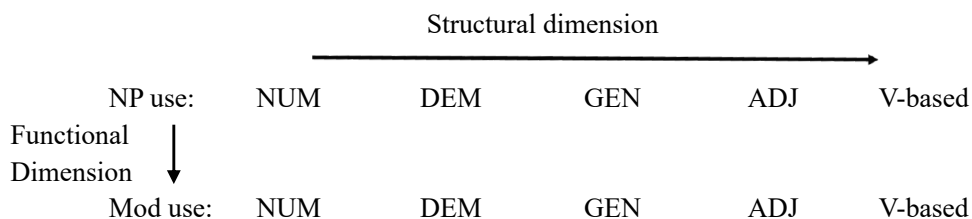
Assamese belongs to the eastern group of the New Indo-Aryan (NIA) family. It is primarily spoken in the state of Assam by 48% of the state’s population (Census of India, 2011). Assamese is surrounded by numerous Tibeto-Burman languages, including Boro, Mishing, Deuri, and Tiwa, as well as a few Indo-Aryan languages, including Bengali and Nepali. It has a significant amount of regional variation, which is mainly classified into two broad groups: Eastern Assamese and Western Assamese (Kakati 1941). Goswami (1982) later proposed a regrouping, adding an intermediate dialect group called Central Assamese. The Eastern Assamese dialect is spoken in the districts of Sivasagar, Jorhat, Dibrugarh, Tinsukia, Majuli, Lakhimpur, and Dhemaji, while Western Assamese is spoken over a large area extending from just east of Guwahati in the south, the Darrang district in the north, and extending westward to the districts of Goalpara and Dhuburi. The Central or Intermediate dialect occupies the area between these two regions (Goswami and Tamuli 2003). Among these three broad dialect groups, Western Assamese is characterized by a rich variety of regional variations.

Figure 1: Dialect map of Assam (Source: d-maps.com)



Unlike western Indo-Aryan languages like Nepali, Maithili, Hindi, Marathi, Gujarati, etc., nouns in Assamese do not have grammatical gender (see Masica 1991: 219–223). However, the language employs a dozen classifiers, making it unique in comparison to other NIA languages. This study reports how these classifiers are used within the structural and functional dimensions of the classifier/gender marking hierarchy proposed by Shibatani (2023: 35).

Figure 2: Hierarchy constraining classifier/gender marking (Shibatani: 2023: 35)



## 2. METHODOLOGY

This report is based on primary data. A two-week fieldwork study was conducted in the north-west and south-west parts of Assam in February 2023. A rigorous elicitation method was used to collect data from a total of 29 informants who speak Western Assamese, to examine whether they differ in their use of classifier marking. The informants who participated in the elicitation sessions varied in terms of age, gender, and educational background. Some were monolingual, while others could speak English and Hindi to some extent. Assamese is the mother tongue of all informants. The fieldwork was conducted in Sualkuchi, Udalguri, Mongoldoi, Barpetta, Bongaigaon, Dhubri, and Goalpara. Data from the eastern and central varieties of the language were also collected from the outset of the project. Since the researcher is a speaker of the eastern variety, both introspective (self-elicited) data and elicited data were collected for the present study.

## 3. ANALYSIS OF DATA

The data collected so far reveals that classifiers in Assamese are employed not only with numerals, but also with other structures such as demonstratives, genitives, adjectives, and verb participles. The data also reveal that the use of classifiers with these structures varies in their functions, namely, NP use and modification use, across dialects, as discussed in the following sub-sections. To demonstrate the distribution of classifiers, two most frequently used classifiers in Assamese, *-zon* and *-kʰon*, are considered here. The former is used to classify male human referents in a respectful sense, while the latter is used to classify flat objects.

### 3.1. Obligatory classifier marking

#### 3.1.1. Numerals

Nearly all classifiers are obligatorily attached to numerals in both NP and Modification functions, as illustrated by the classifiers *-zon* and *-kʰon* in examples (1)–(4). The numeral *ek* ‘one’ becomes *e* when a classifier is added to *ek*.



NP use:

- (1) *gʰɔr-ɔt*      [*e-zɔn*]      *as-e*  
house-LOC    one-CLF.HUM    be-3  
‘There is one (person) at home.’

(Eastern dialect)

- (2) *tebul-ɔt*      [*e-kʰɔn*]      *as-e*  
table-LOC    one-CLF.FLAT    be-3  
‘There is one (book) on the table.’

(Eastern dialect)

Modification use:

- (3) *gʰɔr-ɔt*      [*{e-zɔn}*]      *manuh*]      *as-e*  
house-LOC    one-CLF.HUM    man      be-3  
‘There is one person at home.’

(Eastern dialect)

- (4) *tebul-ɔt*      [*{e-kʰɔn}*]      *kitap*]      *as-e*  
table-LOC    one-CLF.FLAT    book      be-3  
‘There is one book on the table.’

(Eastern dialect)

It should be noted that in all dialects of Assamese that we worked on, classifiers are obligatorily attached to numerals. Furthermore, when a noun phrase contains a numeral, the classifier always forms a constituent with the numeral, not with other elements of the phrase, as illustrated below (Compare (5) with (1)).

- (5) \**gʰɔr-ɔt*      [*{e (ek)}*]      *manuh]-zɔn*      *as-e*  
house-LOC    one      man-CLF.HUM    be-3  
‘There is one person at home.’

### 3.1.2. Demonstratives

The distribution of classifier in the case of demonstrative pronouns is different from numerals. While classifiers are obligatorily attached to demonstratives in NP use across dialects, they are optional in modification use.

#### 3.1.2.1. Classifiers on demonstratives in NP use

The classifier is obligatorily attached to the demonstratives in the NP use across the dialects, as shown below.

- (6) [ei-kʰɔn/\*ei]      mɔ-r  
 this-CLF      1SG-GEN  
 ‘This one is mine.’

(Mongoldoi region)

- (7) [i-kʰen/\*ei]      mɔ-r.  
 this-CLF      1SG-GEN  
 ‘This one is mine.’

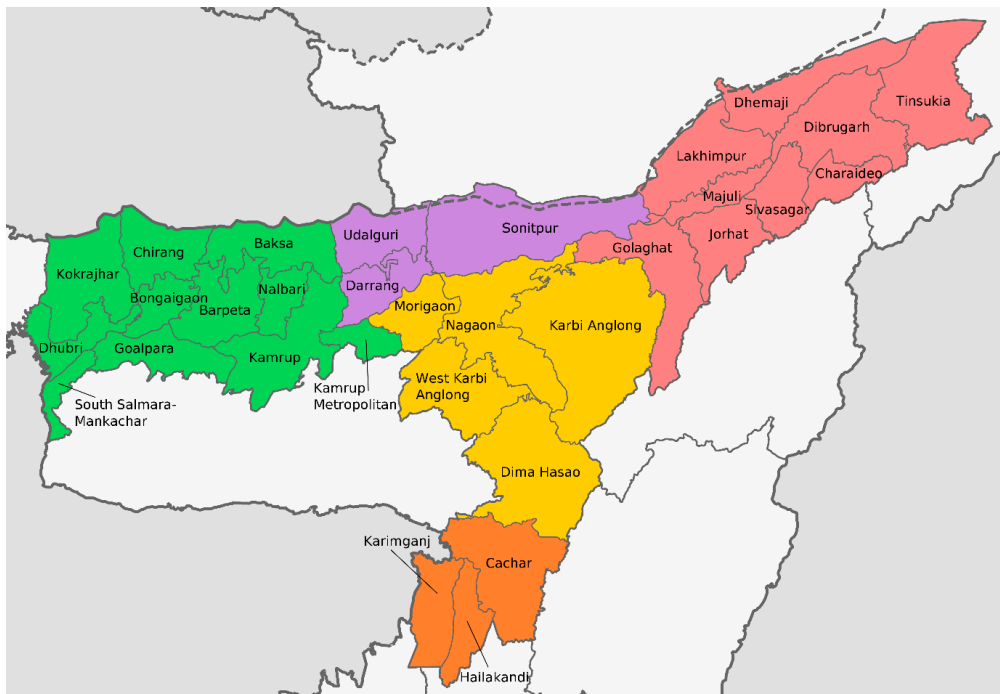
(Barpeta region)

These examples show that the classifier on the demonstrative is obligatory in NP use in the language.

### 3.1.2.2. Classifiers on demonstratives in modification use

Cross-dialectal variation is observed in the usage of classifiers with demonstratives in modification use. Two patterns are observed: obligatory use of classifiers and optional use of classifiers. The former is found in the varieties spoken in the Mongoldoi and Udalguri regions, whereas the latter is found in the varieties spoken in Sualkuchi, Nalbari, Barpeta, Goalpara, Dhubri, and the eastern parts of Assam (see Figure 3 below).

Figure 3: District map of Assam (Source: Wikipedia)



The two patterns are illustrated in (8) through (11) below.

Obligatory classifier marking with demonstratives in modification use:

- (8) a. [*{ei-kʰɔn}*      *kitab*]      *lag-e*  
           this-CLF.FLAT    book      need-3  
           ‘I need this book.’
- b. \**[{ei}*                      *kitab]-kʰɔn*              *lag-e*  
           this-CLF.FLAT      book-CLF.FLAT      need-3  
           ‘I need this book.’
- (Mongoldoi region)

- (9) a. [*{i-kʰan}*              *kitab*]      *mo-r*  
           this-CLF.FLAT    book      1SG-GEN  
           ‘This book is mine.’
- b. \**[{ei}*                      *kitab]-kʰan*              *mo-r*  
           this                  book-CLF.FLAT      1SG-GEN  
           ‘This book is mine.’
- (Udalguri region)

Optional classifier marking with demonstratives in modification use:

The pattern of optional marking varies across dialects. Two patterns are observed: (i) attaching classifier to demonstrative pronoun or (ii) attaching it to the NP containing demonstrative pronoun. In Barpeta, Nalbari, and the eastern parts of Assam, the pattern of attaching classifiers to demonstratives is preferred over pattern of attaching classifier to the NP containing demonstrative pronoun, as shown in the following examples.

- (10) a. [*{i-kʰen/\*ei}*      *kitab*]      *mo-r*  
           this-CLF          book      1SG-GEN  
           ‘This book is mine.’ (Preferred one)
- (Barpeta region)
- b. [*ei kitab]-kʰen*              *mo-r*  
           this    book-CLF.FLAT      1SG-GEN  
           ‘This book is mine.’
- (Barpeta region)
- (11) a. [*{e-kʰan/\*ei}*              *kitab*]      *mo-r*  
           this-CLF.FLAT      book      1SG-GEN  
           ‘This book is mine.’ (Preferred one)
- (Nalbari region)

- b. *[ei kitap]-kʰen* *mo-r*  
 this book-CLF.FLAT 1SG-GEN  
 ‘This book is mine.’  
 (Nalbari region)

The use of classifiers on the NP is not a commonly used pattern in Barpeta, Nalbari, and the eastern parts of Assam. According to our consultants, while this pattern is not impossible, it appears strange. In contrast, classifier on the NP is the more preferred pattern over classifier on demonstratives pattern in the varieties spoken in Goalpara and Bongaigaon regions. According to our consultants, classifier on the demonstrative pattern occurs when a speaker wants to select one item from multiple options. Below are the representative examples.

- (12) a. [*ei-kʰan/\*ei*      *kʰata/boi*]      *mo-r*  
           this-CLF            book/book      1SG-GEN  
           ‘This notebook is mine.’ (Not impossible, but strange)
- b. [*ei*                      *kʰata*]-*kʰan*                      *mo-r*  
           this-CLF            book-CLF.FLAT      1SG-GEN  
           ‘This notebook is mine.’ (Preferred pattern)
- (Goalpara region)

### 3.2. Optional classifier marking

Starting from genitives and continuing through structures further right in the hierarchy, the use of classifiers in both NP and modification contexts becomes optional. This optionality appears to be triggered by the function of specificity. Classifiers become obligatory with genitives, adjectives, and V-based participle structures in both NP and modification uses when the entity denoted by these structures is specific and definite. For non-specific and indefinite referents, classifiers are not allowed to occur with these structures, as shown below.

Context: Which saree will you wear this evening?

- (13) [ma-r]                      pind<sup>h</sup>-im  
mother-GEN                  wear-FUT.1  
'I will wear my mother's.'
- (14) [ma-r                      sari]                  pind<sup>h</sup>-im  
mother-GEN                  saree                  wear-FUT.1  
'I will wear my mother's saree.'
- (Eastern region)

These constructions are used when there are no predefined set of options or choices. If there are an identifiable set and the speaker is asked to choose any one from that known or specific set, then classifiers are obligatory on these structures.

Context: Which saree will you wear this evening?

- (15) [*ma-r-kʰɔn*]                      *pindʰ-im*  
 Mother-GEN-CLF.FLAT      wear-FUT.1  
 ‘I will wear my mother’s one.’

- (16) [{*ma-r*}              *sari*]-*kʰɔn*              *pindʰ-im*  
 Mother-GEN      saree-CLF.FLAT      wear-FUT.1  
 ‘I will wear my mother’s saree.’

(Eastern region)

### 3.2.1. Genitives

Classifiers on genitives are less developed as compared to demonstratives and numerals, both in NP use and modification use. As mentioned above, the presence or absence of classifiers on them in both NP use and modification use is determined by definiteness and specificity of the referent.

#### 3.2.1.1. Classifiers on the genitives in NP use

Classifiers are obligatorily applied to genitives when the NP is definite. Indefinite NPs are not marked by classifiers.

- (17)a. [*to-r-kʰen*]                      *pinh-im*  
 2INF.SG-GEN-CLF.FLAT              wear-FUT.1  
 ‘I will wear yours (saree).’ (Definite)

- b. [*to-r*]                      *pinh-im*  
 2INF.SG-GEN              wear-FUT.1  
 ‘I will wear any (saree) of yours.’ (Indefinite)

(Barpeta region)

#### 3.2.1.2. Classifiers on the genitives in modification use

Classifiers can be applied to genitives only in the varieties spoken in Barpeta and the eastern parts of Assam, and only when a speaker wants to select one item among multiple options.

- (18) [{*toma-r-kʰɔn*}                      *hari*]              *pindʰ-im*  
 2FAM.SG-GEN-CLF.FLAT              saree              wear-FUT.1  
 ‘I will wear your sari.’

(Eastern region)

This pattern is not found in the varieties spoken in other regions, such as Nalbari, Bongaigaon, Goalpara, Dhubri, etc.

- (19) \*[{*tuma-r-k<sup>h</sup>an*} *hari*] *pind<sup>h</sup>-im*  
 2FAM.SG-GEN-CLF.FLAT saree wear-FUT.1  
 ‘I will wear your sari.’  
 (Nalbari region)

The only acceptable pattern in these regions is the use of classifiers on the NP as a whole.

- (20) [*{tuma-r}*      *ħari*]-*k<sup>h</sup>an*      *pind<sup>h</sup>-im*  
2FAM-GEN      saree-CLF.FLAT      wear-FUT.1  
'I will wear your saree.' (Only pattern)  
(Nalbari region)

### 3.2.2. Adjectives

Classifiers do not apply to adjectives in modification use across all dialects. However, in NP use, they mark definite NPs in all dialects.

NP use:

- (21) *məi* [bəga-kʰan] *pindʰ-im*  
 1SG white-CLF.FLAT wear-FUT.1  
 ‘I will wear the white one.’

Modification use:

- (22) a. \**mɔi*    [*{\*bɔga-kʰan}*    *ħari*]    *pindʰ-im*  
 1SG    white-CLF.FLAT    saree    wear-FUT.1  
 ‘I will wear the white saree.’
- b. *mɔi*    [*bɔga*    *ħari*]-*kʰan*    *pindʰ-im*  
 1SG    white    saree-CLF.FLAT    wear-FUT.1  
 ‘I will wear the white saree.’

### 3.2.3. V-based structures

Like adjectives, classifiers do not apply to V-based structures in modification use in all dialects. However, they do mark V-based structures when used as definite NPs.

NP use:

- (23) [almari-*i*            *tʰak-a-to*]            *pin<sup>dh</sup>im*  
 closet-LOC            stay-NMLZ-CLF            wear-1.FUT  
 ‘I will wear the one in the closet.’  
 (Sualkuchi region)

Modification use:

- (24) a. \*[{*almari-t*      \**tʰak-a-tv*}      *sola*]      *pindʰim*  
          closet-LOC      stay-NMLZ-CLF      shirt      wear-FUT.1  
          ‘I will wear the shirt in the closet.’

- b. [{*almari-t*      *tʰak-a*}      *sola*]-*tv*      *pindʰim*  
          closet-LOC      stay-NMLZ      shirt-CLF      wear-FUT.1  
          ‘I will wear the shirt in the closet.’

(Sualkuchi region)

### 3.3. Summary

From the discussion in the foregoing sections, it is clear that classifiers are obligatorily attached to numerals in both NP use and modification use across dialects. Like numerals, classifiers are obligatorily attached to demonstratives in NP use in all dialects. However, unlike numerals, classifiers are optional in modification use and exhibit two patterns of distribution. They are obligatory on demonstratives in the varieties spoken in the Mongoldoi and Udalguri regions, but optional in other regions. Like numerals and demonstratives, classifiers are obligatory on genitives in NP use, but optional in modification use in the Barpeta and eastern regions of Assam. Classifiers can be applied to genitives in modification use when a speaker wants to select one item among several. Otherwise, classifiers tend to mark the whole NP. Classifiers are obligatory on adjectives and V-based participle structures when they serve NP functions, similar to numerals, demonstratives, and genitives. However, unlike these structures, classifiers cannot be applied to adjectives and V-based participle structures in modification use.

Table 1. The structural and functional distribution of classifier marking in Assamese spoken in different regions (Y= permitted, N= not permitted, Y<sup>OPT</sup>= optionally permitted)

| Varieties spoken in different regions | NP use<br>Mod use | NUM    | DEM                   | GEN                   | ADJ    | V-based structure |
|---------------------------------------|-------------------|--------|-----------------------|-----------------------|--------|-------------------|
| Eastern parts                         | NP use<br>Mod use | Y<br>Y | Y<br>Y <sup>OPT</sup> | Y<br>Y <sup>OPT</sup> | Y<br>N | Y<br>N            |
| Barpeta                               | NP use<br>Mod use | Y<br>Y | Y<br>Y <sup>OPT</sup> | Y<br>Y <sup>OPT</sup> | Y<br>N | Y<br>N            |
| Sualkuchi                             | NP use<br>Mod use | Y<br>Y | Y<br>Y <sup>OPT</sup> | Y<br>N                | Y<br>N | Y<br>N            |
| Nalbari                               | NP use<br>Mod use | Y<br>Y | Y<br>Y <sup>OPT</sup> | Y<br>N                | Y<br>N | Y<br>N            |
| Goalpara                              | NP use<br>Mod use | Y<br>Y | Y<br>Y <sup>OPT</sup> | Y<br>N                | Y<br>N | Y<br>N            |
| Dhuburi                               | NP use<br>Mod use | Y<br>Y | Y<br>Y <sup>OPT</sup> | Y<br>N                | Y<br>N | Y<br>N            |
| Udalguri                              | NP use<br>Mod use | Y<br>Y | Y<br>Y                | Y<br>N                | Y<br>N | Y<br>N            |
| Mongoldoi                             | NP use<br>Mod use | Y<br>Y | Y<br>Y                | Y<br>N                | Y<br>N | Y<br>N            |
| Bongaigaon                            | NP use<br>Mod use | Y<br>Y | Y<br>Y <sup>OPT</sup> | Y<br>N                | Y<br>N | Y<br>N            |

#### 4. CONCLUSION

The distribution of classifiers along the structural dimension across dialects of Assamese reflects what Shibatani and Shigeno (2013) predict regarding the development and distribution of classifier marking over various structures in the hierarchy, namely marking originates in NP use and then spreads to modification use. Classifiers are obligatorily attached to numerals in both NP and modification uses, but for all other structures, they are obligatory in a few cases while optional or impossible in other cases, as shown in Table 1. This asymmetry in the distribution of classifiers in Assamese suggests that numerals are more susceptible to classifier marking than any other structures further to the right in the hierarchy. Furthermore, the fact that in modification use classifiers can be applied to numerals, demonstratives, and genitives, but not permitted in the case of adjectives and V-based nominalization in Assamese suggests that N-based nominalizations (numerals, demonstratives, and genitives) are more developed than adjectives and V-based nominalizations (verb-based structures), with respect to the use of classifier marking.

#### ABBREVIATIONS

|     |               |      |             |
|-----|---------------|------|-------------|
| 1   | First Person  | GEN  | Genitive    |
| 2   | Second Person | HUM  | Human       |
| 3   | Third Person  | INF  | Inferior    |
| CLF | Classifier    | LOC  | Locative    |
| FAM | Familiar      | NMLZ | Nominalizer |
| FUT | Future        | SG   | Singular    |



## ACKNOWLEDGMENTS

The present study has been supported by the project “Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective” (PI: Prashant Pardeshi) of National Institute for Japanese Language and Linguistics.

## REFERENCES

- Goswami, Golockchandra. 1982. *Structure of Assamese*. Guwahati: Gauhati University Publication Department.
- Goswami, G. C., and Jyotiprakash Tamuli, J. 2003. Assamese. In George Cardona and Dhanesh Jain. (eds.) *The Indo-Aryan languages*, 391–443. London/New York: Routledge.
- Kakati, Banikanta. 1941. *Assamese: Its formation and development*. Guwahati: LBS Publication, Assam, India.
- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Shibatani, Masayoshi, and Hiromi Shigeno. 2013. Amami nominalizations. *International Journal of Okinawan Studies* 4.1, 107–139.
- Shibatani, Masayoshi. 2019. What is nominalization? Towards the theoretical foundations of nominalization. In Roberto Zariquiey, Masayoshi Shibatani, and David W. Fleck. (eds.) *Nominalization in languages of the Americas*, 15–168. Amsterdam & Philadelphia: John Benjamins.
- Shibatani, Masayoshi. 2021. Rentai shūshoku no bunpō: Ruibetsushi to bunpōsei o chūshin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki* (Nominalization theory and linguistic analysis), 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.

# Distribution of classifier and gender marking in Southern Marathi as compared to Pune Marathi

Manasi Kelkar<sup>1</sup> and Prashant Pardeshi<sup>2</sup>

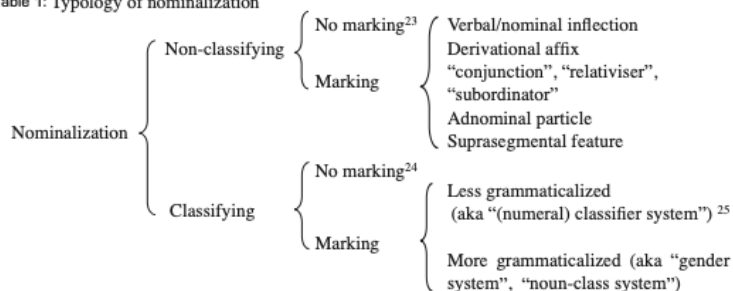
## ABSTRACT

The present study offers a preliminary report on the distribution of the classifier and gender markers in Southern Marathi (an Indo-Aryan language spoken in Tamil Nadu state, India) based on the data collected during our fieldwork conducted in Chennai, Krishnagiri, and Thanjavur districts in Tamil Nadu state in February 2025. Adopting the framework called “dynamic functional typology” proposed in Shibatani (2023) and following the hierarchy (NUM > DEM > GEN > ADJ > V-based) proposed therein (op. cit. 35), we describe the distribution of the classifier and gender markers in Southern Marathi and compare it with the variety of Marathi spoken in Pune, Maharashtra state, India. As for the distribution of the sole classifier in Southern Marathi, it can be attached to numerals only, with a few deviations such as the numeral ‘one’ to which the classifier cannot be attached and special form of numerals ‘two’ to ‘four’ (referred to as the G-form) to which the classifier can be unequivocally attached in the noun-phrase use, but not in modification use, where interspeaker variation is observed. With regard to gender marking, a gender distinction is either marked or neutralized (under specific conditions) in all structural slots on Shibatani’s hierarchy except in genitive noun phrases, where gender distinction is completely lost, in both noun-phrase use and modification use. In sum, while the distribution of the classifier supports the prediction of Shibatani’s hierarchy, the distribution of gender marking deviates from it.

## 1. INTRODUCTION

Shibatani (2023: 21) proposes a typology of nominalization as shown in the figure below and argues that classifier/gender marking are a “classifying” type of nominalization that not only nominalizes the structure in question but also classifies its denotation.

Table 1: Typology of nominalization



<sup>1</sup> Independent Researcher. Email: kelkar.manasi710@gmail.com

<sup>2</sup> National Institute for Japanese Language and Linguistics, Tokyo, Japan. Email: prashant@ninjal.ac.jp

In this paper, we will focus on “classifying” type of nominalizations which are marked by classifier and/or gender marking in the typology of nominalization schematically shown above and offer a preliminary report on the distribution of the classifier and gender markers in Southern Marathi (SM), an Indo-Aryan language spoken in Tamil Nadu state, India, based on the data collected in our fieldwork conducted in three locations: (i) Chennai (in the northern district), (ii) Krishnagiri (in the district at the junction of Karnataka, Andhra Pradesh and Tamil Nadu states) and (iii) Thanjavur (located in the Central district) in the state of Tamil Nadu (see the maps below for details) from February 5th to 10th, 2025. Adopting the framework called “dynamic functional typology” proposed in Shibatani (2023) and following the hierarchy (NUM > DEM > GEN > ADJ > V-based) proposed therein (op. cit. 35), we describe the distribution of the classifier and gender markers in Southern Marathi and compare it with the variety of Marathi spoken in Pune, Maharashtra state, India, Pune Marathi (PM). As for the distribution of the sole classifier in SM, it can be attached to numerals (henceforth, NUM) only, with a few deviations such as the numeral ‘one’ to which the classifier (henceforth, CLF) cannot be attached and special forms of the numerals ‘two’ to ‘four’ (referred to as G-form) to which CLF can be unequivocally attached in NP use, but not in the modification use where interspeaker variation is observed. Regarding gender marking, gender distinction is either marked or neutralized (under specific conditions) in case of all structural slots on Shibatani’s hierarchy except genitive noun phrases (henceforth, GEN), where gender distinction is completely lost, both in NP use and in Mod use. In sum, while the distribution of CLF supports the prediction of Shibatani’s hierarchy, the distribution of gender marking defies it.



Map 1: Tamil Nadu state in India.

Photo credit: CC-by-sa PlaneMad/Wikimedia

[https://en.m.wikipedia.org/wiki/File:India\\_Tamil\\_Nadu\\_locator\\_map.svg](https://en.m.wikipedia.org/wiki/File:India_Tamil_Nadu_locator_map.svg)



Map 2: District map of Tamil Nadu state in India

Source: By C1MM work based on Chumwa - Derivative of Tamil Nadu location map.svg,  
CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=91897514>

A few words about the term “Southern Marathi” that we use in this report are in order. By Southern Marathi we mean the variety of Marathi (Indo-Aryan) spoken in South India, which includes the states of Andhra Pradesh, Karnataka, Kerala, Tamil Nadu and Telangana, where Dravidian languages are spoken. Southern Marathi is thus an Indo-Aryan language spoken in a Dravidian-language-speaking area. In the literature it is commonly known as Thanjavur Marathi, named after the city of Thanjavur (also spelt as Tanjore), the headquarters of the district of Thanjavur in the South Indian state of Tamil Nadu. However, this name is misleading since this variety is spoken by speakers spread throughout South India.

Southern Marathi is a result of the migration of Marathi speakers from their homeland, viz. Maharashtra state, to various states in South India. This migration primarily occurred under the reign of the Maratha Empire, which was founded by Venkoji (also known as Ekoji I), the half-brother of Chatrapati Shivaji Maharaj, who was the founder of the Maratha Empire in Maharashtra. In the late 1600s, Venkoji, then residing in Bangalore, a prominent center in southern India, moved eastward and established the Thanjavur Maratha Kingdom (also referred to as the Thanjavur Bhonsle dynasty) in 1674 in the region corresponding to present-day Tamil Nadu. This migration led to the settlement of Maratha nobles, soldiers, administrators, and artisans in the Cauvery Delta region. Thanjavur was divided into five sub-regions (*Subha*) viz. Tiruvarur, Kumbakonam, Mayladadurai (also known as Mayuram, Mayavaram), Mannargudi, and Pudukkottai. Marathi speakers today are spread throughout these regions. Further, in the beginning of 1679, Chatrapati Shivaji conquered Kolar, Bangalore, Hoskote, Sira in the present-day Karnataka state and Balapoor in Telangana state. He brought with him 24,000 household cavalry, 25,000 hoplites, 10,000 infantries and also Brahmins with immense administrative and diplomatic capabilities. Most of these forces were left behind in the

conquered territory to control and administer new conquests. Marathi speakers in these regions are the descendants of these warriors, administrators and priests. Though Thanjavur was the cultural and political centre for Marathas for a very long time in the past, in the present times Marathi speakers have moved out of Thanjavur and settled in various parts of South India. To accommodate this wider geographical spread, we prefer the nomenclature Southern Marathi over Thanjavur Marathi.

Since the focus of this paper is on nominalizations, it is necessary to take a look at the characteristic grammatical features of nouns in Marathi spoken in its homeland, namely Maharashtra state with Pune variety (PM) as its representative and that spoken in Tamil Nadu state in Southern India (SM).

Nouns in the Pune variety (PM) have three characteristic features, viz. (i) gender, (ii) number and (iii) case (cf. Dhongde and Wali 2009: 39-47). Each noun is typically classified into one of the three genders, viz. masculine, feminine and neuter, irrespective of its animacy (e.g., inanimate noun like *kap* ‘cup’ is masculine, while *ḍāmbār* ‘tar’ is neuter) or natural gender (e.g., *pāl* ‘house lizard’ is feminine, while *ḥital* ‘antelope’ is neuter). Further, nouns in PM inflect for number (singular and plural) and case (direct case [morphologically unmarked or zero marked form] and oblique case [a derived form to which case markers and postpositions are added]). Furthermore, like nouns, declining adjectives in PM also change in their form according to the gender and number of their referent. Definiteness/specificity can be overtly shown by the use of demonstrative pronouns.

In sharp contrast to PM, there is no grammatical gender assignment in Southern Marathi. Gender assignment of nouns in SM is similar to that of Tamil (or Dravidian languages), which is the dominant language spoken in Tamil Nadu state, with which Southern Marathi has a prolonged and extensive contact of over four centuries. Like Tamil, nouns in SM are divided in two groups: human (rational) vs. non-human (irrational). Nouns in the human class have natural gender, viz. masculine and feminine, whereas all non-human (irrational) nouns are assigned neuter gender. *dādigā* and *bāyko* are lexical items meaning ‘male’ and ‘female’. Although there are separate terms referring to ‘son’ (*lōnk*) and ‘daughter’ (*lēnk*), most of the time, Southern Marathi speakers use compounds such as *dādigā lēnkru* (‘male child’) and *bāyko lēnkru* (‘female child’) to refer to son and daughter. The same is the case with other non-human animate nouns as well: *dādigā kutra* (‘dog’) vs. *bāyko kutra* (‘female dog’), *dādigā māṇjārā* (‘male cat’) vs. *bāyko māṇjārā* (‘female cat’) are used because the gender-equivalent vocabulary is rarely created by derivational process. This pattern seems to have emerged as direct translation of Tamil equivalent terms. Though Tamil has lexical gendered nouns for *pəyyən* (‘son’) and *ponnə* (‘daughter’), compounds such as *ān kuḷəṇḍai* (‘male child’) and *pen kuḷəṇḍai* (‘female child’) are preferred in some contexts. For non-human nouns, similar compounds using *ān* (‘male’) and *pen* (‘female’) as modifiers are used: *ān nāyi* (‘male dog’) vs. *pen nāyi* (‘female dog’), *ān pūṇai* (‘male cat’) vs. *pen pūṇai* (‘female cat’) etc. Further, nouns in Southern Marathi variety inflect for number (singular and plural) and case (direct case [morphologically unmarked or zero marked form] and oblique case [a derived form to which case markers and postpositions are added]). Variation in the use of the oblique case is notably observed amongst speakers. Furthermore, like nouns, declining adjectives in Southern Marathi

also change their form according to the gender and number of their referent. Definiteness/specificity can be overtly marked by demonstrative pronouns.

## 2. OVERVIEW OF CLASSIFIER AND GENDER MARKING AND GRAMMATICAL NOMINALIZATIONS IN SOUTHERN MARATHI

### 2.1 Classifier and gender marking in Southern Marathi

Dixon (1982) suggests that the two noun classification systems, viz. classifiers and gender, are distinct categories and they do not coexist in the same language. However, Corbet (2013) has shown that there are a few languages where both co-exist. PM is one such language. As mentioned earlier, unlike PM, nouns in SM do not have grammatical gender assignment. Gender assignment of nouns in SM is similar to that of Tamil where nouns are divided in two groups: human (rational) vs. non-human (irrational). Nouns in human class have natural gender, viz. masculine and feminine, whereas all non-human class nouns are assigned neuter gender.

Regarding “sortal” classifiers, SM, like PM (and Goan Konkani) has only one classifier, which is used for classifying human. There are a few mensural classifiers (aka measure words) in Southern Marathi such as *ek kuṇḍā/maḍka/ghāgar pāṇi* (one pot of water), *don gučča/muṭa durwā* (two bunches of sacred grass), *ek bhāga phulā* (one share/portion of flowers), *ek ḍhiga/gāṇṭoḍa lākuḍ* (one pile/bundle of wood), etc. which will not be discussed in this paper.

### 2.2 Grammatical nominalizations in Southern Marathi

Below we provide forms of numerals, demonstratives, genitives, adjectives and the so-called relative clauses (which we will call as V-based nominalizations following Shibatani (2023)), from the Thanjavur variety of Southern Marathi.

#### [1] Numerals

##### [1a] Cardinal numerals

Cardinal numerals do not have gender distinction. But while referring to human beings, only the numeral ‘one’ shows a three-way gender distinction: *ek.l-{\ā/i/e}* one.L-{\MSG/FSG/NSG}.

Table 1. Cardinal numerals in Southern Marathi

| Cardinal numerals | Form  | Cardinal numerals | Form |
|-------------------|-------|-------------------|------|
| 1                 | eka   | 6                 | sahā |
| 2                 | dona  | 7                 | sāt  |
| 3                 | tina  | 8                 | āṭh  |
| 4                 | čāra  | 9                 | no   |
| 5                 | pāṇča | 10                | dahā |

### [1b] Ordinal numerals

Ordinal numerals have a three-way gender distinction: masculine (M)/feminine (F)/neuter (N).

Table 2. Ordinal numerals in Southern Marathi

| Ordinal number | Form (M/F/N)     | Ordinal number | Form (M/F/N)     |
|----------------|------------------|----------------|------------------|
| 1st            | payl- {ā/i/e~a}  | 6th            | sahāw- {ā/i/e~a} |
| 2nd            | dusr- {ā/i/e~a}  | 7th            | sātw- {ā/i/e~a}  |
| 3rd            | tisr- {ā/i/e~a}  | 8th            | āṭhw- {ā/i/e~a}  |
| 4th            | čauth- {ā/i/e~a} | 9th            | nauw- {ā/i/e~a}  |
| 5th            | pāṇč- {ā/i/e~a}  | 10th           | dahāw- {ā/i/e~a} |

### [2] Demonstratives

As mentioned before, human class have natural gender, viz. masculine and feminine, whereas all non-human class nouns are assigned neuter gender. This gender distinction is reflected in the use of demonstrative pronouns. Demonstratives used for human referents make gender distinction (masculine vs. feminine), demonstratives used for non-human referents use a neuter form. However, the process of gender neutralization in the case of human referents is in progress. We will discuss the distribution of the use of demonstratives in detail in Section 3.

Table 3. Demonstratives in Southern Marathi (the Thanjavur variety)

| Referent | Human referent<br>Gender marked<br>(masculine) | Human referent<br>Gender marked<br>(feminine) | Non-human<br>referent<br>(neuter) |
|----------|------------------------------------------------|-----------------------------------------------|-----------------------------------|
| Proximal | hyāna/hyo                                      | hina                                          | hye                               |
| Distal   | tyāna/tyo                                      | tina                                          | tye                               |

### [3] Genitive noun phrases

In the case of genitive forms, no gender distinction is made based on the referent metonymically evoked by the genitive form in SM, as shown in the table below. This is strikingly different from PM, where gender is obligatorily marked on all forms of all persons and numbers. See Tables 4 and 5 below.

Table 4. Genitives in Southern Marathi (the Thanjavur variety)

| Person | M/F/N |        |
|--------|-------|--------|
|        | SG    | PL     |
| 1      | māḍa  | āmtsa  |
| 2      | tudza | tumtsa |
| 3 (M)  | tyāḍa | tyāḍa  |
| 3 (F)  | tiḍa  | tyāḍa  |
| 3 (N)  | tyāḍa | tyāḍa  |

Table 5. Genitives in Standard Marathi (the Pune variety)

| Personal pronoun | M         |           | F        |           | N         |          |
|------------------|-----------|-----------|----------|-----------|-----------|----------|
|                  | Sg        | Pl        | Sg       | Pl        | Sg        | Pl       |
| 1                | mādzhā    | mādzhe    | mājhi    | mājhyā    | mādzhe    | mājhi    |
| 2                | tudzhā    | tudzhe    | tujhi    | tujhyā    | tudzhe    | tujhi    |
| 3 (M)            | tyātsā    | tyātse    | tyāči    | tyāčyā    | tyātse    | tyāči    |
| 3 (F)            | titsā     | titse     | tiči     | tičyā     | titse     | tiči     |
| 3 (N)            | tyātsā    | tyātse    | tyāči    | tyāčyā    | tyātse    | tyāči    |
| Noun             | N-OBL-tsā | N-OBL-tse | N-OBL-či | N-OBL-čyā | N-OBL-tse | N-OBL-či |

#### [4] Adjectives

Adjectives in SM are divided into two groups on the basis of whether or not they decline according to the number and gender of their referent: (i) declinable (*kāl-ā dādigā*, black-MSG man ‘black man’; *kāl-i bāyko* black-FSG woman ‘black woman’, *kāl-a {pora/kutra/lugda}* black-NSG {boy/dog/saree} ‘black {boy/dog/saree}’), and (ii) non-declinable (*dayāwanta {dādigā/bāyko/pora}* ‘kind {man/woman/boy}’). Irrespective of whether the adjective is declinable or indeclinable, the gender of its referent is overtly marked on the predicate of the sentence in which the adjective in question is used.

#### [5] Verb-based nominalizations

##### [5a] So-called relative clauses

In Standard Marathi (Pune variety), ‘relative-correlative’ strategy and “gap” strategy are used for formation of the so-called relative clause. Broadly speaking, the ‘relative-correlative’ strategy can be used for core (subject, direct object, indirect object) as well as non-core arguments (instrument, location, time, manner, etc.), while the “gap” strategy is primarily used for core arguments (subject and direct object; but see Hook and Pardeshi (2017) for further details). The “gap” strategy, also known as participial type, is considered to be the borrowing from Dravidian languages.

In the Thanjavur variety of Sothorn Marathi, in our elicitation we got spontaneous responses of the participial “gap” type constructions only. We did not come across constructions of ‘relative-correlative’ type in our elicitations, which is attested in Standard Marathi (Pune variety), but further investigation is needed on this issue.

##### [5b] Complement clauses

We have not collected data on noun and verb complement clauses yet.



### 3. DATA: Classifiers and gender marking

#### [1] Numerals

Among the grammatical nominalization to be discussed in this paper, numeral is the only category which can host both classifier and gender marking. Demonstrative, Genitive, Adjective and V-based nominalization on the other hand host gender marking alone.

#### [1a] Cardinal numerals

In Southern Marathi, numerals can be divided into two categories: (A) those that denote humans in and of themselves (L-form of ‘one’, viz. *ek.l* {*ā/i/e*} (one.L- {MSG/FSG/NSG} and G-form of ‘two’, ‘three’, and ‘four’) and (B) those that cannot denote humans in and of themselves (all bare numerals). Let us first discuss the numerals that denote humans in and of themselves: L-form of 1 and G-form of ‘two’, ‘three’, and ‘four’.

#### Numeral ‘one’: gendered form (L-form)

Among numerals, the numeral ‘one’ alone has gendered form, namely, *ek.l*-{*ā/i/e*} one. {MSG/FSG/NSG}, which denotes human referent. We refer to this form as L-form. The classifier *-jan* cannot be attached to the L-form of ‘one’. However, for numerals other than ‘one’, classifier *-jan* should be attached in order to denote human referent.

- (1) a. *kāl*                      *kiti-jan*                      *āl-e?*  
 yesterday    how.many-CLF.M/F    come.PF-MPL/FPL  
 ‘How many people (male + female, or only males, or only females) came yesterday?’

#### NP use

- b. *kāla*                      *ek.l*-{*ā/i/e*}                      *āl*-{*ā/i/e*} [-Honorific]  
 yesterday                  one.HUM- {MSG/FSG/NSG}                  come.PF- {MSG/FSG/NSG}  
 ‘Yesterday one person (MSG/FSG/NSG) (-Hon) came.’
- c. *kāla*                      *ek.l*-{*ā/i/e*}                      *āleyti* [+Honorific]  
 yesterday                  one.HUM- {MSG/FSG/NSG}                  come.PF.EPC.HON  
 ‘Yesterday one person (MSG/FSG/NSG) (+Hon) came.’
- d. *\*kāla*                      *ek.l*-{*a/i/e*}-*jan*                      *āl*-{*ā/i/e*} [-Honorific]  
 yesterday                  one.HUM- {MSG/FSG/NSG}-CLF                  come.PF- {MSG/FSG/NSG}  
*to mean*, ‘Yesterday one (MSG/FSG/NSG) (-Hon) came.’
- e. *\*kāla*                      *ek.l*-{*a/i/e*}-*jan*                      *āleyti* [+Honorific]  
 yesterday                  one.HUM- {MSG/FSG/NSG}-CLF                  come.PF.EPC.HON  
*to mean*, ‘Yesterday one (MSG/FSG/NSG) (+Hon) came.’

- f. \*kāla      **ek**      āl- {ā/i/e}/āleyti  
 yesterday one come.PF- {MSG/FSG/NSG}/come.PF.EPC.HON  
*to mean*, ‘Yesterday one person (MSG/FSG/NSG) (-Hon/+Hon) came.’

The L-form does not permit modification use. Bare numerals are used in Mod use.

#### Mod use

- g. \*kāla      **ek.l- {a/i/e}**      {dādigā/bāyko/pantoji}  
 yesterday one.HUM- {MSG/FSG/NSG} {man/woman/teacher}  
 āl- {ā/i/e}/āleyti  
 come.PF- {MSG/FSG/NSG}/come.PF.EPC.HON  
*to mean*, ‘Yesterday one {man/woman/teacher} (-Hon/+Hon) came.’
- h. \*kāla      **ek.l- {a/i/e}-jan**      {dādigā/bāyko/pantoji}  
 yesterday one.HUM- {MSG/FSG/NSG}-CLF {man/woman/teacher}  
 āl- {ā/i/e}/āleyti  
 come.PF- {MSG/FSG/NSG}/come.PF.EPC.HON  
*to mean*, ‘Yesterday, one {man/woman/teacher} (+Hon/-Hon) came.’
- i. kāla      **ek**      {dādigā/ bāyko/pantoji}  
 yesterday one man.MSG/woman.FSG/teacher.NSG  
 āl- {ā/i/e}/āleyti  
 come.PF- {MSG/FSG/NSG}/come.PF.EPC.HON  
 ‘Yesterday one man/woman/teacher (-Hon/+Hon) came.’

In Southern Marathi, neither the L-form of the numeral ‘one’ nor the bare numeral ‘one’, can host classifier. PM does not have the L-forms for the numeral ‘one’ but it permits the combination of the bare form plus the classifier (‘one’+ CLF).

Table 6. Denoting human using numeral ‘one’ in SM and PM

| Form of numeral<br>‘one’ | Southern Marathi |         | Pune Marathi |         |
|--------------------------|------------------|---------|--------------|---------|
|                          | NP use           | Mod use | NP use       | Mod use |
| L-form                   | OK               | Not OK  | NA           | NA      |
| Bare form                | Not OK           | OK      | Not OK       | OK      |
| L-form + CLF             | Not OK           | Not OK  | NA           | NA      |
| Bare form + CLF          | Not OK           | Not OK  | OK           | Not OK  |

This behaviour of ‘one’ for denoting human in Southern Marathi is similar to Konkani (see Pardeshi and Mopkar, in this volume).

In Sanskrit (Indo-Aryan), the mother language of Southern Marathi, the numeral ‘one’, ‘three’, and ‘four’ in their nominative form have three-way gender distinction (M/F/N), while the numeral ‘two’ has a two-way gender distinction (M vs. F/N). Modern Indo-Aryan languages

(Hindi, Marathi, etc.) have lost gender marking on numerals. From a historical/diachronic perspective, the L-form of Southern Marathi the numeral ‘one’ can be considered as the lone survivor of the old morphologically gender marked form of the numeral ‘one’ (cf. Spanish, where only the numeral ‘one’ has retained the gender distinction (*uno* vs *una*)).

Moving on to numerals ‘two’, ‘three’, and ‘four’ in Southern Marathi, they have a special form, which we call as G-form: *dog* (2.G), *t{i/e}g* (3.G), *ča{o/u}g* (4.G), like PM and Konkani. G-forms of ‘two’, ‘three’, and ‘four’ are similar to L-form of ‘one’ in that they both denote human in and of themselves. They differ, however, in that ‘one’ has gender marked forms (M/F/N), while ‘two’, ‘three’, and ‘four’ do not. They also differ in terms of concatenability with CLF: L-forms do not concatenate with CLF, while G-forms do. Examples of G-forms of numeral ‘two’, ‘three’, and ‘four’, without and with CLF from Southern Marathi are given below.

### G-forms of ‘two’, ‘three’ and ‘four’ (without classifier): *dog/tig/čaug*

- (2) a. *kāla*            *kiti*            *dādige*            *āl-e?*  
          yesterday   how.many    man.PL            come.PF-MPL  
          ‘How many men came yesterday?’

#### NP use

- b. *kāla*            {*doga/tiga/čauga*}            *āl-e*  
          yesterday   {two.G/three.G/four.G}    come.PF-PL  
          ‘{Two/three/four} people came yesterday.’

#### Mod use

- c. *kāla*            {*doga/tiga/čauga*}            {*dādige/bāykā/pantoji*}            *āl-e*  
          yesterday   {two.G/three.G/four.G}    {men/women/teachers}    come.PF-PL  
          ‘{Two/three/four} {men/women/teachers} came yesterday.’

### G-forms of ‘two’, ‘three’ and ‘four’ + CLF ({*dog/tig/čaug*}-jan):

#### NP use

- d. *kāla*            {*doga-jan/tiga-jan/čauga-jan*}            *āl-e*  
          yesterday   {two.G-CLF/three.G-CLF/four.G-CLF}    come.PF-PL  
          ‘{Two/three/four} people came yesterday.’

#### Mod use

- e. *kāla*            {*doga-jan/tiga-jan/čauga-jan*}            {*dādige/bāykā/pantoji*}  
          yesterday   {two.G-CLF/three.G-CLF/four.G-CLF}    {men/women/teachers}  
          *āl-e*  
          come.PF-PL  
          ‘{Two/three/four} {men/women/teachers} came yesterday.’

Cardinal ‘two’, ‘three’ and ‘four’ (bare numerals) in combination with classifier can also be used for referring to human referents, as shown below.

### Numeral (‘two’, ‘three’, ‘four’) + CLF: {don/tin/čār} + jan

- (3) a. kāl                      kiti-**jan**                      āl-e?  
yesterday how.many-CLF.M/F come.PF-MPL/FPL  
‘How many people (male + female, or only males, or only females) came yesterday?’

#### NP use

- b. kāla                      {**don/tin/čār/pāč**}-**jan**                      āl-e  
yesterday {two/three/four/five}-CLF come.PF-PL  
‘Yesterday {two/three/four/five} people came.’

#### Mod use

- c. kāla                      {**don/tin/ čār/pāč**}-**jan**                      {dādige/bāykā/pantoji}                      āl-e  
yesterday {two/three/four/five}-CLF {men/women/teachers} come.PF-PL  
‘Yesterday {two/three/four/five} {men/women/teachers} came.’

Table 7. Denoting human using numeral ‘two’, ‘three’ and ‘four’ in SM and PM

| Form of numeral<br>‘two’, ‘three’ and<br>‘four’ | Southern Marathi (SM) |                                   | Pune Marathi (PM) |         |
|-------------------------------------------------|-----------------------|-----------------------------------|-------------------|---------|
| Use                                             | NP use                | Mod use                           | NP use            | Mod use |
| G-form                                          | OK                    | OK                                | OK                | Not OK  |
| G-form + CLF                                    | OK                    | OK<br>(Not OK for<br>one speaker) | OK                | Not OK  |
| Bare form + CLF                                 | OK                    | OK (Not OK<br>for one<br>speaker) | OK                | Not OK  |

### Numeral ‘five’ and above

For numeral ‘five’ and above, there are no special forms such as the L-form (which is available only for numeral ‘one’) and G-forms (which are available for ‘two’, ‘three’ and ‘four’ ). To denote human beings, the only possible way is to attach classifier to the numeral in question: {cardinal numeral (NUM) + CLF}. Inanimate referents, on the other hand, can be denoted by cardinal numerals alone.

### Numeral + CLF: **pāč** ('five') + **jan**

- (4) a. **kāl**                      **kiti-jan**                      **āl-e?**  
 yesterday how.many-CLF.M/F come.PF-MPL/FPL  
 'How many people (male + female, or only males, or only females) came yesterday?'

#### NP use

- b. **pāč-jan**    **al-e**  
 five-CLF come.PF-MPL/FPL  
 'Five (MPL) came/had come.'
- c. **\*pāč**                      **al-e**  
 five                      come.PF-MPL/FPL  
*to mean*, 'Five (MPL) came/had come.'

#### Mod use

- d. **pāč-jan**    {**dādige/bāykā/pantoji**}                      **āl-e**  
 five-CLF man.MPL/woman.FPL/teacher.PL come.PF-MPL/FPL  
 'Five {men/women/teachers} came/had come.'
- e. **\*pāč**    {**dādige/bāykā/pantoji**}                      **āl-e**  
 five man.MPL/woman.FPL/teacher.PL come.PF-MPL/FPL  
*to mean*, 'Five {men/women/teachers} came/had come.'

PM does not allow the Mod use of {NUM+CLF} while SM permits it as shown in (d) above (except for one speaker). In PM gender marking can be attached to the classifier ({NUM+CLF+GEN}), while in SM, gender marking cannot be attached to CLF.

Let us summarize the distribution of classifier and gender marking with regard to numerals (NUM).

- (a) In SM, like PM, classifier can be attached to numerals only. However, unlike PM, SM does not permit gender marking on the CLF.
- (b) In SM, unlike PM, numeral 'one' has gendered L-forms (M/F/N). This is the only numeral that has gendered form. In this regard, SM is like Konkani (Goan as well as Southern Konkani variety). The L-form is permitted in NP use but not permitted in Mod use. Also, L-form does not host the classifier for humans. The distribution of L form of 'one' in SM is similar to its Konkani counterpart. The distribution of gender and classifier marking on numeral 'one' in SM (and Konkani) shows that they are in complementary distribution: both cannot cooccur. Southern Konkani does not permit combinations of cardinal numeral form of 'one' plus classifier. In SM, this combination is permitted but only in NP use.
- (c) In SM, like PM, and Konkani, numerals 'two', 'three' and 'four', have a special G-form to denote human. However, their distribution differs. In SM, G-forms permit NP use as well as Mod use and also host CLF in both the uses (except for one speaker, who does not allow

the use of CLF in Mod use, like SM). In PM, G-forms, with or without CLF, are permitted only in NP use.

- (d) In SM, cardinal numerals except ‘one’ host classifier, in both NP and Mod use (except for one speaker, who does not allow the use of CLF in Mod use). In PM, in contrast, all numerals host CLF but permit only NP use.

The summary above is based on data collected from a few consultants. A more detailed study is needed to confirm the distribution of classifier and gender marking on numerals.

### [1b] Ordinal numerals

Ordinal numerals in Southern Marathi are not investigated for CLF and gender marking. It needs further investigation.

### [2] Demonstratives

As mentioned before, human class have natural gender, viz. masculine and feminine, whereas all non-human class nouns are assigned neuter gender. This gender distinction is reflected in the use of demonstrative pronouns. Typically, demonstratives used for human referents have gender distinction (masculine and feminine) whereas demonstratives used for non-human referents use a neuter form.

Interestingly, in the case where deference is shown towards human referent, the process of gender neutralization is attested in NP use, as shown in example (7), which is unequivocally accepted by all the speakers we consulted. However, in the case of Mod use, gender neutralization seems to be an ongoing change in that some speakers accept gender neutralized forms while others do not, as shown in the example (10) and (11) below.

#### NP use: No deference (Gender marked, masculine)

- (5) {**hyāna~hyo/tyāna~tyo**}                      šānā/akkalwanta  
       {3MSG.PROX/3MSG.DIST}    intelligent.M/intelligent.N(EPC)  
       ‘{This (3MSG)/That (3MSG)} is intelligent.’

#### NP use: No deference (Gender marked, feminine)

- (6) {**hina/tina**}                                      šāni/akkalwanta  
       {3FSG.PROX/3FSG.DIST}    intelligent.F/intelligent.N(EPC)  
       ‘{This(3FSG)/That(3FSG)} is intelligent.’

#### NP use: With deference (Gender neutralized, masculine or feminine referent)

- (7) {**hyeni/tyeni**}                                      šāne/akkalwanta  
       {3PROX.HON/3DIST.HON} intelligent.N(EPC)  
       ‘{This (MSG.Hon/FSG.Hon)/That (MSG.Hon/FSG.Hon)} is intelligent.’

**Mod use:**

**Masculine referent: gender marked form**

- (8) {**hyāna~hyo/tyāna~tyo**}                      dādigā              šānā/akkalwanta  
      {3MSG.PROX/3MSG.DIST}              man              intelligent.M/intelligent.N(EPC)  
      ‘{This (3MSG)/That (3MSG)} man is intelligent.’

**Feminine referent: gender marked form**

- (9) {**hina/tina**}                                      bāyko              šāni/akkalwanta  
      {3FSG.PROX/3FSG.DIST}              woman              intelligent.F/intelligent.N(EPC)  
      ‘{This(3FSG)/That(3FSG)} woman is intelligent.’

**Masculine referent: gender neutral form**

- (10) ?{**hyeni/tyeni**}                                      dādigā              šānā/akkalwanta  
      {3PROX.HON/3DIST.HON} man              intelligent.M/intelligent.N(EPC)  
      *to mean*, ‘{This (MSG.Hon/FSG.Hon)/That (MSG.Hon/FSG.Hon)} man is intelligent.’

**Feminine referent: gender neutral form**

- (11) ?{**hyeni/tyeni**}                                      bāyko              šāni/akkalwanta  
      {3PROX.HON/3DIST.HON} woman              intelligent.F/intelligent.N(EPC)  
      *to mean*, ‘{This (MSG.Hon/FSG.Hon)/That (MSG.Hon/FSG.Hon)} woman is intelligent.’

Further, neuter forms used for non-human referent, viz. *hye/tye*, can also be used for referring to human beings, if the referent is (i) closely related to and (ii) younger in age than the speaker.

- (12) {**hye/tye**}                                              dādigā/bāyko              akkalwanta  
      {3NSG.PROX/3NSG.DIST} man/woman              intelligent.N(EPC)  
      ‘{This/That} man/woman is intelligent.’

In the case of non-human referent, neuter forms are used. Gender marked forms, which are reserved for human referents, cannot be used.

**NP use: non-human referent**

- (13) {**hye/tye**}                                              šāna/akkalwanta  
      {3NSG.PROX/3NSG.DIST}              intelligent.N/intelligent.N(EPC)  
      ‘{This/That} one (NSG) is intelligent.’
- (14) \*{**hyeni/tyeni**}                                              šāna/akkalwanta  
      {3PROX.HON/3DIST.HON}              intelligent.N/intelligent.N(EPC)  
      *to mean*, ‘{This (MSG.Hon/FSG.Hon)/That (MSG.Hon/FSG.Hon)} one (NSG) is intelligent.’

- (15) \*{hina/tina} šāni/akkalwanta  
 {3FSG.PROX/3FSG.DIST} intelligent.F/intelligent.N(EPC)  
*to mean, ‘{This(3FSG)/That(3FSG)} one (NSG) is intelligent.’*

**Mod use:** non-human referent

- (16) {hye/tye} kutra šāna/akkalwanta  
 {3NSG.PROX/3NSG.DIST} dog intelligent.N/intelligent.N(EPC)  
*‘{This/That} dog is intelligent.’*
- (17) \*{hyeni/tyeni} kutra šāna/akkalwanta  
 {3PROX.HON/3DIST.HON} dog intelligent.N/intelligent.N(EPC)  
*to mean, ‘{This (MSG.Hon/FSG.Hon)/That (MSG.Hon/FSG.Hon)} dog is intelligent.’*
- (18) \*{hina/tina} kutra šāni/akkalwanta  
 {3FSG.PROX/3FSG.DIST} dog intelligent.F/intelligent.N(EPC)  
*to mean, ‘{This(3FSG)/That(3FSG)} dog is intelligent.’*

The distribution of demonstratives in SM is summarized in the table below.

Table 8. Distribution of DEM in Southern Marathi (the Thanjavur variety)

| Referent                   | Human referent<br>Masculine |            |                   |                           | Human referent<br>Feminine |            |                   |                           | Non-human<br>referent<br>Neuter |            |
|----------------------------|-----------------------------|------------|-------------------|---------------------------|----------------------------|------------|-------------------|---------------------------|---------------------------------|------------|
| Deference                  | Without<br>deference        |            | With<br>deference |                           | Without<br>deference       |            | With<br>deference |                           | Not applicable                  |            |
| Proximal                   | hyāna/hyo                   |            | hyeni             |                           | hina                       |            | hyeni             |                           | hye                             |            |
| Distal                     | tyāna/tyo                   |            | tyeni             |                           | tina                       |            | tyeni             |                           | tye                             |            |
| Use                        | NP<br>use                   | Mod<br>use | NP<br>use         | Mod<br>use                | NP<br>use                  | Mod<br>use | NP<br>use         | Mod<br>use                | NP use                          | Mod<br>use |
| Acceptability<br>judgement | OK                          | OK         | OK                | <b>OK/<br/>Not<br/>OK</b> | OK                         | OK         | OK                | <b>OK/<br/>Not<br/>OK</b> | OK                              | OK         |

Summary of DEM: In SM, human class have natural gender, viz. masculine and feminine, whereas all non-human class nouns are assigned neuter gender. This gender distinction is reflected in the use of demonstrative pronouns. Typically, demonstratives used for human referents have gender distinction (masculine and feminine), demonstratives used for non-human referents use a neuter form. In SM, in the case where deference is shown towards human referent, the process of gender neutralization is attested in NP use, while in the case of Mod



use, it seems to be an ongoing change in that some speakers accept gender neutralized forms while others do not. In PM, in contrast to SM, gender marked (M/F/N) forms of demonstratives are used, in NP use and Mod use for human as well as non-human referents.

### [3] Genitive noun phrases

In Southern Marathi, gender on GEN is neutralized in all persons and numbers and in NP use as well as Mod use. This is strikingly different from standard Marathi, where gender is obligatorily marked in all persons and numbers and in NP use as well as Mod use.

#### Animate referent: NP use

- (19) a. **māja**      yēni      **tija**      tyēni  
 1SG.GEN this.HON 3FSG.GEN that.HON  
 ‘Mine is this (MSG.HON), hers is that (MSG.HON).’
- b. **māja**      hyēni      **tyāja**      tyēni  
 1SG.GEN this.HON 3MSG.GEN that.HON  
 ‘Mine is this (FSG.HON), his is that (FSG.HON).’
- c. **māja**      hye      **tyānja**      tye  
 1SG.GEN this.NSG 3PL.GEN that.NSG  
 ‘Mine is this (NSG), theirs is that (NSG).’
- d. **māja**      dallā      hyēni      **tija**      dallā      tyēni  
 1SG.GEN husband this.HON 3FSG.GEN husband that.HON  
 ‘My husband is this, her husband is that.’

#### Animate referent: Mod use

- (20) a. **māja**      bāil      hyēni      **tyāja**      bāil      tyēni  
 1SG.GEN wife this.HON 3MSG.GEN wife that.HON  
 ‘My wife is this, his wife is that.’
- b. **māja**      kutra      hye      **tyānja**      kutra      tye  
 1SG.GEN dog this.NSG 3PL.GEN dog that.NSG  
 ‘My dog is this, their dog is that.’

Summary of GEN: In SM, gender marking on GEN (possessor) is neutralized (i.e., gender of the referent of the possessee is not marked on GEN form of the possessor) in all persons and numbers and also in both NP use and Mod use. In PM, gender marking on GEN is obligatory in all persons and numbers and in NP use as well as in Mod use.

### [4] Adjectives

In Southern Marathi, like standard Marathi, adjectives are divided into two groups: those which decline as per the gender/number of the denotation and those which do not decline as

per the gender/number of the denotation. The examples of both the types in their NP use and Mod use are given below.

Declinable type:

**NP use:**

- (21) a. ami {**kālyā-lā/ kāḷi-lā/ kāḷa**}  
 1PL black.OBL.MSG-ACC/black.OBL.FSG-ACC/black.OBL.NSG.ACC  
 ishṭa paḍ-t-o  
 like fall-IMPFV-1PL  
 ‘We like black one {MSG/FSG/NSG}.’

**Mod use**

- b. ami **kāḷā** mānsālā ishṭa paḍ-t-o  
 1PL black.MSG man.OBL-ACC like fall-IMPFV-1PL  
 ‘We like a black man.’
- c. ami **kāḷi** bāykolā ishṭa paḍ-t-o  
 1PL black.FSG woman.OBL-ACC like fall-IMPFV-1PL  
 ‘We like a black woman.’
- d. ami **kāḷa** {lugḍa/kutra} ishṭa paḍ-t-o  
 1PL black.NSG {saree/dog}.NSG like fall-IMPFV-1PL  
 ‘We like a black {saree/dog}.’

Non-declinable type:

**NP use:**

- (22) a. malā **dayāwanta** āwad-t-{o/e/a}  
 I.DAT kind.3MSG/3FSG/3NSG like-IMPFV-3MSG/3FSG/3NSG  
 ‘I like a kind one (3MSG/3FSG/3NSG).’

**Mod use:**

- b. malā **dayāwanta** dādigā āwad-t-o  
 I.DAT kind.3MSG man.3MSG like-IMPFV-3MSG  
 ‘I like a kind man.’
- c. malā **dayāwanta** bāyko āwad-t-e  
 I.DAT kind.3FSG woman.3FSG like-IMPFV-3FSG  
 ‘I like a kind woman.’
- d. malā **dayāwanta** pore āwad-t-e  
 I.DAT kind.3FSG girl.3FSG like-IMPFV-3FSG  
 ‘I like a kind girl.’

- e. malā      **dayāwanta**      pora      āwad-t-a  
 I.DAT      kind.3NSG      boy.3NSG      like-IMPV-3NSG  
 ‘I like a kind boy.’

In the indeclinable type of adjective, although the adjective does not carry an overt gender marker, the gender of its referent can be understood from the gender marking on the predicate.

Summary of ADJ: In Southern Marathi, in the case of human referent, gender is marked on declinable adjectives while in the case of indeclinable adjective, although the gender of the referent is not marked overtly on the adjective, it can be understood from the gender marking on the predicate, both in NP use and Mod use, like standard Marathi (PM). However, in the case of non-human/inanimate referent, unlike standard Marathi, gender neutralization (treating all non-human entities as neuter) is observed in southern Marathi, both in NP use and Mod use. The pattern of using neuter gender for non-human/inanimate referent in Southern Marathi is a reflection of a shift Southern Marathi has undergone from grammatical gender to natural gender.

## [5] Verb-based nominalizations

### [5a] Relative clauses

As mentioned earlier in Section 2, in the Thanjavur variety of Southern Marathi, we got spontaneous responses of the participial “gap” type constructions, but did not come across ‘relative-correlative’ type constructions in our elicitation. The ‘relative-correlative’ type construction is widely attested in Standard Marathi (Pune variety), but further investigation is needed on this issue. Further, in case gap-type construction in Southern Marathi, we found gender marked forms like standard Marathi as well as gender neutralized forms, which are unique to Southern Marathi.

**Gender marked forms: present participle: -*ṇār*{*ā/i/e*}; past participle: -*lel*{*ā/i/e*}**

### Masculine referent

#### NP use

- (23) a. **khurjīwar**      {**bas-ṇār-ā/ bas-lel-ā**}  
 chair.LOC      sit-PRS.PTCPL-MSG/sit-PST.PTCPL-MSG  
 māja      dallā  
 1SG.GEN.NSG      husband  
 ‘The one (MSG) {sitting/seated} on the chair is my husband.’

#### Mod use

- b. **khurjīwar**      {**bas-ṇār-ā/bas-lel-ā**}      **dādigā**  
 chair.LOC      sit-PRS.PTCPL-MSG/sit-PST.PTCPL-MSG man  
 māja      dallā  
 1SG.GEN.NSG      husband  
 ‘The man {sitting/seated} on the chair is my husband.’

## Feminine referent

### NP use

- (24) a. **khurjīwar** {**bas-ṇār-i/bas-lel-i**}  
chair.LOC sit-PRS.PTCPL-FSG/sit-PST.PTCPL-FSG  
māja bāil  
1SG.GEN.NSG wife  
'The one (FSG) {sitting/seated} on the chair is my wife.'

### Mod use

- b. **khurjīwar** {**bas-ṇār-i/bas-lel-i**} **bāyko**  
chair.LOC sit-PRS.PTCPL-FSG/sit-PST.PTCPL-FSG woman  
māja bāil  
1SG.GEN.NSG wife  
'The woman {sitting/seated} on the chair is my wife.'

## Neuter referent

### NP use

- (25) a. **khurjīwar** {**bas-ṇār-a/bas-lel-a**}  
chair.LOC sit-PRS.PTCPL-NSG/sit-PST.PTCPL-NSG  
māja lēnkru  
1SG.GEN.NSG child.NSG  
'The one (NSG) {sitting/seated} on the chair is my child.'

### Mod use

- b. **khurjīwar** {**bas-ṇār-a /bas-lel-a**} **por**  
chair.LOC sit-PRS.PTCPL-NSG/sit-PST.PTCPL-NSG boy.NSG  
māja lēnkru  
1SG.GEN.NSG child.NSG  
'The boy (NSG) {sitting/seated} on the chair is my child.'

## Gender neutralized forms: *-alte; -ingrāyāca*

## Masculine referent

### NP use

- (26) a. **khurjīwar** **bas-alt-a** māja dallā  
chair.LOC sit-PTCPL-NSG 1SG.GEN.NSG husband  
'The one (NSG) sitting on the chair is my husband.'
- b. **khurjīwar** **bas-ingrāyātsa** māja nawrā  
chair.LOC sit-PTCPL-NSG 1SG.GEN.NSG husband  
'The one (NSG) sitting on the chair is my husband.'

### Mod use

- c. **khurjiwar**    **bas-alt-a**            **dādigā**    māja            dallā  
chair.LOC    sit-PTCPL-NSG    man    1SG.GEN.NSG    husband  
‘The man sitting on the chair is my husband.’
- d. **khurjiwar**    **bas-ingrāyātsa**    **dādigā**    māja            nawrā  
chair.LOC    sit-PTCPL-NSG    man    1SG.GEN.NSG    husband  
‘The man sitting on the chair is my husband.’

### Feminine referent

#### NP use

- (27) a. **khurjiwar**    **bas-alt-a**            māja            bāil  
chair.LOC    sit-PTCPL-NSG    1SG.GEN.NSG    wife  
‘The one (NSG) sitting on the chair is my wife.’
- b. **khurjiwar**    **bas-ingrāyātsa**    māja            bāil  
chair.LOC    sit-PTCPL-NSG    1SG.GEN.NSG    wife  
‘The one (NSG) sitting on the chair is my wife.’

### Mod use

- c. **khurjiwar**    **bas-alt-a**            **bāyko**    māja            bāil  
chair.LOC    sit-PTCPL-NSG    woman    1SG.GEN.NSG    wife  
‘The woman sitting on the chair is my wife.’
- d. **khurjiwar**    **bas-ingrāyātsa**    **bāyko**    māja            bāil  
chair.LOC    sit-PTCPL-NSG    woman    1SG.GEN.NSG    wife  
‘The woman sitting on the chair is my wife.’

### Neuter referent

#### NP use

- (28) a. **khurjiwar**    **bas-alt-a**            māja            lēnkru  
chair.LOC    sit-PTCPL-NSG    1SG.GEN.NSG    child  
‘The one (NSG) sitting on the chair is my child.’

### Mod use

- b. **khurjiwar**    **bas-alt-a**            **por**    māja            lēnkru  
chair-LOC    sit-PTCPL-NSG    boy.NSG    1SG.GEN.NSG    child.NSG  
‘The boy (NSG) sitting on the chair is my child.’

## Non-human referent

### NP use

- (29) a. **kāl**                    **ghe.t-lel-a**                    mol    adik  
yesterday buy.EPN-PST.PTCPL-NSG cost    more  
'The one (NSG) I bought yesterday is expensive.'
- b. **kāl**                    **ghe.t-lat-e**                    **mobāil**                    mol    adik  
yesterday buy.EPN-PTCPL-NSG mobile.NSG cost    more  
'The mobile phone (NSG) I bought yesterday is expensive.'

Following Shibatani (2023), we analyse the so-called relative clauses as V-based nominalizations, which nominalize the structure as well as classify it. In (23), (24), and (25) gender of the denotation, respectively masculine, feminine and neuter, is marked on the V-based nominalizations, both in NP use and in Mod use. This is similar to standard Marathi (Pune variety). On the other hand, in (26), (27) and (28), the gender of the denotation on the V-based nominalization is lost and irrespective of the gender of the denotation, neuter form is used. This is a case of gender neutralization or loss of gender marking. As shown in (29), the neuter form is used for non-human referents, and this form is used in the case where gender neutralization occurs. We speculate that the cause of loss of gender marking on V-based nominalizations in Southern Marathi is intensive contact with Tamil (Dravidian).

Summary of V-based: In Southern Marathi, in the case of human referent, there are two patterns: (i) in which gender is marked on V-based both in NP use and Mod use, like standard Marathi and (ii) in which gender is neutralized, just like in the case of inanimate referents, both in NP use and Mod use. The latter pattern, which is not attested in standard Marathi, arguably shows a shift from grammatical gender to natural gender.

## [5b] Complement clauses

We have not collected data on complement clauses so far.

We will now summarize the data presented above on grammatical nominalizations, viz. numerals (NUM), demonstrative pronouns (DEM), genitives (GEN), adjectives (ADJ) and V-based nominalizations (V-based) in Southern Marathi (Thanjavur variety) in their NP use and Mod use, in the form of Shibatani's hierarchy constraining the distribution of classifier and gender marking (see Shibatani 2023: 35–37).

## CLF marking: Southern Marathi vs. Standard Marathi

Figure A. Classifier marking pattern in Southern Marathi (Thanjavur variety)

NP use: \*OBL/OPTNUM > \*DEM > \*GEN > \*ADJ > \*V-based NMLZ

Mod use: \*OBL/OPTNUM > \*DEM > \*GEN > \*ADJ > \*V-based NMLZ

Figure B. Classifier marking pattern in Standard Marathi (Pune variety)

NP use: NUM > \*DEM > \*GEN > \*ADJ > \*V-based NMLZ

Mod use: \*NUM > \*DEM > \*GEN > \*ADJ > \*V-based NMLZ

\*OBL/OPT= obligatory in case of numerals ‘two’ and above, optional in the case of the G-forms of numerals ‘two’ through to ‘four’ and not permitted in the case of the numeral ‘one’. Acceptability judgement: unequivocally accepted in NP use but there is subject to inter-speaker variation in Mod use: one speaker does not allow the use of CLF in Mod use while others do.

\*= classifier marking not allowed

## Gender marking: Southern Marathi vs. Standard Marathi

Figure C. Gender marking pattern in Southern Marathi (Thanjavur variety)

NP use: MKD/LIMNUM > MKD/NEUTDEM > NEUTGEN > MKD/NEUTADJ > MKD/NEUTV-based NMLZ

Mod use: MKD/LIMNUM > MKD/NEUTDEM > NEUTGEN > MKD/NEUTADJ > MKD/NEUTV-based NMLZ

Figure D. Gender marking pattern in Standard Marathi (Pune variety)

NP use: OBL/\*LIMNUM > OBLDEM > OBLGEN > OBLADJ > OBLV-based NMLZ

Mod use: OBL/\*LIMNUM > OBLDEM > OBLGEN > OBLADJ > OBLV-based NMLZ

LIM= limited to numeral ‘one’ only (L-form)

\*LIM= severely limited [only permitted in non-nominative use of ‘one’ in NP use, G-forms for ‘two’, ‘three’, and ‘four’ (only two-way gender distinction, viz. epicene vs. F) in NP use. Mod use of G-forms is limited in that only a few nouns that express relationships (brother, sister, etc.) are allowed.]

OBL= obligatory

MKD/NEUT= gender marked (i.e., distinguished) in some cases and neutralized in other cases, e.g., in case of DEM, marked in the absence of deference toward human referent, gender neutralized when deference is expressed; in the case of ADJ, marked in the case of human referents and neutralized in the case of non-human referent (treating as neuter); in the case of V-based NMLZ, marked in the case of *-ñār{ā/i/e}/-lē{ā/i/e}* type and neutralized in the case of *-alte*; *-ingrāyāca* type.

It should be added that these summaries are based on a limited data and are tentative. They need to be confirmed in the follow-up work. There are gaps in the data. We have no data on ordinal numbers, complement clauses, lexical nominalization, etc. We need to cast our data net further and refine our description as well as analysis. We also need to do research on the

diachronic development of the use of classifiers and/or gender markers with grammatical nominalizations and also compare grammatical nominalizations in Marathi with their counterparts in neighboring languages, especially as Konkani and Saurashtra, which presents a similar situation: an Indo-Aryan language having grammatical gender system transplanted in an area of Dravidian languages, which have a semantic gender system.

#### **4. ANALYSIS AND DISCUSSION**

As for the distribution of the sole classifier in SM, it can be attached to NUM only, with a few deviations such as the numeral ‘one’ to which CLF cannot be attached and special forms of numerals ‘two’ to ‘four’ (referred to as G-form) to which CLF can be unequivocally attached in NP use, but not in the modification use where interspeaker variation is observed. With regard to gender marking, gender distinction is either marked or neutralized (under specific conditions) in case of all structural slots on Shibatani’s hierarchy except GEN, where gender distinction is completely lost, both in NP use and in Mod use. In sum, while the distribution of CLF supports the prediction of Shibatani’s hierarchy, the distribution of gender marking defies it.

In Southern Marathi, the process of: (i) acquisition (use) of a classifier and (ii) loss of gender, are attested. In many Indo-European languages, numerals have lost gender. In some Indo-European languages numerals have lost gender altogether (English, Hindi-Urdu, etc.), while in others, such as Russian, Spanish, Slovene, Romanian, Bulgarian, Konkani, etc. gender is retained on lower numerals (below ‘five’). In Southern Marathi and Konkani, like Spanish, gender is retained only on the cardinal numeral ‘one’. Hindi has lost gender on DEM in addition to NUM. As for classifier, standard Hindi does not have a classifier, but its southern neighbour, Marathi, has acquired one classifier, like Konkani. Assamese, on the East side, has lost gender altogether and acquired CLF which can be attached to all categories of Shibatani’s hierarchy. These distributions suggest that the process of loss of gender as well as acquisition of classifier starts from the left-most category on Shibatani’s hierarchy and then proceeds rightward along the structural dimension and on the functional dimension, it starts from NP use and then spreads to Mod use.

The process of acquisition of classifier in Marathi (PM and SM) starts from the leftmost end NUM. In SM, attachment of classifier is possible only to NUM like SM but while in SM it is possible both in NP use and in Mod use, it is restricted to NP use in PM. This is as per the prediction of Shibatani’s hierarchy.

The process of loss of gender (or gender neutralization) in SM, however, presents an exception to prediction of Shibatani’s hierarchy. It starts from the leftmost end NUM spreads rightward along the structural dimension and from NP use to Mod use on the functional dimension. The GEN slot poses a problem. In SM, gender marking on GEN is lost altogether but the same has not yet happened in the case of categories to its right, viz. ADJ and V-based, which still permit gender marked forms.

#### **5. CONCLUSION**

In the present study, we offered a preliminary report on the distribution of classifiers and gender markers in Southern Marathi (an Indo-Aryan language spoken in Tamil Nadu state,



India, abbreviated as SM) based on the data collected in our fieldwork conducted in Chennai, Krishnagiri and Thanjavur districts in Tamil Nadu state. Adopting the framework called “dynamic functional typology” proposed in Shibatani (2023) and following the hierarchy (NUM > DEM > GEN > ADJ > V-based) proposed therein (op. cit. 35), we described the distribution of classifiers and gender markers in Southern Marathi and compare it with the variety of Marathi spoken in Pune, Maharashtra state, India (the so-called Standard Marathi, referred to as PM). Our findings are as follows. As for the distribution of the sole classifier in SM, it can be attached to NUM only, with a few deviations such as the numeral ‘one’ to which CLF cannot be attached and special form of the numerals ‘two’ to ‘four’ (referred to as G-form) to which CLF can be unequivocally attached in NP use, but not in the modification use where interspeaker variation is observed. Regarding gender marking, gender distinction is either marked or neutralized (under specific conditions) in case of all structural slots on Shibatani’s hierarchy except GEN, where gender distinction is completely lost, both in NP use and in Mod use. In sum, while the distribution of CLF supports the prediction of Shibatani’s hierarchy, the distribution of gender marking defies it.

## ABBREVIATIONS

|          |                          |
|----------|--------------------------|
| 1        | first person             |
| 3        | third person             |
| ACC      | accusative               |
| ADJ      | adjective                |
| CARD.NUM | cardinal numeral         |
| CLF      | classifier               |
| COMP     | complementizer           |
| COMP.CL  | complement clause        |
| DAT      | dative                   |
| DEM      | demonstrative            |
| DIST     | distal                   |
| EPC      | epicene (gender-neutral) |
| F        | feminine                 |
| FPL      | feminine plural          |
| FSG      | feminine singular        |
| GEN      | genitive                 |
| HON      | honorific                |
| LOC      | locative                 |
| M        | masculine                |
| Mod      | modification             |
| MPL      | masculine plural         |
| MSG      | masculine singular       |
| N        | neuter                   |
| NMLZ     | nominalization           |
| NP       | noun phrase              |
| NSG      | neuter singular          |

|         |                                 |
|---------|---------------------------------|
| NUM     | numeral                         |
| OBL     | oblique                         |
| PF      | perfective                      |
| PL      | plural                          |
| PM      | Pune Marathi (Standard Marathi) |
| PROX    | proximal                        |
| PRS     | present                         |
| PST     | past                            |
| PTCPL   | participle                      |
| PTCP.CL | participle clause               |
| SG      | singular                        |
| SM      | Southern Marathi                |
| V-based | verb-based nominalization       |

## ACKNOWLEDGMENTS

The work in progress that we are going to present today would not have been possible without wholehearted support and cooperation from the following consultants who are the native speakers of Southern Marathi and also know Tamil well: Mr. Ramachandran Rao, Mrs. P. M. Padma (Chennai), Mr. Janardana Rao, Mrs. Laxmikantabai Shinde, Mrs. Geetabai Nikam (Krishnagiri), Mrs. P. Manjula, Mr. R. V. Srinivasan (Thanjavur). We would like to thank our consultants for their cooperation and patience in answering our questions for a prolonged period extending over 2 to 3 hours. We would also like to thank our local coordinators: (i) Mrs. P. M. Padma (Chennai), (ii) Mr. Janardana Rao (Krishnagiri) and (iii) B. Ramachandran (Thanjavur) who helped us in reaching out to the consultants. Special thanks are due to: (i) Prof. Rajesh Kumar and Dr. Milind Brahme of the Department of Humanities and Social Sciences, IIT Madras (Chennai), (ii) Mr. Janardana Rao (Krishnagiri), and (iii) B. Ramachandran (Thanjavur) for their support and cooperation in providing us with space and logistical support to record the elicitation sessions. Finally, we are thankful to Dr. Sonal Kulkarni-Joshi, Head, Department of Linguistics, Deccan College and her colleagues for presenting our work at the Linguistics Seminar held on 13<sup>th</sup> Feb 2025. We are also thankful to the participants for their comments and suggestions.

The work in progress reported has been supported by the NINJAL project “Empirical Study of the Typology of Nominalization—from Theoretical, Fieldwork, Historical and Dialectal Perspective” (PI: Prashant Pardeshi) and by the JSPS KAKENHI Grant-in-Aid for Scientific Research (B) 22H00659 ‘Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers’ (PI: Kazuhiro Kawachi).

## REFERENCES

- Corbett, Greville G. 2013. Number of genders. In Matthew S. Dryer and Martin Haspelmath. (eds.) *WALS Online* (v2020.4) [Data set]. Zenodo.  
<https://doi.org/10.5281/zenodo.13950591>
- Dhongde, Ramesh, and Kashi Wali. 2009. *Marathi* (London Oriental and African Language Library 13). Amsterdam/Philadelphia: John Benjamins

- Dixon, Robert M. W. 1982. *Where have all the adjectives gone?, and other essays in semantics and syntax*. Berlin: Mouton de Gruyter.
- Hook, Peter, and Prashant Pardeshi. 2017. Noun-modifying constructions in Marathi. In Yoshiko Matsumoto, Bernard Comrie, and Peter Sells. (eds.) *Noun-modifying clause constructions in languages of Eurasia: Rethinking theoretical and geographical boundaries*, 293–329. Amsterdam: John Benjamins.
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)

# Nominalizations with classifiers and gender markers in Sinhala: Usage patterns from online sources

Tetsuya Miyagishi<sup>1</sup>

## ABSTRACT

The present study examines the use of classifiers and gender markers in grammatical nominalizations functioning as the heads of noun phrases and as noun modifiers in Sinhala. The data for analysis were collected from the internet and their accuracy was verified by native Sinhala consultants. By examining the usage of classifiers and gender markers—both those typically regarded as correct and those regarded as incorrect—in contemporary Sinhala, we aim to understand the distribution and diffusion hierarchy of classifier and gender marking in the language. The data ultimately support Shibatani's (2021: 524, 2023: 35) hierarchy (NUM > DEM > GEN > ADJ > V-based) for classifier/gender marking.

## 1. INTRODUCTION

The grammatical nominalizations in Sinhala discussed in this paper are those accompanied by classifiers or those accompanied by the nominalizer *ekə*, which can change form depending on gender. In Sinhala, verb infinitives and focus forms can also create noun clauses, but these will only be briefly introduced.

Karunatilake (1998: 13, 14) and Chandralal (2010: 60, 61) identify two types of classifiers in Sinhala: *kenā* (human) and *denā* (animate), and they point out that these classifiers grammatically behave differently. Disanayaka (2001) adds that *namə* (monk) should also be considered a classifier in Sinhala. However, all these studies focus only on classifiers used with numerals and do not analyze the use of classifiers with demonstratives, genitive nouns, adjectives, or verb-based nominalizations. This study categorizes *kenā*, *denā*, and *namə* as the three classifiers in Sinhala, examining their grammatical behaviors and highlighting the differences in their distribution and diffusion hierarchy.

Regarding gender marking, this study places more emphasis on morphology than previous research (Shibatani 2021) and exclusively examines whether gender marking alone influences the nominalization of numerals, demonstratives, genitive nouns, adjectives, and verbs in Sinhala. Our examination reveals that gender marking can be used as a nominalization device only with numerals, demonstratives, and genitive noun phrases (limited to first-person genitive pronouns) on Shibatani's hierarchy (NUM > DEM > GEN > ADJ > V-based), which form a contiguous segment.

The noun phrase structure in Sinhala centers around a head noun, which is morphologically distinguished by singular definite, singular indefinite, and plural forms. Modifiers such as demonstratives, genitives (possessives), and adjectives typically precede the head noun. However, quantifiers, including numerals, exhibit special behavior: the numeral

---

<sup>1</sup> Department of Japanese, Faculty of Letters, Yasuda Women's University, 6-13-1 Yasuhigashi, Asaminami-ku, Hiroshima 731-0153, Japan. email: miyagisi@yasuda-u.ac.jp

*ekə*<sup>2</sup> (meaning ‘one’) and quantifiers such as *goḍak* (meaning ‘many’) and *ṭikak* (meaning ‘few’) can appear both before and after the head noun, while numerals greater than one must follow the head noun. In these cases, it is the numeral or quantifier, rather than the head noun, that receives case marking. Therefore, the numeral or quantifier can also be interpreted as the head of the noun phrase.

In Sinhala grammar, nouns and adjectives are distinguished syntactically and semantically, but not morphologically, and adjectives can sometimes function as nouns without a nominalization suffix.

Singular nouns in Sinhala are morphologically distinguished between definite and indefinite forms, with the indefinite form marked by a suffix (such as *-ek* or *-ak*). There is no such distinction for plural nouns, but specific nouns can be marked using demonstratives, which can also be used with singular nouns.

Non-classifying nominalization markers in Sinhala include *-īmə*, which corresponds to the English *-ing* and is used to nominalize verbs (e.g., *kiyəwə-nəwā* [meaning ‘to read’] > *kiyəw-īmə* [meaning ‘reading’]), and *-kəmə*, which corresponds to the English *-ness* and is used to nominalize adjectives (e.g., *kammæli* [meaning ‘lazy’] > *kammæli-kəmə* [meaning ‘laziness’]). As mentioned earlier, adjectives can also be nominalized without the use of a nominalizing suffix.

Regarding the syntactic devices for forming noun phrases in Sinhala, as previously mentioned, modifier phrases always precede the head noun. In the case of long noun phrases, such as *the destruction of Buddhists by Brahmins* in English, the modifiers are placed before the noun in Sinhala, resulting in a different word order (e.g., *brāhməṇayan wisin bauddhəyan wināśə kirīmə* Lit. ‘Brahmins by Buddhist’s destruction’). Furthermore, Sinhala does not have relative pronouns; instead, it uses the verb in its participle form (relative form) to modify the noun. The verb precedes the head noun and serves to modify it (e.g., *kathā kəṛəṇə ayek* can be literally translated as ‘talking person’, which corresponds to the English ‘the person who is talking’, but in Sinhala, relative pronouns are not used).

Note that the Sinhala data analyzed in this paper was romanized from data obtained by searching in Sinhala script on Google. The validity of the data was then verified by three native speakers of Sinhala.

---

<sup>2</sup> In Sinhala, the vowel /a/ can be pronounced either as [a] or as [ə]. In general, /a/ is clearly pronounced as [a] when it appears in the first syllable of a word, whereas in medial or final positions, it often weakens to [ə]. However, there are exceptions—for example, in the verb */karanawa/* [kəṛənəwə], meaning ‘to do’, the first syllable /kə/ is pronounced as [kə]. Additionally, even in medial or final positions, /a/ is not weakened and remains clearly pronounced as [a] in certain phonological environments, such as after /h/, before /yi/, or within closed syllables.

## 2. OVERVIEW OF CLASSIFIER AND GENDER MARKING AND GRAMMATICAL NOMINALIZATIONS IN SINHALA

### 2.1 Classifier and gender marking in Sinhala

In this section, I will provide an overview of the nominal classification system in Sinhala.

#### 2.1.1 Noun classification based on animacy

In Indo-Aryan languages such as Marathi, Konkani, Gujarati (which have a three-gender system, namely masculine, feminine, neuter like Sanskrit) and Hindi, Panjabi, Sindhi (which have a two-gender system, namely masculine and feminine), gender is a classificatory property of all nouns, whether animate or inanimate. However, the gender system in Sinhala has shifted to a “natural” gender system, as in the neighbouring Dravidian languages (see Masica 1991: 219-223). In Sinhala (which also is an Indo-Aryan language), the distinction between animate and inanimate nouns is significant both morphologically and syntactically. For example, all animate nouns take the genitive case marker *-ge*, while inanimate nouns do not. Furthermore, all animate nouns can optionally take the accusative case marker *-wə*, whereas inanimate nouns cannot. Further, animates and inanimates select different existential verb (*innəwā* vs. *tiyenəwā*) and numeral forms (Masica 1991: 221). According to Karunatilake (1998: 399-403), Sinhala nouns, both regular and irregular, are classified into five major categories, as follows:

#### [1] Regular nouns: Class I to IV inanimate, class V animate

**Class I** nouns end in *-əkə* in the direct case definite singular and *-ekē* in the genitive case definite singular:

kār-əkə ‘the car’  
kār-ekē ‘of the car’

**Class II** nouns end in *-ə*, *-a*, or *-æ* in the direct case definite singular and *-ē* or *-e* in the genitive case definite singular:

|                    |                    |                  |
|--------------------|--------------------|------------------|
| potə ‘the book’    | gaha ‘the tree’    | æhæ ‘the eye’    |
| potē ‘of the book’ | gahē ‘of the tree’ | æhē ‘of the eye’ |

**Class III** nouns end in *-ə* in both the direct and genitive case definite singular:

gedərə ‘the home’  
gedərə ‘of/at the home’

**Class IV** nouns end in *-ē* or *-e* in both the direct and genitive case definite singular:

|                      |                    |
|----------------------|--------------------|
| kūḍe ‘the basket’    | kaḍē ‘the shop’    |
| kūḍe ‘of the basket’ | kaḍē ‘of the shop’ |

**Class V:** All animate nouns are grouped under this class. These take *-ge* as the genitive case marker.

## [2] Irregular nouns:

Irregular nouns include *dē* ‘thing’, *gē* ‘house’, *rē* ‘night’, and so on.

|                         |                        |                   |
|-------------------------|------------------------|-------------------|
| dē ‘the thing’          | gē ‘the house’         | rē ‘the night’    |
| no genitive form for dē | geyi ‘of/at the house’ | rē ‘of the night’ |

All nouns other than those in Class V are inanimate (Karunatillake 1998: 399).

### 2.1.2 Noun classification based on gender marking

Animate nouns, such as those referring to occupations, ranks, animals, and given names, can be differentiated into masculine (including epicene) and feminine forms using different affixes.

The methods of distinguishing biological gender in Sinhala animate nouns can be morphologically classified into three types: suppletion, gender-specific suffixes, and the use of masculine or feminine nouns in part (Gunasekara 1891: 71-77).

#### [1] Suppletion

Quite different words are used to distinguish masculine and feminine nouns as follows:

| <u>Masculine</u>      | <u>Feminine</u>     | <u>Masculine</u> | <u>Feminine</u>  |
|-----------------------|---------------------|------------------|------------------|
| ayiyā ‘elder brother’ | akkā ‘elder sister’ | aśwāyā ‘horse’   | welaṃbā ‘mere’   |
| mahātmāyā ‘gentleman’ | nōnā ‘lady’         | goṇā ‘ox’        | denā ‘cow’       |
| rajā ‘king’           | bisāwa ‘queen’      | monārā ‘peacock’ | sebāḍā ‘pea-hen’ |

#### [2] gender-specific suffixes

The way gender is distinguished through gender-specific suffixes can be further divided into three patterns.

(a) A pattern in which the masculine and feminine suffixes at the word’s end are replaced. The feminine suffixes are more numerous than the masculine ones, and they all share the common feature of containing the vowel *i*, which simply contrasts with the vowel *a* found in masculine suffixes.

Masculine suffixes: -ā, akə(yā)

Feminine suffixes: -ættī, -ætti, -ī, -i, -inī, -ini, -innī, -inni, -issī, -issi, -ikā(wə), -ittī, itti, -iccī, -icci, -nī, -hinnī, -hinni.

The following examples illustrate a pattern in which only the masculine and feminine suffixes are replaced.

| <u>Masculine</u>                 | <u>Feminine</u>                     |
|----------------------------------|-------------------------------------|
| muwā (deer)                      | muwættī (doe)                       |
| taruṇəyā (young man)             | taruṇī (young woman)                |
| yakā (demon)                     | yakinnī (female demon)              |
| roḍiyā (man of the Rodiya caste) | roḍissī (woman of the Rodiya caste) |
| kumārəyā (prince)                | kumārikā(wə) (princess)             |
| mæḍiyā (frog)                    | mæḍittī (female frog)               |
| parəyā (out-caste man)           | pariccī (out-caste woman)           |
| muttā (grandfather)              | muttəṇī (grandmother)               |

The following examples illustrate a pattern in which both the root vowel changes, and the replacement of gender suffixes occur.

| <u>Masculine</u> | <u>Feminine</u>         | <u>Masculine</u> | <u>Feminine</u> |
|------------------|-------------------------|------------------|-----------------|
| nayā (cobra)     | nəyiṇī (female cobra)   | kukuḷā (cock)    | kikiḷī (hen)    |
| waləhā (bear)    | wələhinnī (female bear) |                  |                 |
| kokā (crane)     | kekī (female crane)     |                  |                 |

(b) A pattern in which the vowel of a masculine noun, particularly the initial and final vowels, changes, without the presence of the feminine suffix -ī/-i.

| <u>Masculine</u> | <u>Feminine</u>     | <u>Masculine</u> | <u>Feminine</u>   |
|------------------|---------------------|------------------|-------------------|
| kollā (lad)      | kellə (lass)        | korā (lame man)  | kerə (lame woman) |
| sorā (thief)     | serə (female thief) |                  |                   |

(c) A pattern in which agentive suffixes that nominalize verbs, originally in the masculine forms -ā/-ē, change to the feminine form -i.

| <u>Verb</u>              | <u>Masculine noun</u>                              | <u>Feminine noun</u>          |
|--------------------------|----------------------------------------------------|-------------------------------|
| duwənəwa (to run)        | duwannā/duwannē/duwənuē (runner)                   | duwannī (female runner)       |
| kanəwa (to eat)          | kannā/kannē/kanuē (eater)                          | kannī (female eater)          |
| āṇḍukəṛənəwa (to govern) | āṇḍukəṛannā/āṇḍukəṛannē/<br>āṇḍukəṛənuē (governor) | āṇḍukəṛannī (female governor) |



[3] The use of masculine or feminine nouns in part

This group also includes patterns in which masculine or feminine nouns are added to adjectives or nouns used adjectivally, and patterns in which feminine nouns are added to the crude forms of masculine nouns.

#### Masculine

ingrīsikārəyā [ingrīsi+kārəyā] (Englishman)  
mīharəkā [mī (buffalo) + harəkā (ox)]

#### Feminine

ingrīsikārī [ingrīsi kārī] (English woman)  
mīharəkī [mī (buffalo) + harəkī (cow)]

#### Masculine

aśwəyā (horse)  
eḷuwā (goat)  
mīyā (rat)  
sinhayā (lion)

#### Feminine [crude form of Masculine nouns + female nouns]

aśwəweḷaṁbā [aśwə + weḷaṁbā] (mare)  
eḷudenā [eḷu + denā] (cow)  
mīdenā [mī + denā] (female rat)  
sinhadhēnuwā [sinha + dhēnuwā] (lioness)

### 2.1.3 Sinhala classifiers

According to previous literature (Karunatilake 1998:13-14, Chandralal 2010: 60-61) Sinhala has two sortal classifiers: *kenā* and *denā*. *Kenā* is used when counting people, specifically with the numeral ‘one’ (1). In contrast, *denā* is used for counting both people and animals and is used with numerals greater than one (i.e., ‘two’ or more). Both *denā* and *kenā* are treated as nouns and pronouns in general Sinhalese dictionaries (Carter 1924, de Zoysa 2011, Wijayatunga 2012). They each have indefinite forms: *denek* and *kenek*, which are formed by adding the indefinite suffix *-ek*, like regular count nouns. However, unlike typical count nouns, neither has a plural form.

In my opinion, the two sortal classifiers mentioned above should be supplemented with *namə* and *namak*, which are used to classify monks. Dissanayake (2001) explains that when counting Buddhist monks, the classifiers *denā* and *denek* are replaced by the classifiers *namə* and *namak*.

In Sinhala, as in other languages, mensural classifiers include container nouns like *kōppə* (cup), *petti* (box), *gōni* (bag), and *bōtal* (bottle); shape nouns like *kāli* (piece) and *bindu* (drop); collective nouns like *kuṭṭam* (group or set), *kaṇḍāyam* (team), and *miṭi* (bundle); and unit nouns like *grām* (gram), *mīṭərə* (meter), *līṭərə* (liter), *siyəyətə* (percent), *wināḍi* (minute), *rupiyal* (rupee), and *warə* (times).

In Sinhala, sortal classifiers and mensural classifiers differ not only semantically but also in their grammatical behavior: with sortal classifiers, the numeral precedes the classifier, whereas with mensural classifiers, the numeral follows the classifier. Note that the numerals that precede sortal classifiers are bases, and the numerals and classifiers are fused together. The structure of numerals will be discussed later.

### Sortal classifiers

|                |                         |
|----------------|-------------------------|
| ek-kenā        | tun-denā                |
| one-CLF.person | three CLF.person/animal |
| ‘one person’   | ‘three persons/animals’ |

### Mensural classifiers

|           |                |
|-----------|----------------|
| kōppə ekə | kāli tunə      |
| cup one   | pieces three   |
| ‘one cup’ | ‘three pieces’ |

## 2.2 Grammatical nominalizations in Sinhala

This section will look at how numerals, demonstratives, genitive forms of nouns, adjectives, and verbs in Sinhala are nominalized.

### [1] Numerals

The numerals in Sinhala, both cardinal and ordinal, are based on the numeral base. The numeral base does not function as a noun on its own; when used as a noun, it must be followed by additional elements added to the base.

#### [1a] Cardinal numerals

In Sinhala, cardinal numbers become definite nouns through base forms and gender-distinguishing affixes and further become indefinite nouns through affixes indicating indefiniteness. Animate numerals are divided into the epicene form, which attaches only to the numeral bases one and two, and the feminine form, which attaches only to the numeral base one. Table 1 shows the Sinhala cardinal numerals from 1 to 10.

Table 1. Examples of cardinal numerals in Sinhala

|    | Base <sup>3</sup> | Inanimate  |              | Animate  |            |          |            |
|----|-------------------|------------|--------------|----------|------------|----------|------------|
|    |                   |            |              | Epicene  |            | Feminine |            |
|    |                   | Definite   | Indefinite   | Definite | Indefinite | Definite | Indefinite |
| 1  | ek                | ekə        | ekak         | ekā      | ekek       | ekī      | ekiyak     |
| 2  | de                | dekə       | dekak        | dekā     | dekek      |          |            |
| 3  | tun               | tunə       | tunak        |          |            |          |            |
| 4  | hatəṛə/<br>satəṛə | hatəṛə     | hatərak      |          |            |          |            |
| 5  | pas               | pasə(paha) | pasak(pahak) |          |            |          |            |
| 6  | hayə              | hayə       | hayak        |          |            |          |            |
| 7  | hat               | hatə       | hatak        |          |            |          |            |
| 8  | aṭə               | aṭə        | aṭak         |          |            |          |            |
| 9  | nawə              | nawəyə     | nawəyak      |          |            |          |            |
| 10 | daha              | dahayə     | dahayak      |          |            |          |            |

<sup>3</sup> According to Chandralal (2010: 59), *hatəṛə/satəṛə*, *hayə*, and *aṭə* are regarded as numeral bases.

- (1) *eka-yi deka-yi ekətu kəḷā-mə tuna-yi.*  
 one-and two-and add-TEMP three-AM  
 ‘One and two add up to three.’

(<https://www.facebook.com/groups/883206060190881/posts/1159377019240449/>)

- (2) *karuṇākərə maṭə pah-ak de-nnə!*  
 Please 1.SG.DAT five-INDF give-INF  
 ‘Please give me five. (e. g. apples, pens, books, etc.)’

(<https://si.opentran.net/සිංහල-ජාති-වැඩි-ලංසුවක්-දෙන්න.html/>)

- (3) *ek-ek āwā de-kek āwā. turul kəran hiṭiyā*  
 one-EPI.INDF come.PP two-EPI.INDF come.PST snuggle.PP stay.PST  
 ‘One came, then two came. They were hugging.’

([https://www.youtube.com/watch?v=HHgC6D3i\\_js/](https://www.youtube.com/watch?v=HHgC6D3i_js/))

Additionally, searching the online data reveals examples of the epicene form of the numeral ‘three’. However, these examples are judged to be incorrect by native Sinhala speakers I consulted. It should also be noted that new innovations are rejected by some native speakers but they forget that those who are using them are also native speakers.

- (4) a. *mē nars-la \*tunā ekə ekə vidiy-e fan ga-nnə ayə lu*  
 this nurse-PL three.EPI one one way-LOC fun get-INF person EVD  
*hoṇḍə da?*  
 good Q  
 ‘Is it okay for these three nurses to have fun in different ways?’

(<https://elakiri.com/threads/නර්ස්ලා-තුන්දෙනා.822401/>)

- b. *apē set-ek ek-ek neveyi dek-ek \*tun-ek innəwā*  
 1.PL.GEN set-LOC one-EPI.INDF not two-EPI.INDF three-EPI.INDF stay  
 ‘There is not one but two or three in our group.’

([https://www.facebook.com/story.php?story\\_fbid=819052054962712&id=403815089819746&\\_rdr/](https://www.facebook.com/story.php?story_fbid=819052054962712&id=403815089819746&_rdr/))

### [1b] Ordinal numerals

In Sinhala, ordinal numbers are formed from the numeral base and the affix *weni*. The nominalization of ordinal numbers is marked by the addition of different affixes for inanimate and animate nouns, with each further distinguishing between definite and indefinite forms. Only the ordinal number one uses a special form, different from the numeral base. Table 2 shows the Sinhala ordinal numerals from 1 to 10.

Table 2. Examples of ordinal numerals in Sinhala

|    | Base   | Ordinal adjective | Noun inanimate |               | Noun animate |               |
|----|--------|-------------------|----------------|---------------|--------------|---------------|
|    |        |                   | Definite       | Indefinite    | Definite     | Indefinite    |
| 1  | ek     | paḷəweni          | paḷəweniyə     | paḷəweniyak   | paḷəweniyā   | paḷəweniyek   |
| 2  | de     | deweni            | deweniyə       | deweniyak     | deweniyā     | deweniyek     |
| 3  | tun    | tunweni           | tunweniyə      | tunweniyak    | tunweniyā    | tunweniyek    |
| 4  | hatəṛə | hatəṛəweni        | hatəṛəweniyə   | hatəṛəweniyak | hatəṛəweniyā | hatəṛəweniyek |
| 5  | pas    | pasweni           | pasweniyə      | pasweniyak    | pasweniyā    | pasweniyek    |
| 6  | hayə   | hayəweni          | hayəweniyə     | hayəweniyak   | hayəweniyā   | hayəweniyek   |
| 7  | hat    | hatəweni          | hatəweniyə     | hatəweniyak   | hatəweniyā   | hatəweniyek   |
| 8  | aṭə    | aṭəweni           | aṭəweniyə      | aṭəweniyak    | aṭəweniyā    | aṭəweniyek    |
| 9  | nawə   | nawəyweni         | nawəyweniyə    | nawəyweniyak  | nawəyweniyā  | nawəyweniyek  |
| 10 | daha   | dahaweni          | dahaweniyə     | dahaweniyak   | dahaweniyā   | dahaweniyek   |

- (5) samhitānukūla subhāraṁci satə-en tunweniyə we-yi  
 canonical gospel.PL four-ABL third COP-AM  
 ‘It is the third of the four canonical Gospels.’

(<https://www.wikiwand.com/si/search?q=ජේසුස්+වහන්සේගේ+උන්වනයි/>)

- (6) paḷəweniyā-tə ættəṭəmə ādərē kərā nam kawədəwat deweniy-ek  
 first.ANMT-DAT truly love.PST TOP never second.ANMT-INDF  
 enne næ.  
 come.FOC not  
 ‘If you truly love the first person, there will never be a second one.’

(<https://www.tiktok.com/@nimnawi6356/video/7385163557400349970/>)

## [2] Demonstratives

In Sinhala, demonstratives are classified into four categories: first-person proximal, second-person proximal, distal, and anaphoric as shown in Table 3. The base forms of demonstratives function as both adjectives, modifying nouns, and as nouns themselves. In this case, singular and plural forms are not morphologically distinguished. The base form becomes a noun when affixed. Affixes are distinguished based on categories such as inanimate, human, animate masculine, animate feminine, and animal, with each category having both singular and plural forms.

Table 3. Demonstratives in Sinhala

|                                 |    | 1 <sup>st</sup> person<br>Proximal | 2 <sup>nd</sup> person<br>Proximal | Distal  | Anaphoric |
|---------------------------------|----|------------------------------------|------------------------------------|---------|-----------|
| Base (Adjectival pronoun SG/PL) |    | mē                                 | oyə                                | arə     | ē         |
| Pronoun<br>Inanimate            | SG | mēkə                               | ōkə                                | arəkə   | ēkə       |
|                                 | PL | mēwā                               | ōwā                                | arəwā   | ēwā       |
| Human                           | SG | meyā                               | oyā                                | arəya   | eyā       |
|                                 | PL | meyāla                             | oyāla                              | arəyāla | eyāla     |
| Masculine<br>Human, Animal      | SG | mēkā                               | ōkā                                | arəkā   | ēkā       |
|                                 | PL | mēkālā                             | ōkālā                              | arəkālā | ēkālā     |
| Feminine<br>Human, Animal       | SG | mēkī                               | ōkī                                | arəkī   | ēkī       |
|                                 | PL | mēkīlā                             | ōkīlā                              | arəkīlā | ēkīlā     |

- (7) a. mē potə hoñda-yi.  
this book good-AM  
'This book is good.'

(<https://baduraasapuwa.blogspot.com/2012/06/blog-post.html/>)

- b. mē magē potə.  
this 1.SG.GEN book.SG  
'This is my book.'

([https://grantha.lk/cartquickpro/catalog\\_product/view/id/16705/](https://grantha.lk/cartquickpro/catalog_product/view/id/16705/))

- c. mē magē pot.  
these 1.SG.GEN book.PL  
'These are my books.'

(<https://www.languagepassion.com/learn-english-through-sinhala/>)

- (8) mēkə hoñdə wəɖ-ak.  
this good work-INDF  
'This is good work.'

([https://www.youtube.com/watch?v=5VCHFk\\_aDew/](https://www.youtube.com/watch?v=5VCHFk_aDew/))

- (9) meyā hoñdə yāluw-ek.  
this.person good friend.INDF  
'This is a good friend.'

(<https://www.youtube.com/shorts/OUIml1yZSLQ?app=desktop/>)

- (10) mēkā hoñdə tṛōjan assay-ek.  
this.animal good Trojan horse-INDF  
'This is a good Trojan horse.'

(<https://www.lankadeepa.lk/Caroon/2019-05-07/39-549979/>)

- (11) mēkī me-hemə hiṭiyə kell-ek newē ban.  
 this.F this-like be.PST.VA girl-INDF not EM  
 ‘This is not a girl like this!’

(<https://www.facebook.com/1660214864296548/posts/2746617042322986/>)

- (12) mēkī gəni-yak da?  
 this.F woman-INDF Q  
 ‘Is this a woman?’

(<https://www.youtube.com/watch?v=l4SLUvBiVu0/>)

Table 4 illustrates the case markings of the singular demonstratives in Sinhala.

Table 4. Singular demonstratives

|                                    | Gender    | NOM   | ACC      | DAT      | LOC/GEN  | ABL/INST  |
|------------------------------------|-----------|-------|----------|----------|----------|-----------|
| 1 <sup>st</sup> person<br>Proximal | Inanimate | mēkə  | mēkə     | mēkə-ṭə  | mēke     | mēken     |
|                                    | Human     | meyā  | meyā-wə  | meyā-ṭə  | meyā-ge  | meyā-gen  |
|                                    | Epicene   | mēkā  | mēkā-wə  | mēkā-ṭə  | mēkā-ge  | mēkā-gen  |
|                                    | Feminine  | mēkī  | mēkī-wə  | mēkī-ṭə  | mēkī-ge  | mēkī-gen  |
| 2 <sup>nd</sup> person<br>Proximal | Inanimate | ōkə   | ōkə      | ōkə-ṭə   | ōke      | ōken      |
|                                    | Human     | oyā   | oyā-wə   | oyā-ṭə   | oyā-ge   | oyā-gen   |
|                                    | Epicene   | ōkā   | ōkā-wə   | ōkā-ṭə   | ōkā-ge   | ōkā-gen   |
|                                    | Feminine  | ōkī   | ōkī-wə   | ōkī-ṭə   | ōkī-ge   | ōkī-gen   |
| Distal                             | Inanimate | arəkə | arəkə    | arəkə-ṭə | arəke    | arəken    |
|                                    | Human     | arəya | arəya-wə | arəya-ṭə | arəya-ge | arəya-gen |
|                                    | Epicene   | arəkā | arəkā-wə | arəkā-ṭə | arəkā-ge | arəkā-gen |
|                                    | Feminine  | arəkī | arəkī-wə | arəkī-ṭə | arəkī-ge | arəkī-gen |
| Anaphoric                          | Inanimate | ēkə   | ēkə      | ēkə-ṭə   | ēke      | ēken      |
|                                    | Human     | eyā   | eyā-wə   | eyā-ṭə   | eyā-ge   | eyā-gen   |
|                                    | Epicene   | ekā   | ekā-wə   | ekā-ṭə   | ekā-ge   | ekā-gen   |
|                                    | Feminine  | ēkī   | ēkī-wə   | ēkī-ṭə   | ēkī-ge   | ēkī-gen   |

- (13) a. mamə mēkə aran yaṇəwā.  
 1.SG this take.PP go  
 ‘I’m taking this.’

(<https://gossip.hirufm.lk/95412/2023/10/hot-story.html/>)

- b. mamə mēkə-ṭə sambandhə nē.  
 1.SG this-DAT relation no  
 ‘I am not involved in this.’

(<https://videos.gosiplankanews.com/2023/05/manju-nishanka.html/>)

- c. mēke āsamə koṭasə mok-ak da?  
 this.GEN favorite part what-INDF Q  
 ‘What’s your favorite part of this?’  
 (<https://www.instagram.com/isurudeshan.music/reel/DBq7eG0IJ-B/>)
- d. api mēk-en æt wenəwā.  
 3.PL.NOM this-ABL move away  
 ‘We are moving away from this.’  
 (<https://www.aruna.lk/videos/pIpY0ZzKpSk0wyKTxCTX/>)
- (14) a. mamə meyā-wə dækkā.  
 1.SG.NOM this.person-ACC see.PST  
 ‘I saw him.’  
 (<https://www.youtube.com/watch?app=desktop&v=OpDPywd2-UU/>)
- b. mamə meyā-ṭə demələ uganəwa-nnam.  
 1.SG.NOM this.person-DAT Tamil teach-HORT  
 ‘I will teach him Tamil.’  
 (<https://glosbe.com/es/si/tamil/>)
- c. mamə kawədāwat meyā-gē fōn ekə balannē nā.  
 1.SG.NOM never this.person-GEN phone look.FOC no  
 ‘I never look at his phone.’  
 (<https://www.facebook.com/watch/?v=591402980138140/>)
- d. mamə meyā-gen ætwelā hiṭiyē.  
 1.SG.NOM this.person-ABL leave.PP stay.FOC  
 ‘I was staying away from him.’  
 (Saaravita.lk|| මෙය ආශ්‍රය කරනු ලබන බර්ට් හරි නැ.. එයාගේ කට පොඩ්ඩක් හරි නැ.. මම කතා කරනවා දැක්කොත් බනිනවා)
- (15) a. mamə mēkī-wə kasādə baṇḍinəwā.  
 1.SG.NOM this.person.F-ACC marry  
 ‘I will marry this girl.’  
 (<https://gossip.hirufm.lk/47968/2020/06/nilangas-story.html/>)
- b. mamə mēkī-ṭə kəmati nā.  
 1.SG.NOM this.person.F-DAT like no  
 ‘I don’t like this girl.’  
 (<https://www.wattpad.com/1396720647-has-changed-taekook-15/>)

c. mēkī-gē daruwā magē nemeyi  
 this.F-GEN kid 1.SG.GEN not  
 ‘This woman’s kid is not mine.’

(<https://www.youtube.com/watch?v=KBjWi9YpLhY/>)

d. ehenam mēkī-gen ma paṭan ga-mu.  
 so this.person.F-ABL EM start-HORT  
 ‘So let’s start with this girl.’

(<https://telegra.ph/සලහක-වල-10-21-2/>)

Table 5 illustrates the case markings of the plural demonstratives in Sinhala.

Table 5. Plural demonstratives

|                                    | Gender    | NOM     | ACC        | DAT        | LOC/GEN    | ABL/INST    |
|------------------------------------|-----------|---------|------------|------------|------------|-------------|
| 1 <sup>st</sup> person<br>Proximal | Inanimate | mēwā    | mēwā       | mēwā-ṭə    | mēwā-walə  | mēwā-walin  |
|                                    | Human     | meyāla  | meyāla-wə  | meyāla-ṭə  | meyāla-ge  | meyāla-gen  |
|                                    | Epicene   | mēkālā  | mēkālā-wə  | mēkālā-ṭə  | mēkālā-ge  | mēkālā-gen  |
|                                    | Feminine  | mēkīlā  | mēkīlā-wə  | mēkīlā-ṭə  | mēkīlā-ge  | mēkīlā-gen  |
| 2 <sup>nd</sup> person<br>Proximal | Inanimate | ōwā     | ōwā        | ōwā-ṭə     | ōwā-walə   | ōwā-walin   |
|                                    | Human     | oyā     | oyā-wə     | oyā-ṭə     | oyā-ge     | oyā-gen     |
|                                    | Epicene   | ōkālā   | ōkālā-wə   | ōkālā-ṭə   | ōkālā-ge   | ōkālā-gen   |
|                                    | Feminine  | ōkīlā   | ōkīlā-wə   | ōkīlā-ṭə   | ōkīlā-ge   | ōkīlā-gen   |
| Distal                             | Inanimate | arəwā   | arəwā      | arəwā-ṭə   | arwā-walə  | arə-walin   |
|                                    | Human     | arəyāla | arəyāla-wə | arəyāla-ṭə | arəyāla-ge | arəyāla-gen |
|                                    | Epicene   | arəkālā | arəkālā-wə | arəkālā-ṭə | arəkālā-ge | arəkālā-gen |
|                                    | Feminine  | arəkīlā | arəkīlā-wə | arəkīlā-ṭə | arəkīlā-ge | arəkīlā-gen |
| Anaphoric                          | Inanimate | ēwā     | ēwā        | ēwā-ṭə     | ēwā-walə   | ēwā-walin   |
|                                    | Human     | eyāla   | eyāla-wə   | eyāla-ṭə   | eyāla-ge   | eyāla-gen   |
|                                    | Epicene   | ekālā   | ekālā-wə   | ekālā-ṭə   | ekālā-ge   | ekālā-gen   |
|                                    | Feminine  | ēkīlā   | ēkīlā-wə   | ēkīlā-ṭə   | ēkīlā-ge   | ēkīlā-gen   |

(16) a. ēwā tuppəhi wəḍə.  
 they.INAM hippie work.PL  
 ‘They are stupid things.’

(<https://www.youtube.com/watch?v=DR3nAB8nr8I/>)



b. mamə ēwā labā dunnā.

1.SG.NOM they.INAM give.PST

‘I gave them.’

(<https://www.heladiva.lk/බව්-මම-ඒවා-ලබා-දුන්නා-රනි/>)

c. mamə ēwā-ʔə bayə wunā.

1.SG.NOM they.INAM-DAT afraid become.PST

‘I was afraid of them.’

(<https://www.eethalayanews.com/single.php?id=88/>)

d. ēwā-wələ āyōjəṇəyə karā-ʔə kama-k nā-lu.

They.INAM-LOC investigation do.PST-DAT problem-INDF no-EVD

‘It seems that investing in them is not a problem.’

(<https://lankacnews.com/අමුල්-සමාගමට-දෙන-ඉඩම්-තැක/comment-page-3/>)

e. api-ʔə ēwāwal-in pənəlā ya-nnə bā.

1.PL-DAT they.INAM-ABL escape.PP go-INF impossible

‘We can’t escape from them.’

([https://www.tiktok.com/@sishy\\_\\_\\_\\_y/video/7445671288510876936/](https://www.tiktok.com/@sishy____y/video/7445671288510876936/))

### [3] Genitive noun phrases

The genitive form of Sinhala nouns differs between animate and inanimate nouns. For animate nouns, the suffix *-ge* is added to the end of the word. In the case of inanimate nouns, the form varies between singular and plural: the singular form takes *-e*, while the plural form takes *-walə*.

- (17) a. ʎaməyā-ge sellam baḍu  
child-GEN playing things  
‘the child’s toys’

(<https://www.coaccess.com/si/blog/picnic/>)

- b. ʎamayin-gē sellam baḍu  
child.PL-GEN playing thing.PL  
‘the children’s toys’

(<https://www.terafundchina.com/si/kids-ride-on-for-kids-toy-bp600-product/>)

- (18) a. pāsəl-ē mēsə  
school-GEN desk.PL  
‘the school’s desks’

(<https://www.sikuradapasala.com/blog/>)

- b. pāsal-wələ      lamayi  
 school.PL-GEN child.PL  
 ‘children of the schools’

(<https://www.gossi plankanews.com/2019/05/100less-school.html/>)

The Sinhala pronouns, except for the first-person plural pronoun which ends in *-e*, all other pronouns end with *-ge* as shown in table 6.

Table 6. Genitive forms of pronouns

|                             | SG          | PL              |
|-----------------------------|-------------|-----------------|
| 1 <sup>st</sup>             | magē        | apē             |
| 2 <sup>nd</sup>             | oyāgē/obāgē | oyālāgē/obālāgē |
| 3 <sup>rd</sup> (Epicene)   | eyāgē       | eyālāgē         |
| 3 <sup>rd</sup> (Masculine) | ohugē       | owungē          |
| 3 <sup>rd</sup> (Feminine)  | æyāgē       | æyālāgē         |

- (19) a. magē      sellam      baḍu  
 1.SG.GEN      playing      thing.PL  
 ‘my toys’

(<https://www.youtube.com/watch?v=v1f95tnnSDM/>)

- b. apē      sellam      baḍu  
 I.PL.GEN      playing      thing.PL  
 ‘our toys’

(අපේ සෙල්ලම් බඩු 01 | Dejaya Lanka | Facebook/)

- c. eyā-ge      sellam      baḍu  
 3.SG.GEN      playing      thing.PL  
 ‘his/her toys’

(<https://wol.jw.org/si/wol/d/r125/lp-sn/2010327/>)

The genitive form of Sinhala nouns does not require a noun to modify, and can stand alone to form a noun phrase.

- (20) a. ‘mē! mē! mēkə magē.’      ‘nāe nāe mēkə ayiti      maṭa-yi.’  
 this this this      1.SG.GEN      no no this      belong      1.SG.DAT-AM  
 ‘This! This! This is mine’      ‘No, no, this belongs to me’

([https://archives.dinamina.lk/sithijaya/art.asp?id=2011/10/27/spg15\\_0/](https://archives.dinamina.lk/sithijaya/art.asp?id=2011/10/27/spg15_0/))

- b. warəda ē laməyā-gē nemeyi.  
 fault that child-GEN not  
 ‘The fault is not that child’s (fault).’  
 (<https://www.facebook.com/watch/?v=1812286492555365/>)
- c. mēkə oyā lōkal mārkaṭ e-kē nemeyi. oyā forīn mārkaṭ e-kē.  
 this 2.SG local market-GEN not 2.SG local market-GEN  
 ‘This is not your local market’s (goods), but your foreign market’s (goods).’  
 ([https://www.instagram.com/luna\\_peech/reel/DEFm6LIzM\\_W/](https://www.instagram.com/luna_peech/reel/DEFm6LIzM_W/))
- d. ē ayə apē ēriyā ek-ē nemeyi.  
 that person 1.PL-GEN area-GEN not  
 ‘That person is not our area’s. (That person does not belong to our area.)’  
 (<https://www.bbc.com/sinhala/articles/cerzw011e2mo/>)
- e. eyā mage ṭīm ek-ē nemeyi.  
 3.SG 1.SG.GEN team-GEN not  
 ‘He’s not my team’s. (He’s not on my team.)’  
 (<https://www.youtube.com/channel/UCBigAOAtQOEUYa28bAx0cfw/>)
- f. ē fēsbuk giṇumə æyā-gē nemeyi.  
 that Facebook account 3.SG.F-GEN not  
 ‘That Facebook account isn’t hers.’  
 (<https://www.deshaya.lk/article/43/features/16790/හම්බන්තොට-පුරපතිගේ-නුගේගොඩ-/>)

#### [4] Adjectives

The adjectives in Sinhala do not inflect according to the number, gender, or case of the noun they modify<sup>4</sup>. The nominalization of adjectives in Sinhala can occur either when the adjective forms a noun phrase without any affixes, retaining its original form, or when it is accompanied by an affix. Examples (21a) and (21b) show the Sinhala adjective *barə* (meaning ‘heavy’) used as a modifier and as a predicate, respectively, while example (21c) shows the adjective being used as the subject noun. (21d) is an example where the adjective *barə* is combined with the nominalizing suffix *-kamə*, which is used to abstractly express the state or quality conveyed by the adjective. (21e) and (21f) show examples where numeral nouns function as nominalizers for adjectives.

<sup>4</sup> In Hindi and Marathi some adjectives decline as per the number, gender and case of the noun they modify and some do not. Sinhala seems to have done away with the agreement between adjective and noun.

- (21) a. *barə bæg rægenə yāmə-ṭə udaw kəra-nnə.*  
 heavy bag.PL carry.PP going-DAT help-INF  
 ‘Help carry heavy bags.’  
 (<https://abovethehimalaya.com/si/අන්තපූරණ-වටයේ-දින-7-කෙටි-ගමන.html/>)
- b. *mē bæg bara-yi.*  
 this bag.PL heavy-AM  
 ‘These bags are heavy.’  
 (බැග් නිර්මාණ කලාවේ අළුත් දේවල් සමගින් ජීවිතය දිනූ ගෝමිකා – Colombo Journal)
- c. *æyə-gē barə pramāṇəy-en grām 15-k-in aḍu wiya.*  
 she-GEN weight scale-INST grams 15-INDF-ABL drop.PST  
 ‘Her weight dropped by 15 grams on the scale.’  
 (<https://samsungmagazine.eu/si/2025/03/03/srovnani-galaxy-a56-a-galaxy-a55/>)
- d. *jīwitē barəkəmə saha ādərəyə*  
 life.GEN heaviness and love  
 ‘the burden of life and love’  
 (<https://www.tiktok.com/@nethmianu12/video/7468133923512470792/>)
- e. *bayik ek-ak gannəwā nam hoṇḍə barə ekak ga-nnə.*  
 bike one-INDF buy TOP good heavy one-INDF buy-INF  
 ‘If you buy a bike, buy one that is a good weight.’  
 (<https://elakiri.com/threads/finally-im-gonna-buy-the-new-pulsar-220.640148/page-4/>)
- f. *maṭə nam barə ekə de-nnə.*  
 1.SG.DAT TOP heavy one give-INF  
 ‘Give me the heavy one.’  
 (<https://www.facebook.com/photo.php?fbid=159847920049981&id=102324669135640&set=a.102324985802275/>)

## [5] Verb-based nominalizations

According to Shibatani (2019), verb-based nominalizations can be divided into lexical nominalization and grammatical nominalization, each of which can further be categorized into event nominalization and argument nominalization. The main purpose of this paper is to describe grammatical nominalization in Sinhala; however, I will also briefly discuss lexical nominalization of Sinhala verbs. This section will examine how nominalization is performed using the Sinhala verb *kiyāwəṇəwā* (to read) as an example.

- (22) mamə pot kiyəwənəwā  
1.SG book.PL read  
'I read books.'

(<https://www.youtube.com/watch?v=J24C00FpeqY/>)

Lexical nominalization is formed by adding suffixes to the verb root. In event nominalization, the suffix *-īmə* is added, while in argument nominalization, the suffixes *-nnā* (singular) and *-nnō* (plural) are used.

- (23) kiyəwīmə yanu buddhiyə labā gənīmə-ṭə atyəwəśyə de-yək-i  
reading called wisdom gaining-DAT essential thing-INDF-AM  
'What we call reading is essential for gaining intelligence.'

(<https://www.wijeya.lk/essays/කියවීමේ-වැදගත්/187-48457/>)

- (24) a. pot kiyəwannā lōkəyə jayə gani.  
book.PL reader.SG world conquer  
'The book reader conquers the world.'

(<https://www.youtube.com/channel/UCHsID47UtVmYIHwCIK06fVw/>)

- b. pot kiyəwannō menmə pot liyannō-da suwiśēṣī wē.  
book.PL reader.PL as book.PL writer.PL-also unique  
'Book readers and book writers are both unique.'

(<https://divaina.lk/කියවීමට-පෙරහුරුවක්/>)

#### [5a] Verb Nominalization with the Nominalizer *ekə* and its variants

Karunatillake (1998: 97) introduces *ekə* as a nominalizer. As shown in Table 1, *ekə* generally represents the numeral 'one' used to count inanimate nouns. However, its variants, such as *ekā* (Epicene) and *ekī* (Feminine), which are used to count animate nouns, also function as nominalizers. *Ekə* and its variants are used with verb modifiers. (25) is an example of event nominalization, and (26) is an example of argument nominalization.

- (25) [[raṭē taruṇəyō pot kiyəwənə] ekə] hoṇḍa-yi.  
country.GEN youth.PL book.PL read.VA NMLZR good-AM  
'It is good for the country's youth to read books.'

(<https://www.youtube.com/watch?v=WusSUJnXz6Y/>)

- (26) a. [[φ oya dē kiyəwənə] ekā] tērum gannəwā.  
that thing read.VA NMLZR.EPI understand  
'The person who reads that will understand.'

(<https://medamidula.blogspot.com/2013/05/blog-post.html/>)

- b. “magē      ḍenimə”    mamə    kiyəwənə    ekak                      tamayi    hæmədāmat  
1.SG.GEN jeans      1.SG read.VA    NMLZR.INDF    is                      always  
“‘My jeans’ is something I read all the time.’”

([https://piyapathdenna.blogspot.com/2014/03/blog-post\\_17.html/](https://piyapathdenna.blogspot.com/2014/03/blog-post_17.html/))

### [5b] Verb nominalization through the infinitive form

The infinitive form of Sinhala verbs (the *-nnə* form) can be used to create noun clauses. The infinitive is used for event nominalization as shown in (27).

- (27) mamə [pot kiyəwa-nnə] kæməti-yi.  
1.SG book.PL read-INF like-AM  
'I like to read books.'

(ප්‍රසිද්ධකාලය | Wijeya.lk)

The infinitive form in (27) can also be replaced by the dative form of the suffix *-īmə* as shown in (28).

- (28) mamə [pot kiyəwīmə-ʔə] kæməti-yi.  
1.SG book.PL reading-DAT like  
'I like to read books.'

(ඉංග්‍රීසි(English) ව්‍යාකරණයේ(Grammar) Gerunds (ක්‍රියා නාම) යනු මොනවාද? – Learn English with Amila I. Jayathissa)

### [5c] Verb nominalization through the focused form

The focused form of Sinhala verbs (the *-nne* form) can also be used to create noun clauses. This focused form is used for argument nominalization, as shown in (29).

- (29) a. [φ    pot        kiyəwannē] ugat        ayə-nē.  
           book.PL read-FOC    educated person-TAG  
           ‘It is an educated person who reads books.’

(<https://www.bbc.com/sinhala/sri-lanka-45652189/>)

- b. [mamə ɸ kiyəwannē]    tahanam kərəpu    pot.  
1.SG    read-FOC    ban.PFT.VA    book.PL  
'They are the banned books that I am reading.'

([https://www.instagram.com/hasitha\\_b.a.\\_/p/CfRLY6vtKVy/](https://www.instagram.com/hasitha_b.a._/p/CfRLY6vtKVy/))

#### [5d] Nominalization through the gerund form

The gerund form of Sinhala verbs (the *-imə* form of a verb) can also be used to create noun clauses. This focused form is used for event nominalization, as shown in (30) and (31).

- (30) [pot kiyəwīmə] hoñda-yi.  
book.PL reading good-AM  
'To read books is good.'

(<https://www.sithmalyaya.lk/article/buddhist/discussion/288/>)

- (31) koḷaṃba nagərey-ē goḍənaegīmə  
Colombo city-GEN construction  
'the construction of the city of Colombo'

(<https://biz.adaderana.lk/sri-lanka-apparel-brands-association-slaba/>)

### [5e] Relative clauses

Relative clauses in Sinhala are formed when the verb in its attributive form modifies the head noun, which is omitted in the relative clause.

- (32) [obē daruwā φ kiyəwənə] pot monəwā da?  
2.SG.GEN child read.VA book.PL what.PL Q  
'What are the books that your child read?'

(<https://play.google.com/store/apps/details?id=com.bookbot.report&hl=si/>)

- (33) [φ pot kiyəwənə] ayə nam dannəwā.  
book.PL read.VA person TOP know  
'As for the persons who read books, I know them.'

(<https://www.youtube.com/watch?v=bers4Mvjp-w/>)

### [5f] Complement clauses

Complement clauses in Sinhala are formed in various ways. (34) are examples where complement clauses are formed by the noun-modifying form of verbs.

- (34) a. [[guruḷugōmī æp ek-en nomilē pot kiyəwənə] widiyə]  
Gurulugomi app-INST free book.PL read.VA method  
'how to read books for free with the Gurulugomi app'

([https://www.youtube.com/shorts/OYlkH9i\\_x1g/](https://www.youtube.com/shorts/OYlkH9i_x1g/))

- b. [[japānə-yē kāntārə tiyenə] bawə] obə dannəwā da?  
Japan-LOC deserts exist.VA fact 2.SG know Q  
'Do you know the fact that there are deserts in Japan?'

(<https://aayubo.com/jp/Travel/aubeN83543s/> Accessed on July 21, 2024)

- c. [[salli hambə kərənə] ekə] mamə dæn nawattəlā.  
money earn.VA NMLZR 1.SG now stop.PP  
'I've stopped making money for now.'

(<https://www.lifetraveler.lk/2021/04/26/kandiah-balendra-management-style/>)

(35) is an example where a complement clause is formed by the infinitive form of a verb.

- (35) maṭə [gedərə ya-nnə] ōnə.  
 1.SG.DAT house go-INF want  
 ‘I want to go home.’

(<https://www.facebook.com/reel/1489240861700007/>)

(36) is an example where a complement clause is formed by the gerund form of a verb.

- (36) mamə [pot kiyəwīmə-tə] kəməti-yi.  
 1.SG book.PL reading-DAT like-AM  
 ‘I like reading books.’

(<https://blog.plag.ai/si/proofreading-essay-tips-to-improve-your-writing/>)

Examples (37) and (38) show where complement clauses are formed by quotative markers.

- (37) tātte maṭə [[sudu dik kalisam-ak ōnə] kiyəḷā] kiwwā.  
 Father.VOC 1.SG.DAT white long trousers-INDF want QM tell.PST  
 ‘Dad! (He) told me that he wanted white trousers.’

(<https://sinhala.asianmirror.lk/gossip/rambukkana-3/>)

- (38) ohu “səmə denā-tə ma puluwan” yi kiwwā.  
 he every CLF.ANMT-DAT EM able QM tell.PST  
 ‘He told me, “Anyone can do it”.’

(<https://www.facebook.com/groups/750730455601663/posts/879963216011719/>)

Example (39) shows where the complement clause is formed by an interrogative.

- (39) [meyā-tə layisan dunnē kawudə] dannē nā.  
 this person-DAT license give.PST.FOC who know.FOC no  
 ‘I don’t know who gave him the license.’

(<https://hiruvideos.hirufm.lk/videos-watch-2105-මෙයාට-ලයිසන්-දුන්නේ-කවුද-දන්නේ-නැ.html/>)

Example (40) shows where the complement clause is formed by a zero complementizer.

- (40) [minisun mærenəwā] mamə dækkā.  
 people die 1.SG see.PST  
 ‘I saw people die.’

(<https://entrevue.fr/si/zapping-jai-vu-des-gens-mourir-le-temoignage-glacant-de-lea-bloquee-sur-un-pont-durant-les-inondations-en-espagne/>)





b. ek-kenek                      wenuwen    nīti    nawa-nnə    bā. [NP use]  
 one-CLF.human.INDF    for                rules    change-INF    impossible  
 ‘Rules can’t be changed for the sake of just one person.’  
 (<https://www.news19.lk/laws-cannot-be-bent-for-one-person-minister-namal/>)

(42) a. apē                      minissu    ek-kenā                      şopin bæg                      tunə, hatəre                      bađu  
 1.PL.GEN    man.PL    one-CLF.human    shopping bag.PL    three four.LOC    thing.PL  
 arəgenə    yanəwā. [Mod Use]  
 take.PP    go  
 ‘Our guys will take three or four shopping bags.’  
 (<https://www.facebook.com/groups/774948999803183/posts/d41d8cd9/1550866485544760/>)

b. ekē    pantiy-ē    iñdan    innə                      yāluwō                      ek-kenek                      innəwā.  
 that class-LOC    from    exist.VA    friend.PL    one-CLF.human.INDF    exist  
 ‘There is one friend who comes from the first class (Grade 1).’ [Mod Use]  
 ([https://maranaya.blogspot.com/2011/01/blog-post\\_28.html/](https://maranaya.blogspot.com/2011/01/blog-post_28.html/))

c. ehi    ek    koṭiy-ek    nowə    koṭi                      dennek                      ma    siṭinu  
 there    one    tiger-INDF    not    tiger.PL    two.CLF.ANMT EM    stay.PST  
 duṭu                      wiṭə. [Mod Use]  
 see.PST.VA    when  
 ‘when I saw not one tiger there, but two’  
 (<https://www.musesbooks.com/Content/UploadFiles/PDF/10ffe188-f145-4577-abcc-7bfaf922a775.pdf/>)

### [1b] Ordinal numerals

Ordinal numerals are nominalized with the classifier *kenā/kenek*. They are used in noun phrases, as shown in (43a, b), but are not used in modification, as shown in (43c).

(43) a. pawul-e                      deweini    kenā                      adə    gedərə    genāwā. [NP use]  
 family-GEN    second    CLF.human    today    home    bring.PST  
 ‘The second member of the family was brought home today.’  
 ([https://www.tiktok.com/@sri.lanka.\\_agro\\_m/video/7454927179214998791/](https://www.tiktok.com/@sri.lanka._agro_m/video/7454927179214998791/))

b. kawədāwat    deweni    ken-ek                      laṅgəṭə    ga-nnə    epā. [NP use]  
 never                      second    CLF.human-INDF    nearby    take-INF    no  
 ‘Never take a second person closer.’  
 (<https://www.facebook.com/reel/1099748888303411/>)

- c. maṁ deweni (\*kenā) ḷamayā. [Mod. use]  
 1.SG second (CLF.human) child  
 ‘I am the second child.’

(<https://www.tiktok.com/@hirutv/video/7226624579345599745/>)

## [2] Demonstratives

The Sinhala classifier *kenā* can take demonstratives, but it can only be used in noun phrase constructions, as shown in (44a). When a demonstrative modifies a noun, *kenā* must be omitted, as shown in (44b).

- (44) a. mē kenā kawudə kiyəla obə-tə kiya-nnə bā! [NP use]  
 this CLF.human who QM 2.SG-DAT tell-INF impossible  
 ‘You can’t tell who this person is.’  
 (<https://gossip.hirufm.lk/69418/2021/12/umariya-sinhawansha.html/>)

- b. mē (\*kenā) pudgəlayā [Mod. use]  
 this (CLF.human) person  
 ‘this person’

## [3] Genitive noun phrases

The classifier *kenā* in genitive noun phrases is acceptable in NP use, as shown in (45a, b), but not in modification use, as shown in (45c, d).

- (45) a. magē kenā yāluw-ek ekkə. [NP use]  
 1.GEN CLF.human friend-INDF together  
 ‘My partner is with a friend.’  
 (<https://www.facebook.com/100957924911718/posts/149542516719925/>)

- b. maṭə obə-gē ken-ek wanə awəśyā-yi. [NP use]  
 1.SG.DAT 2.SG-GEN CLF.human-INDF become want-AM  
 ‘I want to be yours.’  
 (<https://www.facebook.com/groups/349001202263956/permalink/1224008384763229/>)

- c. magē (\*kenā) səmiyā [Mod. use]  
 1.SG.GEN (CLF.human) husband  
 ‘my husband’

- d. obə-gē (\*ken-ek) mituriy-ak [Mod. use]  
 2.SG-GEN (CLF.human-INDF) friend.F-INDF  
 ‘your girlfriend’

#### [4] Adjectives

The classifier *kenā* on adjectives is acceptable in NP use, as shown in (46a, b), but not in modification use, as shown in (46c, d).

- (46) a. *kæməti-mə mahatə kenā kawudə?* [NP use]  
 Prefer-EMP fat CLF.human who  
 ‘Who is your favorite fat person?’  
 (<https://www.facebook.com/gatlcri/posts/686256945387902/>)
- b. *oyā narəkə ken-ek neweyi nēdə.* [NP use]  
 2.SG bad CLF.human-INDF not TAGQ  
 ‘You’re not a bad person, right?’  
 (<https://www.facebook.com/watch/?v=552137072916305/>)
- c. *mahatə (\*kenā) minihā* [Mod. use]  
 fat (CLF.human) man  
 ‘the fat man’
- d. *narəkə (\*ken-ek) kāntāw-ak* [Mod. use]  
*bad* (CLF.human-INDF) woman-INDF  
 ‘a bad woman’

#### [5] Verb-based nominalizations

##### [5a] Relative clauses

The classifier *kenā* on relative clauses is acceptable in NP use, as shown in (47a, b), but not in modification use, as shown in (47c, d).

- (47) a. [*φ maṭə ādərəyə kərənə*] *kenā magē ballā-ṭə ādərəyi* [NP use]  
 1.SG.DAT love.VA CLF.human 1.SG.GEN dog-DAT love  
 ‘The person who loves me loves my dog.’  
 (<https://kingstep.ru/si/cheating-husband/svet-da-odoleet-tmu-na-latyni-latinskie-krylatye-vyrazheniya-istoriya-perevod/> Accessed on January 7, 2025)
- b. [*mamə φ goḍak ādərəyə kərənə*] *ken-ek gænə sihinə*  
 1.SG much love.VA CLF.human-INDF about dream  
*mæwwā.* [NP use]  
 make.PST  
 ‘I had a dream about someone I love a lot.’  
 (<https://si.joellemena.com/සිහින-අර්ථ-නිරූපණය/මම-ආදරය-කරන-කෙනෙකුට-සිහිනයකින්-දැකීම/>)

- c. [φ maṭə ādərəyə kərənə (\*kenā)] miturā [Mod. use]  
 1.SG.DAT love.VA (CLF.human) friend  
 ‘the friend who loves me’
- d. [mamə φ goḍak ādərəyə kərənə (\*ken-ek) ] mituriy-ek [Mod. use]  
 1.SG much love.VA (CLF.human-INDF) friend.F-INDF  
 ‘a girlfriend whom I love very much’

### [5b] Complement clauses

The classifier *kenā* cannot form complement clauses.

### 3.1.2 Classifier *denā/denek*

#### [1] Numerals

The classifier *denā* is used for counting both people and animals, and it is used with numerals greater than one (‘two’ or more). Both NP use and modification use are acceptable, as shown in (48).

- (48) a. ē de-denā mul guruwəru yə. [NP use]  
 that two-CLF.ANMT original teacher.PL AM  
 ‘Those two are the original teachers.’  
 (<https://www.budusarana.lk/budusarana/2024/12/08/tmp.asp?ID=vision02/>)
- b. mē bōṭṭuw-ē denn-ek witərayi ya-nnə ōnə. [NP use]  
 this boat-LOC two.CLF. ANMT-INDF only go-INF must  
 ‘Only two people are allowed on this boat.’  
 (<https://www.gossi plankanews.com/2024/11/negambo-incident.html/>)
- c. monə hitəkin-də mē ali dennā mæruwē. [Mod use]  
 what thought.INDF-Q this elephant.PL two.CLF ANMT kill.PST-FOC  
 ‘What was their intention in killing these two elephants?’  
 ([https://www.lankadeepa.lk/latest\\_news/මොන-හිත-කින්ද-මේ-අලි-දෙන්නා-මැරුවේ/1-31712/](https://www.lankadeepa.lk/latest_news/මොන-හිත-කින්ද-මේ-අලි-දෙන්නා-මැරුවේ/1-31712/))
- d. ūru pəṭaw tun denā [Mod use]  
 pig.PL young three CLF. ANMT  
 ‘the three little pigs’  
 (<https://www.youtube.com/watch?v=GRrQrPZcJnQ/>)

#### [2] Demonstratives

On the internet, there are numerous instances in which the classifier *denā* is employed to nominalize demonstratives in NP use, but not in modification use, as shown in (49a). However, according to my native Sinhala speaker consultants, there is a prevailing view that such usages

are incorrect. (49b) is an ungrammatical example provided by the author, but it would be correct if *denā* were omitted.

- (49) a. ?? *īṭəpasse ē denā kār ekə dihā-ṭə āwā*. [NP use]  
 then that CLF.ANMT car direction-DAT come.PST  
 ‘Then that one came toward the car.’  
 (<https://www.wattpad.com/1159249877-founding-love-knj-completed-/>)

- b. [*ē (\*denā)*] minisun [Mod use]  
 that (CLF.ANMT) person.PL  
 ‘those people’

### [3] Genitive noun phrases

Although infrequent, there are instances on the internet in which the genitive forms of first- and second-person pronouns are nominalized with the classifier *denā*, as shown in (50). However, according to my native Sinhala speaker consultants, these instances are generally regarded as incorrect.

- (50) a. ?? *magē den-ek desə balan obə kīwē*  
 1.SG.GEN CLF.ANMT-INDF direction look.PP 2.SG tell.PST.FOC  
*ek wacəṇəy-ak*  
 one word-INDF  
 ‘(just) one word you said when you looked at my people’  
 (<https://www.tiktok.com/@usershanu698madhu/video/7275002888785628418/>)  
 Accessed on January 7, 2025)

- b. ?? *apē muhudə-ṭə pənəlā apē den-ek mæruwā*  
 1.PL.GEN sea-DAT jump.PP 1.PL.GEN CLF.ANMT-INDF kill.PST  
*indīyā dhīwəṛəyan-ṭə erehiwə yāpənay-ē dhīwəṛə wirōdhātā*  
 India fisherman.PL-DAT against Jaffna-GEN fisheries remonstrance  
 ‘Our sea was invaded, and our people were killed. A protest against Indian fishermen in Jaffna.’  
 (<https://divaina.lk/අපේ-මුහුදට-පැනලා-අපේ-දෙන/>)

- c. ?? *obagē den-ek swāmīn wahansē-gē baləyā ālēpəyā*  
 2.SG.GEN CLF.ANMT-INDF Lord-GEN power anointing  
*læbuwā*.  
 receive.PST  
 ‘Your people have received the anointing of the Lord’s power.’  
 ([https://www.facebook.com/watch/live/?ref=watch\\_permalink&v=354032100777066/](https://www.facebook.com/watch/live/?ref=watch_permalink&v=354032100777066/)) Accessed on January 7, 2025)

#### [4] Adjectives

On the internet, there are examples where adjectives are nominalized by attaching *denā* in NP constructions, as shown in (51a, b). However, according to the consultants' judgment, such examples do not appear to be correct. (51c) is an ungrammatical example provided by the author, but it would be correct if *denā* were omitted.

- (51) a. ?? mokədə poḍi den-ek ekkə kata kərənəwa-tə waḍā  
 what small CLF.ANMT-INDF with talk-DAT than  
 loku kaṭṭiy-ak ekka katā kəra-ddi api-tə enə adəhas  
 big batch-INDF with talk-when 1.PL-DAT come.VA intention  
 pramāṇəyə wəḍi-yi. [NP use]  
 degree more-AM  
 'That's because discussing in a larger group generates more ideas than in a smaller group.'

([https://www.facebook.com/watch/live/?ref=watch\\_permalink&v=475887212190712/](https://www.facebook.com/watch/live/?ref=watch_permalink&v=475887212190712/) Accessed on January 7, 2025)

- b. ?? ē poḍi denā pudumə duk-ak wiṇḍinnē [NP use]  
 that small CLF.ANMT awful sorrow-INDF suffer-FOC  
 'Those little children are suffering a lot.'

(<https://wkwijeratna.blogspot.com/2015/06/blog-post.html/>)

- c. [poḍi (\*denā)] lamayi [Mod use]  
 small (CLF.ANMT) child.PL  
 'the small children'

#### [5] Verb-based nominalizations

Examples can be found on the internet where verbs are nominalized by adding *denā* in NP constructions, as shown in (52a, b). However, based on the consultants' judgment, these examples seem to be incorrect. (52c, d) are ungrammatical examples provided by the author, but they would be correct if *denā/denek* were omitted.

- (52) a. mēwā sələsum kərənə denā ēwā kərannē  
 these plan.VA CLF.ANMT those do.FOC  
 janətāwə-tə set-ak sāda-nnə noweyi  
 people-DAT well-being-INDF make-INF not  
 'Those who plan these things are not doing them to benefit the people.'

([https://www.lankadeepa.lk/latest\\_news/ජලාශ-පිරිසි-කැල-නි-ගත-නම්/1-136342/](https://www.lankadeepa.lk/latest_news/ජලාශ-පිරිසි-කැල-නි-ගත-නම්/1-136342/))

- b. ættatəmā ādārē kārənā denek monā praśne āwā-t  
truly love.VA CLF.ANMT.INDF what problem come.PST-also  
ekəṭṭā hiṭṭiyā.  
together stay  
‘People who truly love each other stay together no matter what the problems are.’  
(<https://www.facebook.com/reel/1192365795589732/>)
- c. [mēwā sālāsūm kārənā (\*denā)] ayā [Mod use]  
these plan.VA (CLF.ANMT) person.SG  
‘the person who plans these’
- d. [ādārē kārənā (\*denek)] ayā [Mod use]  
love.VA (CLF.ANMT.INDF) person.SG  
‘people who did crazy things’

### 3.1.3 Classifier *namə/namak*

*Namə/namak* is utilized not only to denote ‘name’ but also as a numeral classifier for counting monks. A search conducted on the internet revealed instances where the addition of *namə/namak* to numerals, demonstratives, genitive nouns, adjectives, and verbs results in nominalization. The consultants also affirmed that these examples are correct.

#### [1] Numerals

The classifier *namə/namak* can nominalize numerals and can be used both in NP constructions and for modification purposes. Additionally, in the case of the numeral ‘one’, it is acceptable to omit the numeral, and *namak* alone is sufficient.

- (53) a. tun nama-k terun-gē adhikārənā wyapāśəmānay-en  
three CLF.monk-INDF high priest-GEN judicial healing-INST  
no satuṭu wūwāhu, [NP use]  
no satisfy.PST.HON  
‘Three monks were not happy with the Thera’s judicial conduct.’  
(<https://www.thripitakaya.org/sutta/index/24631/>)
- b. poḍi hamuduruwan tun nama-k sasun gatā kirimā [Mod use]  
small monk-PL three CLF.monk-INDF calm passing time  
‘to calm down three little monks’  
(<https://www.saaravita.lk/events/108/540/>)



c. meməbhikṣūn wahansē-lā tun namə śrī laṃkāwē nama-k  
 this monk Sir-PL three CLF.monk Sri Lanka-GEN name-INDF  
 dināgenə siṭinə himiwərun bawədə kiya-ti. [Mod use]  
 win.PP be.VA monk.PL fact say-3.PL  
 ‘It is said that these three monks are well-known monks in Sri Lanka.’  
 ([https://www.lankadeepa.lk/latest\\_news/හම-තන-නමක-ලභ-හර-අල-පට්ට/1-364326/](https://www.lankadeepa.lk/latest_news/හම-තන-නමක-ලභ-හර-අල-පට්ට/1-364326/))

(54) a. swāmīn wahansē-lā ekə nama-k we-nnə puḷuwan [Mod use]  
 Buddhist monk-PL one CLF.monk-INDF be-INF possible  
 ‘It might be just one monk.’  
 (<https://buddhistbrotherhood.net/Dana+paramitha+wadamu/>)

b. ættəṭəmə satuṭuyi technology eka-t ekkə update wenə laṃkā-wē  
 really happy technology-too with update.VA Sri Lanka-GEN  
 hamuduruwan nama-k daki-nnə. [Mod use]  
 monk.PL CLF.monk-INDF see-INF  
 ‘I’m really happy to see a Sri Lankan Buddhist monk who is up to date with  
 technology.’  
 (<https://www.facebook.com/watch/?v=1286119952631292/>)

## [2] Demonstrative

The classifier *namə/namak* can nominalize demonstratives and can be used in NP constructions, but not for modification purposes.

(55) a. budurəjāṇan wahansə, api ē namə bæhəra-tə gene-mu. [NP use]  
 Buddha Sir 3.PL this CLF.monk outside-DAT take-HORT  
 ‘Lord Buddha, let us take that monk with us.’  
 (<https://www.facebook.com/vikasithaapi/>)

## [3] Genitive

The classifier *namə/namak* is capable of nominalizing genitive case nouns and can be used in NP constructions, though it is not used for modification purposes. The classifier *namə/namak* following genitive case nouns can be omitted, as shown in (56b).

(56) a. tawat apē nama-k innəwā. [NP use]  
 more 1.PL-GEN CLF.monk-INDF stay  
 ‘There is another monk of us.’  
 (<https://www.facebook.com/watch/?v=622078266995192/>)

b. dēwəḍattə hāmuduruwō apē da?  
 Dewadatta monk 1.PL-GEN Q  
 ‘Is Venerable Dewadatta ours?’  
 ([https://www.facebook.com/watch/live/?ref=watch\\_permalink&v=](https://www.facebook.com/watch/live/?ref=watch_permalink&v=)

#### [4] Adjectives

The classifier *namə/namak* can nominalize adjectives and is used in NP constructions, although it is not employed for modification purposes.

- (57) a. poḍi namə      duken      pansələ-tə      waḍi. [NP use]  
           little CLF-monk sorrowfully temple-DAT go.HON  
           ‘The little monk goes to the temple in sorrow.’  
           (<https://www.facebook.com/photo.php?fbid=1168100703246805&id=911339815589563&set=a.912574602132751/>)

#### [5] Verb-based nominalizations

The classifier *namə/namak* can be used to nominalize verbs and appears in NP constructions, though it is not utilized for modification purposes.

- (58) a. sanghayā wisin    nam kərənu    labənə      nama-k [NP use]  
           Sangha    by      name.PASS    receive.VA    CLF.monk.INDF  
           ‘a Buddhist monk who is named by the Sangha’  
           (<https://www.ada.lk/woman/religion/ආසනයේ-සුගති/4-287732/>)

### 3.2 Gender markers

The distribution of gender marking in nominalization, both in noun phrases and in modification, has already been demonstrated through several examples in Shibatani (2021). Based on these examples, gender marking in nominalization in Sinhala applies to numerals, demonstratives, genitive noun phrases, adjectives, and verbs. These forms are limited to noun phrases and are not used in modification.

[Numerals]

(59) NP use

- a. ek-ek                      āwā.  
     one-EPI. INDF    came  
     ‘One (M/F) came.’
- b. ek-iyak                  āwā.  
     one-F.INDF        came  
     ‘One.F came.’
- c. ek-ak                     āwā.  
     one-INAN.INDF    came  
     ‘One.INAN        came.’

Mod. use

- a’. ekə/\*ek-ek                  shishyaek  
     one/one-EPI.INDF    male.student  
     ‘one male student’
- b’. ekə/\*ek-iyak              guruwariyak  
     one/one-F.INDF    female.teacher  
     ‘one female teacher’
- c’. ekə/\*ek-ak                potak  
     one/one-INDF    book  
     ‘one book’

(Shibatani 2021: 476)

[Demonstratives]

(60) NP use

- a. mē-kī    magē    duwə  
     this-F    my        daughter  
     ‘This.F is my daughter.’
- b. ē-kā    magē    putā  
     that-M    my        son  
     ‘That.F is my daughter.’

Mod. use

- a’. mē (\*-kī) kellā  
     this (-F)    girl  
     ‘this girl’
- b’. ē (\*-kā) kollē  
     that (-M)    boy  
     ‘that boy’

(Shibatani 2021: 480)

[Genitive noun phrases]

(61) NP use

- a. ē    kellā    magē    ekī  
     that girl    my        one.F  
     ‘That.F girl is mine.F.’
- b. ē    kollā    magē    ekā  
     that boy    my        one.M  
     ‘That boy is mine.’

Mod. use

- b’. magē (\*-kī)    duwə  
     my (-F)        daughter  
     ‘my daughter’
- b’. magē (\*-ekā)    putā  
     my (-one.M)    son  
     ‘my son’

(Shibatani 2021: 482)

[Adjective]

(62) NP use

- [usa]-ekā/-ekī                  āwā.  
     tall-one.EPI/-one.F    came  
     ‘The tall one/one.F came.’

Mod. use

- [usa (\*-ekā/\*-ekī)]              kolla/kella  
     tall (-one.M/-one.F)    boy/girl  
     ‘tall boy/girl’

(Shibatani 2021: 487)

[Verb-based]

(63) NP use

[[oyā gahapu]-ekā/-ekī]                      magē putā/duwa  
you hit.PFT.VA -one.M/-one.F my son/daughter  
'The one whom you hit is my son/daughter.'

Mod. use

[[oyā gahapu] (\*-ekā/\*-ekī)]                      kolla/kella  
you hit.PFT.VA (-one.M/-one.F) boy/girl  
'the boy/girl whom you hit'

(Shibatani 2021: 487)

However, we cannot draw a definitive conclusion about the distribution of gender marking in nominalization based solely on these examples. First, in Sinhala, numerals typically follow the noun when used as modifiers. While the numeral 'one' can appear both before and after the noun, numerals higher than one only appear after the noun. In the examples provided, correctness judgments apply only to cases where the numeral precedes the noun. Additionally, the gender markings attached to genitive nouns, adjectives, and verbs in these examples are specifically associated with the numeral 'one', and thus should be considered morphologically distinct from numerals and demonstratives that are marked purely with gender. For these reasons, this paper will examine whether numerals with gender marking can follow the noun, and investigate whether genitive nouns, adjectives, and verbs that are purely marked for gender can be nominalized.

The results of the investigation confirm that numerals with gender marking indeed follow the noun and modify it, and that first-person genitive pronouns are nominalized exclusively with a gender suffix. Examples of each are provided below.

## [1] Numerals

### [1a] Gender marking of Numeral 'one'

The numeral 'one' with gender marking can be used both as a noun and as a modifier.

[a] ekā/ekek: masculine/epicene

(64) a. ættə    katā kərənə ekā    awasānē narəkə ekā    wenəwā! [NP use]  
truth tell.VA one.EPI finally bad one.EPI become  
'The one who tells the truth ends up being the bad guy!'

([https://www.facebook.com/groups/608244889852652/posts/1566078737402591/?\\_rdr/](https://www.facebook.com/groups/608244889852652/posts/1566078737402591/?_rdr/))

b. metənə eke-k innəwā. [NP use]

here one.M-INDF stay

‘There is one (man or animal) here.’

(<https://www.facebook.com/61554970935659/videos/40-squad-full-rush-pubgm-/1069696598347294/>)

[b] ekī/ekiy-ak: feminine

(65) a. kawudə arə lassənə ekī [NP use]

who that beautiful one.F

‘Who is that beautiful woman?’

(<https://atampahura.blogspot.com/2019/05/>)

b. ape skole-t ekiy-ak innəwā. [NP use]

1.PL.GEN school-too one.F.INDF stay

‘There is a girl at our school.’

(<https://www.tiktok.com/@adaderanasinhala/video/7153811648027856130?lang=ar/>)

There are several instances on the internet where a numeral, accompanied by a gender marker, follows the noun and serves to modify it.

[a] ekā/ekek: masculine/epicene

(66) a. lamayi eke-k æwillā. [Mod. use]

child.PL one.EPI-INDF come.PP

‘One of the children has arrived.’

(<https://www.facebook.com/ROGoNFIRE.GG/videos/hi-boyz/1208951270167375/>)

Accessed on January 7, 2025)

b. lamayi ekā dekə innə, pawul kædiccə, kaṇawændum weccə  
child.PL one.EPI two.EPI exist.VA family broken.VA widowhood become.VA  
gæṇu daruwo hat denek ma maṭə innəwā. [Mod. use]

daguther.PL seven CLF.human EM 1.SG.DAT exist

‘I have seven daughters, all of whom have lost their families, become widows, and have one or two children each.’

(<https://tripitaka.online/sutta/1656/>)

c. ali ekā dennā kæl-en matuwennə gatta. [Mod. use]

Elephant.PL one.EPI two.CLF.ANMT jungle-ABL emerge.PST

‘One or two elephants emerged from the jungle.’

(<https://elakiri.com/threads/වැඩියෙක්ම-බය-වෙව්ව-දවස-mega-thread.1637335/page-5/>)

[b] ekī/ekiy-ak: feminine

- (67) akkəḷā      ekiy-ak      nay-ek      gāṇiy-ek      wenə      amutu  
elder sister.PL one.F-INDF snake-INDF woman-INDF become.VA peculiar  
katāwa-k      bala-nṇə-tə      āsə wī      tibū      atərə, [Mod. use]  
story.INDF see-INF-DAT like PP be.VA while  
'When one older sister wanted to watch a mysterious TV show about a snake turning  
into a woman, ...'

(<https://www.facebook.com/kiriputhagekolama/>)

### [1b] Gender marking of Numeral 'two'

The numeral 'two' with gender marking can also be used both as a noun and as a modifier. However, this numeral is commonly used in conjunction with the numeral 'one'.

- (68) a. eke-k      āwā      deke-k      āwā.      turul kəran      hiṭiyā. [NP use]  
1.EPI-INDF come.PST 2.EPI-INDF come.PST hug.PP stay.PST  
'One came, then two came, and they hugged each other.'

([https://www.youtube.com/watch?v=HHgC6D3i\\_js/](https://www.youtube.com/watch?v=HHgC6D3i_js/))

- b. wasəloyō eke-k      deke-k      nisā      waṭinā-mə      minisun  
outcaste one.EPI-INDF two.EPI-INDF because valuable-most people  
ikmanin      nətiwē.  
quickly lose

'The most valuable people are quickly lost because of one or two infidels.'

([https://www.facebook.com/story.php?story\\_fbid=10236999047544642&id=1495534492&\\_rdr/](https://www.facebook.com/story.php?story_fbid=10236999047544642&id=1495534492&_rdr/) Accessed on January 7, 2025)

- c. gonnu eke-k      deke-k      hari e-yi      bayəwe-nṇə      epā.  
bull.PL one.EPI-INDF two.EPI-INDF just come-AM afraid-INF do not  
'Don't be afraid, one or two bulls will come.'

(කොරෝනා සමග සටන් වදින්නන්ට නෙළුම් කූලුනේ උත්තමාවරය  
අද - Lankadeepa Online)

### [2] Demonstratives

Several examples of demonstratives with gender marking have already been presented in this paper. The following are additional examples of their use in noun phrases.

[a] demonstrative + ā: masculine

- (69) a. ēkā      innəwā      etənə.  
that.M stay there  
'He's there.'

(<https://lowthuruarana.net/දහම්-ලිපි/>)

b. magē ayiyə da malli da mēkā.

1.SG elder brother Q younger brother Q this.M

‘Is this my brother or younger brother?’

([https://www.tiktok.com/@\\_nethu.\\_/video/7465894801150381320/](https://www.tiktok.com/@_nethu._/video/7465894801150381320/))

[b] demonstrative + ī: feminine

(70) a. ēkī innəwā maṭə dænenəwā.

that.F stay 1.SG.DAT feel

‘I feel that she is there.’

(<https://telegra.ph/සහනන-එන-යවන-09-04/>)

b. mēkī maṭə gæhuwā.

this.F 1.SG.DAT hit.PST

‘This one.F hit me.’

(<https://www.youtube.com/watch?v=uw1i6xDIPI0/>)

### [3] Genitive noun phrases

Upon conducting an online search, it was found that genitive nouns with gender marking are only attested with first-person pronouns, and not with second-person or third-person pronouns, nor with common nouns<sup>5</sup>. According to a personal letter from Prof. Dileep Chandralal, *magē-kā*, *apē-kā*, *magē-kī*, and *apē-kī* in examples (71) and (72) are all shortened forms of the genitive + numerals *ekā/ekī*.

[a] gentive+kā: masculine

(71) a. subə-mə subə upandinay-ak wē-wā magē-kā...! [NP use]

good-most good birthday-INDF become-OPT 1.SG.GEN-M

‘Happy birthday, my man!’

(<https://www.facebook.com/watch/?v=1753856355378443/>)

b. monā una-t apē-kā henə hoṇḍə kollā. [NP use]

what become.PST-too 1.PL.GEN-M very good boy

‘No matter what, our guy is a good boy.’

(<https://telegra.ph/වසසනය-පපන-ලල-මල-08-09-3/>)

[b] gentive+kī: feminine

(72) a. āḍəreyi magē-kī. [NP use]

love 1.SG.GEN-F

‘I love you, my angel.’

(<https://www.tiktok.com/@tashniofficial/video/7422687791022165256/>)

<sup>5</sup> When the acceptability of gender marking on personal pronouns was evaluated by two native speakers of Sinhala, one considered it unacceptable for the first, second, and third person pronouns, whereas the other considered it acceptable for all of them.

b. apē-kī koll-ek we-nnə gihillā, [NP use]

1.PL.GEN-F boy-INDF become-INF go.PP

‘Our daughter has become a boy,’

(<https://www.wattpad.com/1428606647>~~පිනිපොදු~~completed-

හතලිස්එක්-වන-පිටින~ /page/3/)

#### [4] Adjectives

Examples of gender marking that independently nominalize adjectives can be found a little bit on the internet. However, my consultants do not accept such examples as correct ones.

(73) ?? laggələ wəli horā. æ bolə apē loku-kā wagē. [NP use]

Laggala sand thief Ah 2.SG 1.PL.GEN big-M like

‘The sand thief in Laggala. Ah! You are like our big son.’

([https://www.facebook.com/permalink.php?story\\_fbid=440431973460908&id=](https://www.facebook.com/permalink.php?story_fbid=440431973460908&id=286946378809469&comment_id=440763150094457/)

286946378809469&comment\_id=440763150094457/ Accessed on January 7, 2025)

Although not common, there is an example of such usage in the naming of a pet cat. *Cūṭikī*<sup>6</sup> is a word formed by adding the feminine suffix *-kī* to the adjective *cūṭi* (meaning small). However, it does not refer to a general small female animal; it is actually the name of a pet cat.

(74) kaṁkoṭā, saha cūṭikī [NP use]

kankota and chutiki

‘Kanokota (name of a cat) and chutiki (name of a cat)

(<https://www.youtube.com/watch?v=M6Tsvzz8Z9c/>)

#### [5] verb-based

Examples of gender marking that independently nominalize verbs seem to be much fewer than those for adjectives on the internet. The following example is the only one I was able to find on the internet, and according to my consultants, such examples are considered incorrect.

[Verb+kā masculine/epicene]

(75) ?? hariyəṭə wiḥilu kəṛənə-kā<sup>7</sup> wagē ṭikak ikman unə banī

exactly joke do.VA-M like a little hurry become.VA Bunny

e tanə-ṭə giyā. [NP use]

that place-DAT go.PST

‘Bunny, who was a little hasty, went there, just like a joker.’

(<https://www.wattpad.com/1373187891-the-bunny-jjk-6/> Accessed on January 7, 2025)

<sup>6</sup> According to a personal letter from Prof. Dileep Chandralal, *cūṭikī* is a shortened form of the adjective *cūṭi* combined with the numeral *ekī*.

<sup>7</sup> According to a personal letter from Prof. Dileep Chandralal, *kəṛənəkā* is a shortened form of the verb *kəṛənə* combined with the numeral *ekā*.



#### 4. ANALYSIS AND DISCUSSION

The Sinhala data collected from the internet this time did not contain any counterexamples to Shibatani's hierarchy constraining and predicting the distribution classifiers and gender markers. The new insights we were able to point out regarding the nominalization hierarchy of Sinhala classifiers and gender markers are summarized in the following Tables 7 and 8.

Table 7. The summary of the nominalization of Sinhala classifiers

| Classifier        | Classifying  | NP use           |     |     |     |         | MOD use  |
|-------------------|--------------|------------------|-----|-----|-----|---------|----------|
|                   |              | NUM              | DEM | GEN | ADJ | V-based |          |
| <i>kenā/kenek</i> | human        | OK (= 1)         | OK  | OPT | OK  | OK      | NUM only |
| <i>denā/denek</i> | human/animal | OK ( $\geq 2$ )  | ??  | ??  | ??  | ??      |          |
| <i>namə/namak</i> | monk         | OK (1: optional) | OK  | OPT | OK  | OK      |          |

Table 8. The summary of the nominalization of Sinhala gender markers

| Gender marker | Classifying    | NP use               |     |                               |     |         | MOD use  |
|---------------|----------------|----------------------|-----|-------------------------------|-----|---------|----------|
|               |                | NUM                  | DEM | GEN                           | ADJ | V-based |          |
| -ā/-ek        | epicene, masc. | OK ( $2 \leq$ ), ??3 | OK  | OPT (=1 <sup>st</sup> person) | PNF | ??      | NUM only |
| -ī/-ak        | fem.           | OK (= 1)             |     |                               |     |         |          |

PNF: Proper noun formation

Common features in the nominalization of classifiers and gender markers in Sinhala include the fact that only numerals extend from noun phrase usage to modifying usage (Mod. Use). This might be related to the fact that numerals in Sinhala, unlike demonstratives, genitive nouns, adjectives, and verbs, appear after the modified noun.

As for the characteristics of Sinhala classifiers, prior research has pointed out the forms *kenā/kenek* attached to the numeral 1 and *denā/denek* attached to numerals greater than 1. However, it was found that some classifiers, such as *namə/namak*, do not have such numeral restrictions. Furthermore, classifiers like *kenā/kenek* and *namə/namak*, which are limited to humans, extend nominalization not only to numerals but also to demonstratives, genitive nouns, adjectives, and verbs. On the other hand, *denā/denek*, used for humans and animals, can only nominalize numerals. It is ironically noteworthy that *namə/namak*, which has not been much discussed in previous research on Sinhala classifiers, has the most extensive distribution and function in nominalization within Sinhala classifiers. The observation of varying degrees of extension within the classifiers of a single language is an important point. However, while not widely used in standard contexts, some internet users have attempted to apply *denā/denek* in nominalization outside of numerals. Conducting large-scale surveys and statistical analysis of such phenomena would be valuable for testing Shibatani's hierarchical theory.

Next, regarding gender marking in Sinhala, previous studies have reported that numerals and demonstratives can take gender markers when nominalized. In this study's internet-based investigation, nominalization was also observed with first-person genitive pronouns. However, interviews with two native Sinhala speakers yielded completely opposing judgments: one stated that gender marking is possible for genitive nouns of all persons (first,

second, and third), while the other claimed it is not possible for any of them. This highlights the risk of drawing incorrect conclusions from internet data alone, due to accidental gaps or nonstandard usage. Accordingly, a more detailed examination of genitive nouns in Shibatani's hierarchical theory is needed, with particular attention to person distinctions. Additionally, while proper noun formation from adjectives was confirmed in the internet data, no such examples were found for verbs. Since Shibatani's theory posits a clear boundary between adjectives and verbs, further investigation into proper noun formation across these word classes is necessary to assess the validity of this boundary.

Furthermore, the data from this study is valuable in clarifying the regional and genealogical characteristics of nominalization involving classifiers and gender markers in numerals. Pardeshi (2024) pointed out that in Indo-Aryan languages spoken in India such as Marathi, Konkani, and standard Hindi, the numerals that allow nominalization with classifiers and gender markers are limited and show different patterns. In this study, it was noted that in Sinhala, there are classifiers *kenā/kenek* and gender markers *-ī/-ak* specific to numeral 1, classifiers *denā/denek* excluding the numeral 'one', classifiers *namā/namak* that attach to all numerals and allow omission of the numeral 'one', and gender markers *-ā/-ek* that apply to numerals 1 and 2. Additionally, although not accepted by consultants in this study, examples were found where gender markers *-ā/-ek* attached to the numeral 'three'. Thus, the numerals attached to classifiers and gender markers are fluid, and there seems to be various individual and regional variations even within languages of the same family.

As for future research, as mentioned above, one focus will be on nominalization with classifiers *denā/denek* for non-numeral items such as demonstratives, genitive nouns, adjectives, and verbs; the relationship between gender markers and the nominalization of genitive nouns and person markers; and proper noun formation with adjectives and verbs under gender marking. Surveys will be conducted to assess the degree of appropriateness for these areas, using methods that involve data collection outside the internet. Additionally, research should be conducted on the nominalization of classifiers and gender markers in Sinhala's contact language, Sri Lankan Tamil, and compared with Sinhala.

## 5. CONCLUSION

The Sinhala language features both classifiers and gender markers, each indicating different situations depending on the extent of nominalization. However, it has been confirmed that nominalization generally follows Shibatani's hierarchical theory. In this study, by incorporating internet data—despite its potential unreliability—I was able to conduct an analysis that anticipates future changes in Sinhala nominalization. Through this analysis, I also uncovered several new perspectives that could help evaluate Shibatani's theory. These include issues related to person in genitive nouns and questions about the boundaries between adjectives and verbs in the nominalization hierarchy. Moving forward, I plan to develop hypotheses based on these perspectives and conduct research to test their validity.

## ABBREVIATIONS

|       |                      |
|-------|----------------------|
| 1     | first person         |
| 2     | second person        |
| 3     | third person         |
| ABL   | ablative             |
| ACC   | accusative           |
| ADJ   | adjective            |
| AM    | assertion marker     |
| ANMT  | animate              |
| CLF   | classifier           |
| COP   | copula               |
| DAT   | dative               |
| DEM   | demonstrative        |
| EM    | emphatic marker      |
| EPI   | epicene              |
| EVD   | evidential clitic    |
| F     | feminine             |
| FOC   | focused form of verb |
| GEN   | genitive             |
| HON   | honorable            |
| HORT  | hortative marker     |
| INAM  | inanimate            |
| INDF  | indefinite           |
| INF   | infinitive           |
| INST  | instrument           |
| LOC   | locative             |
| M     | masculine            |
| NMLZR | nominalizer          |
| NOM   | nominative           |
| NUM   | numeral              |
| OPT   | optative             |
| PASS  | passive              |
| PFT   | perfect              |
| PL    | plural               |
| PP    | perfect participle   |
| PST   | past                 |
| Q     | question             |
| QM    | quotation marker     |
| SG    | singular             |
| TAG   | tag question         |
| TEMP  | temporal marker      |
| TOP   | topic marker         |
| VA    | verbal adjective     |

V-based  
VOC

verb based  
vocative

## ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Professor Emeritus Dileep Chandralal of the University of Okinawa, who read the manuscript, pointed out the issues, and provided valuable comments. I also extend my thanks to Ms. Devika Weerasinghe and Ms. Thisara Ruwanthi, master's course students at the Graduate School of Yasuda Women's University, for their cooperation in evaluating the suitability of example sentences on the internet. Any issues with the content of this paper are the responsibility of the author.

The present study has been supported by grants from the Japan Society for the Promotion of Science (KAKENHI Project IDs: 22H00659 and 23K21931, 'Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers', PI: Kazuhiro Kawachi), (KAKENHI Project ID: 20K13095, 'Development of a Japanese language teaching grammar book for Sinhala-speaking learners aimed at efficient learning and intercultural understanding', PI: Tetsuya Miyagishi), and from the National Institute for Japanese Language and Linguistics ('Empirical study of the typology of nominalization—from theoretical, fieldwork, historical, and dialectal perspectives'), PI: Prashant Pardeshi).

## REFERENCES

- Carter, Charles. 1924. *Carter's Sinhalese English dictionary*. Colombo: The Baptist Missionary Society.
- Chandralal, Dileep. 2010. *Sinhala*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- De Zoysa, R.W.B. 2011. *Godage Sinhalese-English dictionary*. Colombo: S. Godage & Brothers.
- Disanayaka, J.B. 2001. Counting Buddhist monks. *The Sunday Times*, 25th March 2001. <https://www.sundaytimes.lk/010325/plus3.html#3LABEL2>
- Gunasekara, Abraham Mendis. 1891. *Comprehensive grammar of the Sinhalese language*. Colombo: G. J. A. Skeen, Government Printer.
- Karunatilake, W.S. 1998. *An introduction to spoken Sinhala, 2nd edition*, Colombo: Gunasena.
- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Pardeshi, Prashant 2024. On the distribution and function of classifier and gender marking in Marathi with side glances at a few Indo-Aryan languages: A nominalization perspective. Joint Meeting for the "Empirical Study of the Typology of Nominalization - From Theoretical, Fieldwork, Historical and Dialectal Perspective" Project and the "Cross-Linguistic Studies on Grammatical Nominalizations: With a Focus on Classifiers and Gender Markers" Project. National Institute for Japanese Language and Linguistics (NINJAL), (14 Sept 2024) at the University of Tokyo (Hongo Campus), Tokyo, Japan
- Shibatani, Masayoshi. 2021. Rentai shuushoku no bunpoo: Ruibetsushi to bunpoosei o chuushin ni (Grammar of adnominal modification: Centering on classifiers and gender). In Sung-Yeo Chung and Masayoshi Shibatani. (eds.) *Taigenka riron to gengo bunseki (Nominalization theory and linguistic analysis)*, 459–555. Osaka: Osaka University Press.

- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)
- Wijayatunga, H. 2012. *Gunasena great Sinhala dictionary*. Colombo: Gunasena.

# Distribution of classifier and gender marking in Konkani

Prashant Pardeshi<sup>1</sup> and Govind Mopkar<sup>2</sup>

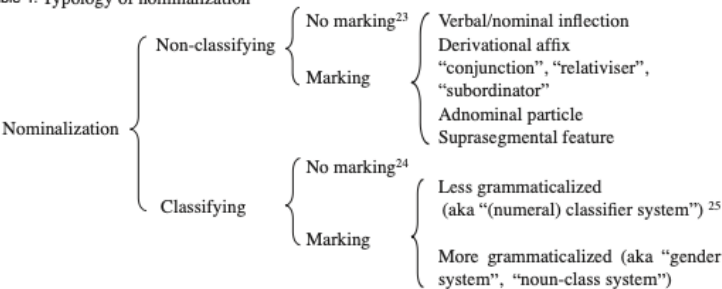
## ABSTRACT

The present study provides an initial account of the distribution of classifiers and gender markers in Konkani (an Indo-Aryan language spoken along the west coast of India) based on the fieldwork conducted in February and March 2024 in four locations: (i) the northern tip (Pernem taluka) of the state of Goa, (ii) southern tip (Canacona taluka) of the state of Goa, (iii) Mangalore, Karnataka state, and (iv) Kozhikode, Kerala state. Adopting the framework called “dynamic functional typology” proposed in Shibatani (2023) and following the hierarchy (NUM > DEM > GEN > ADJ > V-based) proposed therein (p.35), we describe the distribution of classifiers and gender markers in Konkani spoken in Goa (referred to as GK) and compare it with its counterpart spoken in Kozhikode, Kerala state, India (referred to as KK). We claim that (numeral) classifiers and grammatical genders in Konkani nominalize the structures to which they are attached and at the same time classify what is being denoted by those structures. Further, we demonstrate that the distribution of classifier and gender marking in Konkani lends support to Shibatani’s (2023:35) hierarchy constraining classifier/gender marking.

## 1. INTRODUCTION

Shibatani (2023: 21) proposes a typology of nominalization, as shown in the figure below, and argues that classifier/gender marking is a “classifying” type of nominalization that not only nominalizes but also classifies its denotation.

Table 1: Typology of nominalization



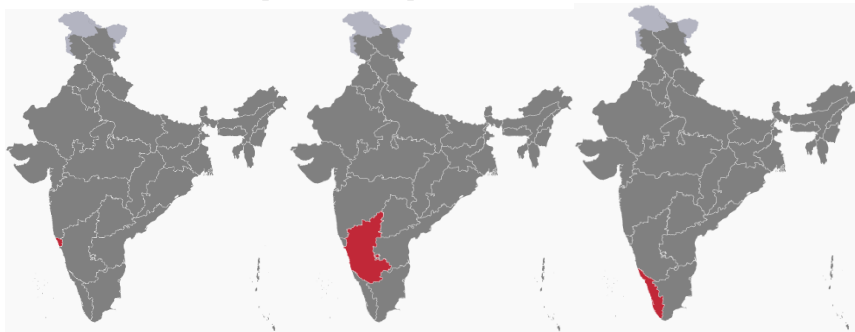
In this paper we will focus only on the “classifying” type of nominalizations, which are overtly marked by a (numeral) classifier or by grammatical gender. Specifically, we will offer a preliminary report on the distribution of classifiers and gender markers in Konkani (Indo-

<sup>1</sup> National Institute for Japanese Language and Linguistics, Tokyo, Japan. Email: prashant@ninjal.ac.jp

<sup>2</sup> Deccan College Post-Graduate and Research Institute, Pune, India.

Aryan language) spoken along the west coast of India within the typological framework called “dynamic functional typology” proposed in Shibatani (2023). We claim that (numeral) classifiers and grammatical genders in Konkani nominalize the structures to which they are attached and at the same time classify what is being denoted by those structures. Further, we will demonstrate that the distribution of classifier and gender marking in Konkani lends support to Shibatani’s (2023:35) hierarchy constraining classifier/gender marking.

The data presented in the paper were collected during our fieldwork on Konkani spoken conducted from 25 February to 13 March 2024 in four locations along the west coast of India: (i) the northernmost tip of the state of Goa state (Pernem taluka), (ii) southernmost tip of the state of Goa state (Canacona taluka), (iii) Mangalore, Karnataka state and (iv) Kozhikode, Kerala state (see the maps and description below for details).



Map 1. Goa (left), Karnataka (middle) and Kerala (right) states in India.

Sources:

Goa: <https://en.wikipedia.org/wiki/Goa>

Karnataka: <https://en.wikipedia.org/wiki/Karnataka>

Kerala: <https://en.wikipedia.org/wiki/Kerala>

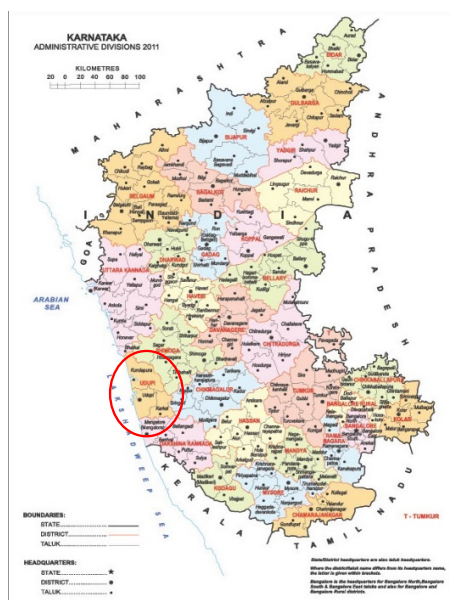
The field sites within each state are shown on the maps below.



Map 2. Field sites in Goa State

Source: <https://en.m.wikipedia.org/wiki/File:Goamap.png>

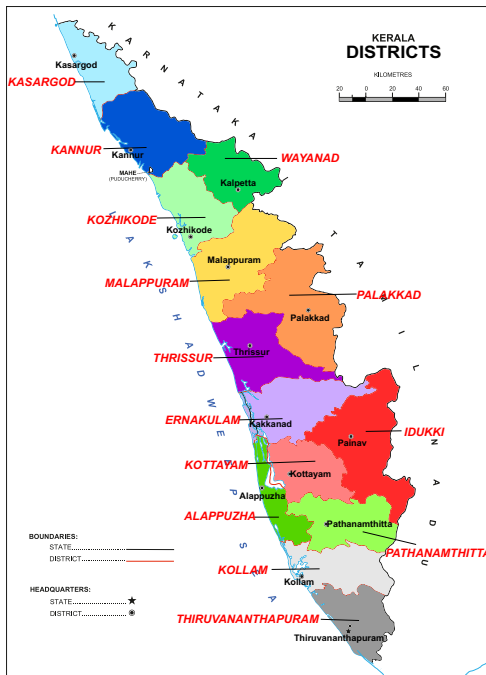
Goa: Fieldwork was conducted at the northernmost tip (Harmal, Pernem taluka) and the southernmost tip (Chaudi, Canacona taluka) of Goa from 25 February to 3 March 2024 with Govind Mopkar, partly joined by Ruta Paradkar in Pernem.



Map 3. Field sites in Karnataka State (Udupi, Mangaluru)

Karnataka state, India: Fieldwork conducted in Mangalore (Southern) [Fieldwork on the Kudubi variety of Konkani in Mangalore (southern Karnataka) was conducted from 5 to 9 March 2024 with Niranjan Upoor.





Map 4. Field sites in Kerala State

Source: [https://upload.wikimedia.org/wikipedia/commons/5/5b/Political\\_map\\_of\\_Kerala.svg](https://upload.wikimedia.org/wikipedia/commons/5/5b/Political_map_of_Kerala.svg)  
 Kerala state, India: Fieldwork was conducted in Kozhikode (Calicut, northern Kerala) from 11 to 13 March 2024 with Dr. Niranjan Uppoor.

Before moving on to the classifier and gender marking, we will offer a brief description of the characteristic features of nouns in Konkani. Nouns in Konkani have three characteristic features, namely (i) gender, (ii) number, and (iii) case (see also Dhongde (2022: 47-58). A noun in Konkani is divided into three genders, namely masculine, feminine, and neuter, irrespective of their animacy (e.g., an inanimate noun such as *kap* ‘cup’ is neuter, while *pelɔ* ‘steel glass’ is masculine) or natural gender (e.g., *pāl* ‘house lizard’ is feminine, while *čhital* ‘antelope’ is neuter). Nouns in Konkani also inflect for number: singular and plural. Further, nominals in Konkani take case markers. They have two case forms: direct case (morphologically unmarked or zero marked form) and oblique case (a derived form to which overt case markers are added). Furthermore, it should be noted that declining adjectives also undergo change in their form according to the gender and number of their denotation and take case markers like nouns. Definiteness/specificity can be overtly marked with demonstrative pronouns.

Konkani also has non-classifying nominalizers (non-classifying nominalization markers), which will not be discussed in this paper. For non-classifying nominalizers, see Dhongde (2022: Chapter 7, Word formation).

## 2. OVERVIEW OF CLASSIFIER AND GENDER MARKING AND GRAMMATICAL NOMINALIZATIONS IN KONKANI

### 2.1 Classifier and gender marking in Konkani

Dixon (1982) suggests that the two noun classification systems, namely classifiers and gender, as separate categories that do not coexist in the same language. However, Corbett (2013) has shown that there are a few languages where both co-exist. Konkani is one such language. As mentioned earlier, nouns in Konkani are divided into three grammatical genders: masculine (M), feminine (F), and neuter (N). Further, Konkani has only one “sortal” classifier, which is used for classifying humans, such as its northern neighbour, Marathi. There are also a few mensural classifiers (or measure words) in Konkani, such as *bhājyeči tsuḍi* (a bunch of vegetable), *čaplānči dzoḍi* (a pair of shoes), and *nustyātsō vāṭo* (a portion of fish), among others, which will not be discussed in this paper.

### 2.2 Grammatical nominalizations in Konkani

Below, we provide forms and/or examples of each type of grammatical nominalization from the Pernem variety of Konkani.

#### [1] Numerals

##### [1a] Cardinal numerals

Table 1. Cardinal Numbers in Konkani (the Pernem variety)

| Cardinal numerals | Form  | Cardinal numerals | Form |
|-------------------|-------|-------------------|------|
| 1                 | ek    | 6                 | sa   |
| 2                 | don   | 7                 | sāt  |
| 3                 | tin   | 8                 | āṭh  |
| 4                 | čār   | 9                 | ṇau  |
| 5                 | pānts | 10                | dha  |

##### [1b] Ordinal numerals

Ordinal numerals have a three-way gender distinction: M/F/N, as illustrated in the following table.

Table 2. Ordinal numerals in Konkani (the Pernem variety)

| Ordinal number | Form (M/F/N)    | Ordinal number | Form (M/F/N)   |
|----------------|-----------------|----------------|----------------|
| 1st            | pail- {o/i/ā}   | 6th            | savv- {o/i/yā} |
| 2nd            | dusr- {o/i/ā}   | 7th            | sātv- {o/i/yā} |
| 3rd            | tisr- {o/i/ā}   | 8th            | āṭhv- {o/i/yā} |
| 4th            | čaut- {o/i/ā}   | 9th            | ṇauv- {o/i/yā} |
| 5th            | pātsv- {o/i/yā} | 10th           | dhā- {o/i/yā}  |

Note: The neuter gender marker for the Canacona variety is /e/ and for other varieties it is /e/.

## [2] Demonstratives

In all varieties of Konkani we worked on, demonstratives are marked for three genders: masculine, feminine, and neuter.

Table 3. Demonstratives in Konkani (the Pernem variety)

|           | Proximal |        | Distal   |        |
|-----------|----------|--------|----------|--------|
|           | Singular | Plural | Singular | Plural |
| Masculine | ho       | he     | to       | te     |
| Feminine  | hi       | hyo    | ti       | tyo    |
| Neuter    | hyā      | hī     | tā       | tī     |

## [3] Genitive noun phrases

Table 4. Genitive noun phrases in Konkani (the Pernem variety)

| Personal pronoun | M                  |                  | F                |                   | N         |                  |
|------------------|--------------------|------------------|------------------|-------------------|-----------|------------------|
|                  | Sg                 | Pl               | Sg               | Pl                | Sg        | Pl               |
| 1                | mhādzo             | mhāje            | mhāji            | mhājyo            | mhādza    | mhāji            |
| 2                | tudzo              | tuje             | tuji             | tujyo             | tudza     | tuji             |
| 3                | te{ts/dz}o         | te{č/j}e         | te{č/j}i         | te{č/j}yo         | tetsa     | te{č/j}i         |
|                  | ti{ts/dz}o         | ti{č/j}e         | ti{č/j}i         | ti{č/j}yo         | titsa     | ti{č/j}yo        |
|                  | te{ts/dz}o         | te{č/j}e         | te{č/j}i         | te{č/j}yo         | tetsa     | te{č/j}i         |
| Noun             | N-Obl-<br>{ts/dz}o | N-Obl-<br>{č/j}e | N-Obl-<br>{č/j}i | N-Obl-<br>{č/j}yo | N-Obl-tsa | N-Obl-<br>{č/j}i |

Note: Genitive markers across varieties tend to vary with /č/, /j/ and /gel/ forms.

## [4] Adjectives

In Konkani, adjectives can be divided into two groups: (i) those that decline according to the number, gender, and case of the noun they modify and (ii) those that do not decline. Example of these two types are given below.

Declinable type

### NP use

- (1) mhākā                      dhav-{o/i/(y)ā}                      āvāḍṭā  
 1SG.M.DAT                  white-{M/F/N}                      like.IFV  
 ‘I like white one (M/F/N).’

### Mod use

- (2) mhākā                      dhav-{o/i/(y)ā}                      {mobāil/sāḍi/ghaḍyāl}                      āvāḍṭā  
 1SG.M.DAT                  white-{M/F/N}                      mobile/saree/watch                      like.IFV  
 ‘I like white {mobile/saree/watch}.’

Non-declinable type

**NP use**

- (3) mhākā                      goḍ                      āvāḍṭā  
1SG.M.DAT                  sweet(M/F/N)      like.IFV  
'I like sweet one (M/F/N).'

**Mod use**

- (4) mhākā                  goḍ                  āmo/barfi/udik                  āvāḍṭā  
1SG.M.DAT      sweet                  mango/burfi/water      like.IFV  
'I like sweet {mango/burfi/water}.'

**[5] Verb-based nominalizations**

**[5a] Relative clauses**

In the Pernem variety, the relative-correlative strategy is predominantly used to form so-called relative clauses or what we refer to as V-based nominalizations.

**NP use** [subject of intransitive verb, feminine referent]

- (5) (j-i)                  soḥā-cher                  bas-l-yā                  t-i  
REL-FSG      sofa-LOC                  sit-PST-FSG      COREL-FSG  
mhādz-i                  maitriṇ  
1SG.GEN-FSG      friend.FSG  
'[The one (FSG) sitting on the sofa] is my friend.'

**Mod use** [subject of intransitive verb, feminine referent]

- (6) (j-i                  bāil)                  soḥā-cher                  bas-l-yā                  t-i  
REL-FSG                  woman.FSG      sofa-LOC                  sit-PST-FSG      COREL-FSG  
bāil                  mhādz-i                  maitriṇ  
woman-FSG                  1SG.GEN-FSG      friend.FSG  
'[The woman sitting on the sofa] is my friend.'

**[5b] Complement clauses**

We have not collected data on noun and verb complement clauses yet.

**3. DATA**

**3.1 Classifiers and gender marking**

**[1] Numerals**

The numeral is the only category that can potentially host both classifier and gender marking, and their distribution is as follows. The numeral 'one' is unique in that it hosts only gender marking (*ek.l-}{o/i/ā}*: one.L-}{MSG/FSG/NSG}). All other numerals, except 'one', host a

combination of a classifier plus gender marking: NUM (other than ‘one’) + *dz{ā/a}n* (in the case of male referents) or NUM (other than ‘one’) + *dz{ā/a}nā* (in the case of female referents or epicene [male + female] referents).

### [1a] Cardinal numerals

- (7) *kāl*                      *kitki-dz{ā/a}nā*                      *ill-t̃?*  
yesterday      how.many-CLF.M/F      come.PF-M/F  
‘How many people (male + female, or only females) came yesterday?’

#### NP use

- (8) *ek.l-{o/i/ā}-(ts)*                      *ill-{o/i/ā}*  
one.L-{MSG/FSG/NSG}-EMPH      come.PF-{MSG/FSG/NSG}  
‘(Only) one (MSG/FSG/NSG) came.’

#### Mod use

- (9) *ek.l-{o/i/ā}-(ts)*                      *dāldo/bāil/cheḍu*  
one.L-{MSG/FSG/NSG}-EMPH      man.MSG/woman.FSG/girl.NSG  
*ill-{o/i/ā}*  
come.PF-{MSG/FSG/NSG}  
‘(Only) one (MSG/FSG/NSG) {man/woman/girl} came.’

We looked at four varieties of Konkani namely the Pernem variety, the Canacona variety, the Kudubi variety, and the Kozhikode variety. The gender marked L-form of the numeral ‘one’ is attested in all of them. However, there is an interesting difference. While the Pernem, Canacona and Mangalore (Kudubi) varieties have a three-way gender distinction with the gendered or L-form (M/F/N) of the numeral ‘one’, the Kozhikode variety has only a two-way gender distinction (M/F). In the Pernem, Canacona and Mangalore (Kudubi) varieties of Konkani, an unmarried girl is treated as neuter (grammatical gender), while the same in the Kozhikode variety is treated as feminine (semantic gender). The loss of the neuter L-form of the numeral ‘one’ in the Kozhikode variety may be due to an on-going shift from a grammatical gender system to a semantic/biological gender system. One can speculate that this on-going shift is a contact-induced change, under the influence of Malayalam (Dravidian), but the same does not seem to have happened yet in the Kudubi variety, which is in also close contact with Kannada (Dravidian). A finer-grained work on this is needed.

In Sanskrit (Indo-Aryan), the mother language of Konkani, numerals such as one, three and four have three-way gender distinction (M/F/N), while numeral two has a two-way gender distinction (M vs. F/N) in their nominative form. Modern Indo-Aryan languages such as Hindi and Marathi have lost overt morphological gender marking on numeral. From a historical/diachronic perspective, the L-form of Konkani numeral one can be considered as the lone survivor of the old morphologically gender marked form of numeral one (Cf. In Spanish also, only the numeral one has retained the gender distinction (*uno* vs *una*)). In the Pernem variety, we found another for namely T-form for the numeral one which shows three-way

gender distinction: (*ek.t*-{*o/i/ā*}: one.T-{MSG/FSG/NSG}), which in addition to the quantity unit meaning also means ‘alone’. Whether this T-form is the same as the T-form in Bangla and Assamese needs further investigation.

Moving on to other numerals, numerals ‘two’, ‘three’, and ‘four’ in Konkani (such as Marathi) have a special form, which we call as G-form: *dog* (2.G), *t{i/e}g* (3.G), *ch{o/u}g* (4.G). G-forms of two/three/four are like L-form of ‘one’ in that they both are **gendered**. They differ, however, in terms of concatenability with CLF: L-forms do not concatenate with CLF, while g-forms do. Examples of NP use and Mod use of the G-forms of numeral ‘two’, ‘three’, and ‘four’, without and with CLF, from the Kudubi variety are given below.

### Male referent

- (10) *kāl*            *kitle*            *dārle*            *ill-e?*  
 yesterday   how.many   man.PL   come.PF-MPL  
 ‘How many men came yesterday?’

### G-form : *dog/tig/chug* (Masculine referent)

#### NP use

- (11) *dog/tig/chug*                    *il-e/ill-e*  
 two(g)/three(g)/four(g) come.PST-MPL/come-PSTPF-MPL  
 ‘Two/three/four (MPL) came/had come.’

#### Mod use

- (12) *dog/tig/chug*                    *dārl-e*                    *il-e/ill-e*  
 two(g)/three(g)/chug(g)        man.MPL                come.PST-MPL/come-PSTPF-MPL  
 ‘Two/three/four men (MPL) came/had come.’

### G-form + CLF ({*dog/tig/chug*}-dan): (Masculine referent)

#### NP use

- (13) *dog/tig/chug-dan*                    *il-e/ill-e*  
 two(g)/three(g)/four(g)-CLF come.PST-MPL/come-PSTPF-MPL  
 ‘Two/three/four (MPL) came/had come.’

#### Mod use

- (14) *dog/tig/chug-dan*            *dārl-e*            *il-e/ill-e*  
 two(g)/three(g)/four(g) man.MPL   come.PST-MPL/come-PSTPF-MPL  
 ‘Two/three men (MPL) came/had come.’

**Female [women]/Neuter [girls] referent**

- (15) kāl kitle bāil.manushā/cheḍ-vā ill-ĩ?  
 yesterday how.many woman.PL/girl-PL come.PF-FPL  
 ‘How many girls/women came yesterday?’  
 (cheḍu (=girl) is neuter in singular number while it is feminine in the plural number  
 (cheḍvā (= girls) is feminine hence the question is same for FPL as well as NPL)

**G-form : dog/tig/chug (Feminine and neuter)****NP use**

- (16) dog/tig/chug il-ĩ/ill-ĩ (il-i/ill-i)  
 two(g)/three(g)/four(g) come.PST-F&NPL/come-PSTPF-F&PL  
 ‘Two/three/four (FPL/NPL) came/had come.’

**Mod use**

- (17) dog/tig/chug bāil.manushā/cheḍ-vā il-ĩ/ill-ĩ (il-i/ill-i)  
 two(g)/three(g)/four(g) woman.PL/girl-PL come.PST-F&NPL/come-PSTPF-F&PL  
 ‘Two/three/four men (MPL) came/had come.’

**G-form+ CLF: dog/tig/chug+dan (Feminine and neuter)****NP use**

- (18) dog/tig/chug-dan il-ĩ/ill-ĩ (il-i/ill-i)  
 two(g)/three(g)-four(g)-CLF come.PST-F&NPL/come-PSTPF-F&PL  
 ‘Two/three/four (FPL/NPL) came/had come.’

**Mod use**

- (19) dog/tig/chug-dan bāil.manushā/cheḍ-vā  
 two(g)/three(g)/four(g)-CLF woman.PL/girl-PL  
 il-ĩ/ill-ĩ (il-i/ill-i)  
 come.PST-F&NPL/come-PSTPF-F&PL  
 ‘Two/three/four women (FPL)/girls (NPL) came/had come.’

Finally, cardinal numerals in combination with classifier, also can be used for referring to human referents, as shown below, which are answers to the question in (10).

**Numeral + CLF: {don/tin/chār} + dan****NP use**

- (20) don/tin/chār-dan il-ĩ/ill-ĩ (il-i/ill-i)  
 two/three/four-CLF come.PST-F&NPL/come-PSTPF-F&PL  
 ‘Two/three/four (FPL/NPL) came/had come.’

### Mod use

- (21) don/tin/chār-dan      bāil.manushā/cheḍ-vā    il-ĩ/ill-ĩ (il-i/ill-i)  
two/three/four-CLF    woman.PL/girl-PL      come.PST-F&NPL/come-PSTPF-F&PL  
'Two/three women (FPL)/girls (NPL) came/had come.'

### Numeral 'five' and above

For the numeral 'five', there are no special forms such as the L/T-form (which is available only for numeral one) and G-forms (which are available for two, three and four). To denote human beings, the only possible form is: {cardinal numeral (NUM) + CLF}. Numerals in and of themselves cannot be used for denoting human beings, except in limited contexts such as newspaper headlines. Inanimate referents, on the other hand, can be denoted by cardinal numerals alone.

### Numeral + CLF: pāts ('five') + dan: Masculine referent

#### NP use

- (22) **pāts-dan**    il-e/ill-e  
five-CLF    come.PST-MPL/come-PSTPF-MPL  
'Five (MPL) came/had come.'

#### Mod use

- (23) **pāts-dan**    **dārl-e**                      il-e/ill-e  
five-CLF    man.MPL                      come.PST-MPL/come-PSTPF-MPL  
'Five men (MPL) came/had come.'

### Feminine and neuter referent

#### NP use

- (24) **pāts-dan**    il-ĩ/ill-ĩ (il-i/ill-i?)  
five-CLF    come.PST-F&NPL/come-PSTPF-F&PL  
'Five (FPL/NPL) came/had come.'

#### Mod use

- (25) **pāts-dan**    **bāil.manushā/cheḍ-vā**      il-ĩ/ill-ĩ (il-i/ill-i?)  
five-CLF    woman.PL/girl-PL                      come.PST-F&NPL/come-PSTPF-F&PL  
'Five women (FPL)/girls (NPL) came/had come.'

**Summary of numerals (NUM):** For denoting human beings, cardinal numerals in and of themselves cannot be used. They should be either marked with gender (L/T-form of numeral one, G-form of numeral two, three and four) or used in combination with the classifier. Numeral one is unique in having a gendered form (L-form) which specifies the gender of the denotation in the Konkani spoken in the coastal region starting from northern tip of Goa (Pernem) to northern Kerala (Kozhikode). Moving southward from northern Goa up to Karnataka



(Mangalore), three-way gender distinction (M/F/N) is maintained, while further moving south and entering in Kerala (Kozhikode) the neuter form seems to have been lost. This change might reflect an on-going shift from grammatical gender to semantic gender, arguably under the influence of Malayalam (Note that Dravidian languages have semantic gender and not grammatical gender). A similar shift is also seen on the loss of gender marking on classifier in Kozhikode Konkani (The loss of gender on the CLF is also observed in Agonda village in Canacona). Further, classifier cannot be attached to the gendered L-form of one.

Numerals two, three, and four have G-forms that denote human beings. To these, a human classifier can be optionally attached. The G-forms (with and without CLF and in both usages) are unequivocally accepted.

As for numeral five and above, the [NUM + CLF] form is unequivocally acceptable. On the acceptability of bare numerals (NUM alone) to denote human beings, judgements of speakers are divided. These observations summarized above are based on data collected from a few consultants. A more detailed study is needed to refine the distribution of classifier and gender marking on numerals.

### [1b] Ordinal numerals

In Konkani, ordinal numerals are necessarily gender marked, as shown in Table 2 (Section 2.2). Consider the following examples.

#### NP use

- (26) pail-yā      bāk-ā-r                      bas-lel-i              dusr-i  
       first-OBL   bench-OBL-LOC           sit-PST.PF-F.SG   second-F.SG  
       mhā-j-i              āvay  
       1.GEN.F.SG        mother  
       ‘The second one sitting on the first bench is my mother.’

#### Mod use

- (27) pail-yā      bāk-ā-r                      bas-lel-i              dusr-i              bāil  
       first-OBL   bench-OBL-LOC           sit-PST.PF-F.SG   second-F.SG        woman  
       mhā-j-i              āvay  
       1.GEN.F.SG        mother  
       ‘The second woman sitting on the first bench is my mother.’

### [2] Demonstratives

In all varieties of Konkani we worked on, demonstratives are 3-way gender marked: masculine (M), feminine (F), and neuter (N). They are unequivocally acceptable in NP use and Mod use. Classifier cannot be attached to demonstratives.

In the case of animate referents, gender distinction is maintained in both usages. Examples from the Kozhikode variety are given below.

## Animate

### Masculine referent

#### NP use

- (28) **y- $\{o/*e\}$**                       t.a-gel-yā-pashi                      thoru                      āsa  
this- $\{MSG/NSG\}$                       that.MSG-GEN-OBL-COMP                      fat.MSG                      be.PRS  
'This one (MSG) is fatter than that one (MSG).'

#### Mod use

- (29) **y- $\{o/*e\}$**                       **dādlo**                      t- $\{o/e/yā\}$                       dādł-yā-pashi  
this- $\{MSG/NSG\}$                       man.MSG                      that- $\{MSG/NSG/OBL\}$                       man-OBL-COMP  
thoru                      āsa  
fat.MSG                      be.PRS  
'This man is fatter than that man.'

### Feminine referent

#### NP use

- (30) **y- $\{i/*e\}$**                       t-i-gel-yā-pashi i-pashi                      thoru                      āsa  
this- $\{FSG/NSG\}$                       that-FSG-GEN-OBL-COMP                      fat.FSG                      be.PRS  
'This one (FSG) is fatter than that one (FSG).'

#### Mod use

- (31) **y- $\{i/*e\}$**                       **bāil.manshi**                      t- $\{i/e/yā\}$                       bāil.manshi-pashi  
this- $\{FSG/NSG\}$                       woman.FSG                      that- $\{FSG/NSG/OBL\}$                       woman.OBL-COMP  
thoru                      āsa  
fat.FSG                      be.PRS  
'This woman is fatter than that woman.'

### Neuter referent

#### NP use

- (32) **y-e**                      t.a-gel-yā/t-yā-ch-yā-pashi                      thoru                      āsa  
this-NSG                      that.NSG-GEN-OBL-COMP                      fat.FSG                      be.PRS  
'This one (NSG) is fatter than that one (NSG).'

#### Mod use

- (33) **y-e**                      **supe**                      t- $\{e/yā\}$                       suṇ-yā-pashi                      thoru                      āsa  
this-NSG                      dog.NSG                      that- $\{NSG/OBL\}$                       dog-OBL-COMP                      fat.FSG                      be.PRS  
'This dog is fatter than that dog.'

In the case of inanimate referents, however, there is a tendency to neutralize the gender in the Mod use. In the NP use, gender neutralization is possible but retaining gender is preferred as shown in the following examples.

## Inanimate

### Masculine referent

#### NP use

- (34) **y- $\{o/e\}$**                       t-a-j-yā-pashi                      porn-o  
this- $\{MSG/NSG\}$               that-MSG-GEN-OBL-COMP              old-MSG  
'This one (MSG) is older than that one (MSG).'

#### Mod use

- (35) **y- $\{o/e\}$**                       **mancho**                      t-e                      manch.ā-pashi              porn-o  
this- $\{MSG/NSG\}$               cot.MSG                      that-NSG              cot.OBL-COMP              old-MSG  
'This cot (MSG) is older than that cot (MSG).'

### Feminine referent

#### NP use

- (36) **y- $\{i/e\}$**                       t-e-j-yā-pashi                      porn-i  
this- $\{FSG/NSG\}$               that-NSG-GEN-OBL-COMP              old-FSG  
'This one (FSG) is older than that one (FSG).'

#### Mod use

- (37) **y- $\{i/e\}$**                       **gāḍi**                      t-e                      gāḍi-yā-pashi                      porn-i  
this- $\{FSG/NSG\}$  vehicle.FSG              that-NSG              vehicle-OBL-COMP              old-FSG  
'This vehicle (FSG) is older than that vehicle (FSG).'

### Neuter referent

#### NP use

- (38) **y-e**                      t-e-j-yā-pashi                      porn-e  
this-NSG              that-NSG-GEN-OBL-COMP              old-NSG  
'This one (NSG) is older than that one (NSG).'

#### Mod use

- (39) **y-e**                      **ghaḍiyār**                      t-e                      ghaḍiyār-ā-pashi                      porn-e  
this-NSG              watch/clock.NSG              that-NSG              watch/clock-OBL-COMP              old-NSG  
'This watch/clock (NSG) is older than that watch/clock (NSG).'

We did not notice gender neutralization in the Pernem variety, as shown in the examples below. However, further work is needed to see whether it happens in other varieties as well.

### Masculine referent

#### NP use

- (40) **h-o**            mhārak/dhav-o            (āsā)            āṇi  
this-MSG   expensive/white-MSG   be.PRS   and  
**t-o**            savāy/ kāl-o            (āsā)  
that-MSG   cheap/black-MSG   be.PRS  
‘This one (MSG) is expensive/white and that one (MSG) is cheaper/black.’

#### Mod use

- (41) **h-o**            **mobāil**            mhārak/dhav-o            (āsā)            āṇi  
this-MSG   mobile.MSG   expensive/white-MSG   be.PRS   and  
**t-o**            **mobāil**            savāy/kāl-o            (āsā)  
that-MSG   mobile.MSG   cheap/black-MSG   be.PRS  
‘This mobile is expensive/white and that mobile is cheaper/black.’

### Feminine referent

#### NP use

- (42) **h-i**            mhārak/dhav-i            (āsā)            āṇi  
this-FSG   expensive/white-FSG   be.PRS   and  
**t-i**            savāy/kāl-i            (āsā)  
that-FSG   cheap/black-FSG   be.PRS  
‘This one (FSG) is expensive/white and that one (FSG) is cheaper/black.’

#### Mod use

- (43) **h-i**            **sāḍi**            mhārak/dhav-i            (āsā)            āṇi  
this-FSG   saree.FSG   expensive/white-FSG   be.PRS   and  
**t-i**            **sāḍi**            savāy/kāl-i            (āsā)  
that-FSG   saree.FSG   cheap/black-FSG   be.PRS  
‘This saree is expensive/white and that saree is cheaper/black.’

### Neuter referent

#### NP use

- (44) **h-yā**            khup vhoḍl-ā/dzun-ā            (āsā)            āṇi  
this-NSG   very big-NSG/old-NSG   be.PRS   and  
**t-ā**            dhākt-ā/navy-ā            (āsā)  
that-NSG   small-NSG/new-NSG   be.PRS  
‘This one (NSG) is very big/old, and that one (NSG) is small/new.’

### Mod use

- (45) **h-yā**      **medz**      khup vhoḍl-ā/dzun-ā      (āsā)      āṇi  
this-NSG    table.NSG    very big-NSG/old-NSG    be.PRS    and  
**t-ā**      **medz**      dhākt-ā/navy-ā      (āsā)  
that-NSG    table.NSG    small-NSG/new-NSG      be.PRS  
‘This table is very big/old, and that table is small/new.’

### [3] Genitive noun phrases

In all varieties of Konkani we worked on, genitives are 3-way gender marked: masculine (M), feminine (F), and neuter (N). They are unequivocally acceptable in NP use and Mod use. Classifiers cannot be attached to genitives. Examples from the Kozhikode variety are given below.

### Animate referent

#### NP use

- (46) **magel-o**      tugel-yā-pashi      buddivant āsa  
1SG.GEN-MSG    2SG.GEN-OBL-COMP    intelligent    be.PRS  
‘My (MSG) is more intelligent than your (MSG).’
- (47) **magel-i**      tugel-yā-pashi      buddivant āsa  
1SG.GEN-FSG    2SG.GEN-OBL-COMP    intelligent    be.PRS  
‘My (FSG) is more intelligent than your (FSG).’
- (48) \***magel-o-jaṇ**      **tugel-o-jaṇ-ā-pashi**      buddivant  
1SG.GEN-MSG-CLF    2SG.GEN-MSG-CLF-OBL-COMP    intelligent  
āsa  
be.PRS  
*to mean*, ‘My (MSG) is more intelligent than your (MSG).’
- (49) \***magel-i-jaṇ**      **tugel-i-jaṇ-ā-pashi**      buddivant  
1SG.GEN-FSG-CLF    2SG.GEN-FSG-CLF-OBL-COMP    intelligent  
āsa  
be.PRS  
*to mean*, ‘My (FSG) is more intelligent than your (FSG).’

### Mod use

- (50) **magel-o**      **chello**      tugel-o      chell-yā-pashi  
1SG.GEN-MSG    son.MSG    2SG.GEN-MSG    son-OBL-COMP  
buddivant āsa  
intelligent    be.PRS  
‘My son (MSG) is more intelligent than your son (MSG).’

- (51) **magel-i**                      **chelli**                      **tugel-i**                      chelli-pashi  
 1SG.GEN-FSG daughter.FSG 2SG.GEN-FSG daughter.OBL-COMP  
 buddivant āsa  
 intelligent be.PRS  
 ‘My daughter (FSG) is more intelligent than your daughter (FSG).’
- (52) \***magel-o-jaṇ**                      chello                      **tugel-o-jaṇ**                      chell-yā-pashi  
 1SG.GEN-MSG-CLF son.MSG 2SG.GEN-MSG-CLF son-OBL-COMP  
 buddivant āsa  
 intelligent be.PRS  
*to mean*, ‘My son (MSG) is more intelligent than your son (MSG).’
- (53) \***magel-i-jaṇ**                      chelli                      **tugel-i-jaṇ**                      chelli-pashi  
 1SG.GEN-FSG-CLF daughter.FSG 2SG.GEN-FSG-CLF daughter.OBL-COMP  
 buddivant āsa  
 intelligent be.PRS  
*to mean*, ‘My daughter (FSG) is more intelligent than your daughter (FSG).’

### Inanimate referent

#### Masculine

##### NP use

- (54) **magel-{o/e}**                      **tugel-yā-pashi**                      porn-o  
 1SG.GEN-{M/ EPC} 2SG.GEN-EPC-OBL-COMP old-MSG  
 ‘My (MSG) is older than your (MSG).’

##### Mod use

- (55) **magel-{e/o}**                      **mancho**                      **tugel-e**                      **manch-ā-pashi**                      porn-o  
 1SG.GEN-{EPC/M} cot.MSG 2SG.GEN-EPC cot-OBL-COMP old-MSG  
 ‘My cot (MSG) is older than your cot (MSG).’

#### Feminine

##### NP use

- (56) **magel-{i/e}**                      **tugel-yā-pashi**                      porn-i  
 1SG.GEN-{F/EPC} 2SG.GEN-EPC-OBL-COMP old-FSG  
 ‘My (FSG) is older than your (FSG).’

##### Mod use

- (57) **magel-{e/i}**                      **gāḍi**                      **tugel-yā**                      **gāḍi-pashi**  
 1SG.GEN-{EPC/F} vehicle.FSG 2SG.GEN-EPC vehicle-OBL-COMP  
 porn-i  
 old-FSG  
 ‘My vehicle (FSG) is older than your vehicle (FSG).’

## Neuter

### NP use

- (58) **magel-e**                      **tugel-yā-pashi**                      porn-e  
1SG.GEN-NSG                      2SG.GEN-EPC-OBL-COMP                      old-NSG  
'My (NSG) is older than your (NSG).'

### Mod use

- (59) **magel-e**                      **ghaḍiyār**                      **tugel-yā**                      **ghaḍiyār-ā-pashi**  
1SG.GEN-NSG    watch/clock.NSG                      2SG.GEN-EPC    watch/clock-OBL-COMP  
porn-e  
old-NSG  
'My watch/clock (NSG) is older than your watch/clock (NSG).'

**NP use** of inanimate referent: neuter gender form used

- (60) y-e                      **magel-e**  
this-EPC    1SG.GEN-EPC  
'This is mine (MSG/FSG/NSG).'

**NP use** of animate referent: gender of the referent marked (M or F)

- (61) y-o                      **magel-o**                      y-i                      **magel-i**  
this-MSG    1SG.GEN-MSG                      this-FSG    1SG.GEN-FSG  
'This is mine (MSG).'

In the Kozhikode Konkani, note that the gender distinction on the GEN is maintained both in NP use as well as Mod use in the case of animate referents. However, in the case of inanimate referents, there is a tendency to neutralize the gender on the GEN in the Mod use. In the NP use, gender neutralization is possible but retaining gender is preferred as shown in the examples above. We did not notice gender neutralization in the Canacona variety, as shown in the examples below.

## Inanimate referent

### Masculine

#### NP use

- (62) h-o                      **mhāgel-o**                      āni  
this-MSG    1SG.GEN-MSG    and  
t-o                      mhāge                      **bhāv-ā-l-o**  
that-MSG    1SG.GEN                      brother-OBL-GEN-MSG  
'This is mine (MSG) and that is my brother's (MSG).'

**Mod use**

- (63) h-ɔ            **mhāgel-o**            **mobāil**            āni  
                          this-MSG   1SG.GEN-MSG            mobile.MSG            and  
                          t-ɔ            mhāge            **bhāv-ā-l-ɔ**            **mobāil**  
                          that-MSG   1SG.GEN            brother-OBL-GEN-MSG            mobile.MSG  
                          ‘This is my mobile and that is my brother’s mobile.’

**Feminine****NP use**

- (64) h-i            **mhāgel-i**            āni            t-i  
                          this-FSG            1SG.GEN-FSG            and            that-FSG  
                          mhāgel(-yā)            **āvay-l-i**  
                          1SG.GEN(-OBL)            mother-GEN-FSG  
                          ‘This is mine (FSG) and that is my mother’s (FSG).’

**Mod use**

- (65) h-i            **mhāgel-i**            **sāḍi**            āni  
                          this-FSG            1SG.GEN-FSG            saree.FSG            and  
                          t-i            mhāgel(-yā)            **āvay-l-i**            **sāḍi**  
                          that-FSG            1SG.GEN(-OBL)            mother-GEN-FSG            saree.FSG  
                          ‘This is my saree and that is my mother’s saree.’

**Neuter****NP use**

- (66) h-ẽ            **mhāge-l-ẽ**            ani            t-ẽ  
                          this-NSG            1SG.GEN-NSG            and            that-NSG  
                          mhāgel(-yā)            / mhāg-yā            **bhāv-ā-l-ẽ**  
                          1SG.GEN(-OBL)/ 1SG.GEN-OBL            brother-OBL-GEN-NSG  
                          ‘This is mine (NSG) and that is my brother’s (NSG).’

**Mod use**

- (67) h-ẽ            **mhāge-l-ẽ**            **pen**            ani            t-ẽ  
                          this-NSG   1SG.GEN-NSG            pen.NSG            and            that-NSG  
                          mhāgel(-yā)            / mhāg-yā            **bhāv-ā-l-ẽ**            **pen**  
                          1SG.GEN(-OBL)/ 1SG.GEN-OBL            brother-OBL-GEN-NSG            pen.NSG  
                          ‘This is my pen and that is my brother’s pen.’

As for gender neutralization in the case of inanimate referent, we noticed difference between the Canacona and the Kozhikode variety. However, further work is needed to understand more about gender neutralization through comparative work with other varieties.



#### [4] Adjectives

Adjectives in Konkani come in two varieties: those which decline for gender and number and those which do not. Here we will report on the declinable adjectives. In all varieties of Konkani we worked on, declinable adjectives have 3-way gender marked: masculine (M), feminine (F), and neuter (N). They are unequivocally acceptable in NP use and Mod use. Examples from the Kudubi variety are given below.

##### Masculine referent

###### NP use

- (68) **dhav-o**                      **kāl-y-ā-pekshā**                      **mhārag**                      (āsā)  
white-MSG                      black-OBL-COMP                      expensive                      be.PRS  
'The white one (MSG) is costlier than the black one (MSG).'

###### Mod use

- (69) **dhav-o**                      **mobāil**                      **kāl-y-ā**                      **mobāil-ā-pekshā**  
white-MSG                      mobile.MSG                      black-OBL                      mobile.MSG-OBL-COMP  
**mhārag**                      (āsā)  
expensive                      be.PRS  
'The white mobile (MSG) is costlier than the black mobile (MSG).'

##### Feminine referent

###### NP use

- (70) **dhav-i**                      **kāl-y-ā-pekshā**                      **mhārag**                      (āsā)  
white-FSG                      black-OBL-COMP                      expensive                      be.PRS  
'The white one (FSG) is costlier than the black one (FSG).'

###### Mod use

- (71) **dhav-i**                      **sāḍi**                      **kāl-y-ā**                      **sāḍ-ye-pekshā**  
white-FSG                      saree.FSG                      black-OBL                      saree.FSG-OBL-COMP  
**mhārag**                      (āsā)  
expensive                      be.PRS  
'The white saree (FSG) is costlier than the black saree (FSG).'

##### Neuter referent

###### NP use

- (72) **dhākt-ā**                      **vhodl-y-ā-pekshā**                      **mhārag/savāy**                      (āsā)  
small-NSG                      big-OBL-NSG-COMP                      expensive/cheap                      be.PRS  
'The small one (NSG) is {expensive/cheaper} than the big one (NSG).'

### Mod use

- (73) **dhākt-ā**                      **pen**                      vhoḍl-y-ā                      pen-ā-pekshā  
small-NSG                      pen.NSG                      big-OBL-NSG                      pen.NSG-OBL-COMP  
mhārag/savāy                      (āsā)  
expensive/cheap                      be.PRS  
'The small pen is {expensive/cheaper} than the big pen.'

Classifier cannot be attached to adjectives. Noun homophonous with CLF can host the adjective, as shown in the examples from the Kozhikode Konkani.

- (74) \***thor-jaṇ**    chelle                      **bhāgel-jaṇ**                      chell-yā-n-pashi                      buddivant    āsa  
fat-CLF    boy.PL                      slim-CLF                      boy-OBL-PL-COMP                      intelligent    be.PRS  
*to mean*, 'Fat boys are more intelligent than slim boys.'

- (75) **thor**                      **jaṇ**                      **bhāgel**                      **jaṇ-ā-n-pashi**                      buddivant    āsa  
fat                      people                      slim                      people-OBL-PL-COMP                      intelligent    be.PRS  
'Fat people are more intelligent than slim people.'

In the Kudubi variety examples of adjectives above, gender neutralization on ADJ in the case of inanimate referent was not observed. However, in the Kozhikode variety, gender neutralization on ADJ in the case of inanimate referent is observed, both in NP use and Mod use, as shown in the examples below.

### Masculine referent

#### NP use

- (76) **dhav{-o/e}**                      kāl-yā-pashi                      porn{-o/e}  
white.SG{-MSG/NSG}                      black-OBL-COMP                      old-{MSG/NSG}  
'The white one (MSG/NSG) is older than that black one (MSG/NSG).'

#### Mod use

- (77) **dhav{-o/e}**                      mancho                      kāl-e                      manch.ā-pashi  
white.SG{-MSG/NSG}                      small.bed.MSG                      black-OBL                      small.bed.OBL-COMP  
porn{-o/e}  
old-{MSG/NSG}  
'The white small bed (MSG/NSG) is older than that black small bed (MSG/NSG).'

### Feminine referent

#### NP use

- (78) **dhav{-i/e}**                      kāl-yā-pashi                      porn{-i/e}  
white.SG{-FSG/NSG}                      black-OBL-COMP                      old-{FSG/NSG}  
'The white one (FSG/NSG) is older than that black one (FSG/NSG).'

**Mod use**

- (79) **dhav{-i/e}**                      **gāḍi**                      *kāl-yā*                      *gāḍi-pashi*  
 white.SG {-FSG/NSG} vehicle.FSG                      black-OBL vehicle.OBL-COMP  
 porn-{i/e}  
 old-{FSG/NSG}  
 ‘The white vehicle (FSG/NSG) is older than that black vehicle (FSG/NSG).’

**Neuter referent****NP use**

- (80) **dhav{-e}**                      *kāl-yā-pashi*                      *porn-e*  
 white.SG {-NSG}                      black-OBL-COMP                      old-NSG  
 ‘The white one (NSG) is older than the black one (NSG).’

**Mod use**

- (81) **dhav{-e}**                      **ghaḍiyār**                      *kāl-yā*                      *ghaḍiyār-ā-pashi*  
 white.SG {-NSG} watch/clock.NSG                      black-OBL watch/clock-OBL-COMP  
 porn-e  
 old-NSG  
 ‘The white watch/clock (NSG) is older than the black watch/clock (NSG).’

However, in the case of indefinite referent, gender is expressed in the Kozhikode variety.

**Masculine referent****NP use**

- (82) *makkā*                      *ek*                      **dhav-o**                      *jāyi*  
 1SG.DAT                      one                      white-MSG                      want  
 ‘I want a white one (MSG).’

**Mod use**

- (83) *makkā*                      *ek*                      **dhav-o**                      **mancho**                      *jāyi*  
 1SG.DAT                      one                      white-MSG                      cot.MSG                      want  
 ‘I want a white cot (MSG).’

**Feminine referent****NP use**

- (84) *makkā*                      *ek*                      **dhav-i**                      *jāyi*  
 1SG.DAT                      one                      white-FSG                      want  
 ‘I want a white one (FSG).’

**Mod use**

- (85) *makkā*                      *ek*                      **dhav-i**                      **gāḍi**                      *jāyi*  
 1SG.DAT                      one                      white-FSG                      vehicle.FSG                      want  
 ‘I want a white vehicle (FSG).’

## Neuter referent

### NP use

- (86) makkā                      ek                      **dhav-e**                      jāyi  
1SG.DAT                      one                      white-NSG                      want  
'I want a white one (NSG).'

### Mod use

- (87) makkā                      ek                      **dhav-e**                      **ghaḍiyār**                      jāyi  
1SG.DAT                      one                      white-NSG                      watch/clock.NSG                      want  
'I want a white watch/clock (NSG).'

Gender neutralization on ADJ (both in Mod use and NP use) seems to be happening in the Kozhikode variety, in the case of definite/specific inanimate referents. We did not notice it in the Kudubi variety, as shown in the examples above. However, further work is needed to see whether it happens in other varieties as well.

Finally, classifier cannot be attached to adjectives. Noun homophonous with CLF can host the adjective.

- (88) \***thor-jaṇ**    chelle                      **bhāgel-jaṇ**    chell-yā-n-pashi                      buddivant    āsa  
fat-CLF    boy.PL                      slim-CLF    boy-OBL-PL-COMP                      intelligent    be.PRS  
*to mean*, 'Fat boys are more intelligent than slim boys.'

- (89) **thor**                      **jaṇ**                      **bhāgel**                      **jaṇ-ā-n-pashi**                      buddivant    āsa  
fat                      people                      slim                      people-OBL-PL-COMP                      intelligent    be.PRS  
'Fat people are more intelligent than slim people.'

## [5] Verb-based nominalizations

### [5a] Relative clauses

The so-called relative or adjectival clause in Konkani can be formed by (i) the 'non-reduction strategy' (relative-correlative type) where head noun appears in the modifying clause, and (ii) the gap strategy (participle type) where there is no trace of the head noun in modifying clause (absence of trace or gap). In the varieties of Konkani spoken in Goa (Pernem and Cancona) speakers' spontaneous response is the relative-correlative type construction. Participle types are not provided but confirmation is needed whether they are grammatically possible. The examples of relative-correlative type from the Pernem variety are given below.

**NP use** [subject of intransitive verb, Masculine referent]

- (90) (**dz-o**)      sofā-cher    bas-l-ā                    **t-o**                    mhādz-o  
REL-MSG sofa-LOC    sit-PST-MSG      COREL-MSG      1SG.GEN-MSG  
mitra  
friend  
‘[The one (MSG) sitting on the sofa] is my friend.’

**Mod use** [subject of intransitive verb, Masculine referent]

- (91) (**dz-o**      bhurg-o)    sofā-cher                    bas-l-ā                    **t-o**  
REL-MSG child-MSG sofa-LOC                    sit-PST-MSG                    COREL-MSG  
bhurg-o                    mhādz-o                    mitra  
child-MSG                    1SG.GEN-MSG                    friend  
‘[The boy sitting on the sofa] is my friend.’

**NP use** [subject of intransitive verb, feminine referent]

- (92) (**j-i**)            sofā-cher                    bas-l-yā                    **t-i**  
REL-FSG sofa-LOC                    sit-PST-FSG                    COREL-FSG  
mhādz-i                    maitriṇ  
1SG.GEN-FSG friend.FSG  
‘[The one (FSG) sitting on the sofa] is my friend.’

**Mod use** [subject of intransitive verb, feminine referent]

- (93) (**j-i**            bāil)                    sofā-cher                    bas-l-yā                    **t-i**  
REL-FSG woman.FSG sofa-LOC                    sit-PST-FSG                    COREL-FSG  
bāil                    mhādz-i                    maitriṇ  
woman-FSG                    1SG.GEN-FSG friend.FSG  
‘[The woman sitting on the sofa] is my friend.’

**NP use** [subject of intransitive verb, neuter referent]

- (94) (**dz-ā**)      sofa-cher                    bas-l-ā                    **t-ā**  
REL-NSG sofa-LOC                    sit-PST-NSG                    COREL-NSG  
ṣāṇā                    (āṣā)  
clever                    be.PRS  
‘[The one (NSG) sitting on the sofa] is clever.’

**Mod use** [subject of intransitive verb, neuter referent]

- (95) (**dz-ā**      cheḍu)      sofa-cher                    bas-l-ā                    **t-ā**  
REL-NSG girl-NSG sofa-LOC                    sit-PST-NSG                    COREL-NSG  
cheḍu                    ṣāṇā                    (āṣā)  
girl-NSG clever                    be.PRS  
‘[The girl (NSG) sitting on the sofa] is clever.’

In a sharp contrast to varieties spoken in Goa, participle type construction is the spontaneous response of the consultants of the Kozhikode variety, as shown below. Note that classifier cannot be attached to V-based nominalizations.

**NP use** [subject of intransitive verb, masculine referent]

- (96) **bank-ā-ri**                      **bashil-o-(\*jaṇ)**                      m{a/i}gel-{o/e}                      dost  
 bench-OBL-LOC                      sit.PST-MSG-CLF                      1SG.GEN-MSG/EPC                      friend  
 ‘The one (MSG) sitting on the bench is my friend.’

**Mod use** [subject of intransitive verb, masculine referent]

- (97) **bank-ā-ri**                      **bashil-o-(\*jaṇ)**                      **chell-o**                      m{a/i}gel-{o/e}                      dost  
 bench-OBL-LOCsit.PST-MSG-CLF                      child-MSG 1SG.GEN-MSG/EPC                      friend  
 ‘The boy (MSG) sitting on the bench is my friend.’

**NP use** [Subject as a referent and feminine]

- (98) **bank-ā-ri**                      **bashil-i-(\*jaṇ)**                      m{a/i}gel-{i/e}                      dost  
 bench-OBL-LOCsit.PST-FSG-CLF                      1SG.GEN-FSG/EPC                      friend  
 ‘The one (FSG) sitting on the bench is my friend.’

**Mod use** [Subject as a referent and feminine]

- (99) **bank-ā-ri**                      **bashil-i-(\*jaṇ)**                      **chell-i**                      m{a/i}gel-{i/e}                      dost  
 bench-OBL-LOCsit.PST-FSG-CLF                      child-FSG 1SG.GEN-FSG/EPC                      friend  
 ‘The girl (FSG) sitting on the bench is my friend.’

**NP use** [Subject as a referent and neuter]

- (100) **bank-ā-ri**                      **bashil-e**                      m{a/i}gel-e  
 bench-OBL-LOC                      sit.PST-NSG                      1SG.GEN-NSG  
 ‘[The one (NSG) sitting on the bench] is mine.’

**Mod use** [Subject as a referent and neuter]

- (101) **bank-ā-ri**                      **bashil-e**                      **suṇe**                      m{a/i}gel-e  
 bench-OBL-LOC                      sit.PST-NSG                      dog-NSG 1SG.GEN-NSG  
 ‘The dog (NSG) sitting on the bench is mine.’

**NP use** [Direct object as a referent and masculine]

- (102) **āve**                      **kāli**                      **dekhil-o**                      buddivant                      āsa  
 1SG.NOM yesterday                      see.PST-MSG                      intelligent                      be.PRS  
 ‘The one (MSG) I saw yesterday is intelligent.’

**Mod use** [Direct object as a referent and masculine]

- (103) **āve**            **kāli**            **dekhil-{o/e}**            **chell-o**  
 1SG.NOM yesterday see.PST- {MSG/EPC} child-MSG  
 buddivant āsa  
 intelligent be.PRS  
 ‘The boy (MSG) I saw yesterday is intelligent.’

**NP use** [Direct object as a referent and feminine]

- (104) **āve**            **kāli**            **dekhil-i**            buddivant āsa  
 1SG.NOM yesterday see.PST-FSG intelligent be.PRS  
 ‘The one (MSG) I saw yesterday is intelligent.’

**Mod use** [Direct object as a referent and feminine]

- (105) **āve**            **kāli**            **dekhil-{i/e}**            **chell-i**  
 1SG.NOM yesterday see.PST- {FSG/EPC} child-FSG  
 buddivant āsa  
 intelligent be.PRS  
 ‘The girl (FSG) I saw yesterday is intelligent.’

**NP use** [Direct object as a referent and neuter]

- (106) **āve**            **kāli**            **dekhil-e**            buddivant āsa  
 1SG.NOM yesterday see.PST-NSG intelligent be.PRS  
 ‘The one (NSG) I saw yesterday is intelligent.’

**Mod use** [Direct object as a referent and neuter]

- (107) **āve**            **kāli**            **dekhil-e**            **suṇe**            buddivant āsa  
 1SG.NOM yesterday see.PST-NSG dog-NSG intelligent be.PRS  
 ‘The dog (NSG) I saw yesterday is intelligent.’

Examples (96) through (107) demonstrates that in the Kozhikode variety gender neutralization occurs in the Mod use but not in the NP use of the participial type V-based nominalizations. Whether this happens in other varieties as well needs further investigation.

In the Kozhikode variety, no REL-COREL type construction was attested. Even after providing coined versions of REL-COREL type construction, speakers reject them. They instead use COREL in both clauses as shown below.

**Mod use** [Subject as a referent and masculine]

- (108) **\*j-o**            chell-o            bank-ā-ri            bashil-o            āsa  
 REL-MSG child-MSG bench-OBL-LOC sit.PST-MSG be.PRS  
**t-o**            magel-o            dost  
 COREL-MSG 1SG.GEN-MSG friend  
*to mean*, ‘The boy (MSG) sitting on the bench is my friend.’

- (109) (t-o)      chell-o              bank-ā-ri              bashil-o              āsa  
           that-MSG child-MSG      bench-OBL-LOC      sit.PST-MSG      be.PRS  
           **t-o**              magel-o              dost  
           that-MSG      1SG.GEN-MSG      friend  
           ‘The boy (MSG) who is sitting on the bench is my friend.’

**Mod use** [Subject as a referent and feminine]

- (110) \*j-i      chell-i      bank-ā-ri              bashil-i              āsa  
           REL-FSG child-FSG bench-OBL-LOC      sit.PST-FSG      be.PRS  
           **t-i**              magel-i              dost/dost-iṇi  
           COREL-FSG 1SG.GEN-FSG      friend/friend-F  
           *to mean*, ‘The girl (FSG) who is sitting on the bench is my friend.’

- (111) (t-i)      chell-i      bank-ā-ri              bashil-i              āsa  
           that-FSG child-FSG bench-OBL-LOC      sit.PST-FSG      be.PRS  
           **t-i**              magel-i              dost/dost-iṇi  
           that-FSG 1SG.GEN-FSG      friend/friend-F  
           ‘The girl (FSG) who is sitting on the bench is my friend.’

**Mod use** [Subject as a referent and neuter]

- (112) \*j-e      suṇe      bank-ā-ri              bashil-e              āsa  
           REL-FSG dog-NSG bench-OBL-LOC      sit.PST-NSG      be.PRS  
           **t-e**              suṇe              magel-e  
           COREL-NSG dog-NSG 1SG.GEN-NSG  
           *to mean*, ‘The dog (FSG) that is sitting on the bench is my friend.’

- (113) (t-e)      suṇe      bank-ā-ri              bashil-e              āsa  
           that-FSG dog-NSG bench-OBL-LOC      sit.PST-NSG      be.PRS  
           **t-e**              suṇe              magel-e  
           that-NSG dog-NSG 1SG.GEN-NSG  
           ‘The dog (FSG) that is sitting on the bench is my friend.’

In a sharp contrast to Konkani varieties spoken in Goa, the variety spoken in Kozhikode does not use REL-COREL type construction at all. The construction using distal pronouns twice seem to be used instead of REL-COREL. Gender is marked on the REL and COREL pronoun in REL-COREL construction and on the participle form of the verb in the participial construction. Both these constructions permit NP use where there is no head to agree with. Previous research treats these constructions as clauses or sentences. Pardeshi and Shibatani (2023) argue that these constructions are grammatical nominalizations, which denote things, link nouns. Classifier cannot be attached to the participial form of the verb in the participial construction both in NP use and Mod use.

Construction using wh-pronoun given below was constructed and shown to the Kozhikode variety speakers. One consultant accepted it while the other rejected it.



(114) **kench-o**    chell-o                    bank-ā-ri                    bashil-o                    āsa  
                   who-MSG child-MSG                    bench-OBL-LOC                    sit.PST-MSG                    be.PRS  
                   t-o                    magel-o                    dost  
                   that-MSG 1SG.GEN-MSG friend  
                   ‘The boy (MSG) who is sitting on the bench is my friend.’

(115) **kench-i**    chell-i                    bank-ā-ri                    bashil-i                    āsa  
                   who-FSG child-FSG bench-OBL-LOC                    sit.PST-FSG                    be.PRS  
                   t-i                    magel-i                    dost/dost-ini  
                   that-FSG 1SG.GEN-FSG                    friend/friend-F  
                   ‘The girl (FSG) who is sitting on the bench is my friend.’

(116) **kench-e**    suṇe                    bank-ā-ri                    bashil-e                    āsa                    t-e  
                   who-NSG dog-NSG bench-OBL-LOC                    sit.PST-NSG                    be.PRS                    that-NSG  
                   suṇe                    magel-e  
                   dog-NSG 1SG.GEN-NSG  
                   ‘The dog (NSG) that is sitting on the bench is mine.’

## [5b] Complement clauses

We have not collected data on complement clauses so far.

We will now summarize the data presented above on grammatical nominalizations in all varieties of Konkani we studied in their NP use and Mod use in the form of Shibatani’s hierarchy constraining the distribution of classifier and gender marking (see Shibatani 2023: 35–37). It should be added that these summaries are based on a limited data and are tentative. They need to be confirmed in the follow-up work.

## CLF marking

Figure 1. Classifier marking pattern in Konkani (all varieties)

NP use:    \*OBL NUM >    \*DEM >    \*GEN >    \*ADJ >    \*V-based NMLZ  
 Mod use:    \*OBL NUM >    \*DEM >    \*GEN >    \*ADJ >    \*V-based NMLZ

(\* = no marking or impossible; \*OBL = obligatory with few exceptions such as numeral 1 to which CLF cannot be attached and G- forms of numeral 2 through to 4 to which CLF can be optionally attached)

## Gender marking

Figure 2. Gender marking pattern in Konkani (the Pernem variety)

NP use:    LIM NUM > OBL DEM > OBL GEN > OBL ADJ > OBL V-based NMLZ  
 Mod use:    LIM NUM > OBL DEM > OBL GEN > OBL ADJ > OBL V-based NMLZ

Figure 3. Gender marking pattern in Konkani (the Kozhikode variety)

NP use: <sup>LIM</sup>NUM ><sup>OBL</sup>DEM ><sup>OBL</sup>GEN ><sup>OBL</sup>ADJ ><sup>OBL</sup>V-based NMLZ

Mod use: <sup>LIM</sup>NUM ><sup>OBL/NEUT</sup>DEM ><sup>OBL/NEUT</sup>GEN ><sup>OBL/NEUT</sup>ADJ ><sup>OBL/NEUT</sup>V-based NMLZ

(<sup>LIM</sup> = limited to a few numerals 4 and below, <sup>OBL</sup>= obligatory; <sup>OBL/NEUT</sup>= obligatory in some cases while neutralization allowed in other cases)

#### 4. ANALYSIS AND DISCUSSION

From the tentative summaries provided above it is clear that the distribution of CLF and gender marking in Konkani can be explained in terms of Shibatani's (2023: 35) hierarchy.

In Konkani, there are two processes observed: (i) loss of gender and (ii) acquisition (use) of classifier, both of which are limited to the first category on Shibatani's hierarchy, namely numeral. In many Indo-European languages, numeral have lost gender: in many languages all numerals have lost gender altogether (English, Hindi-Urdu, among others), while in others, such as Russian, Spanish, Slovene, Romanian, Bulgarian, Konkani, among others, gender is retained on lower numerals (below 5). In Konkani, like Spanish, gender is retained only on cardinal numeral one. Hindi has lost gender on DEM in addition to NUM. As for classifier, standard Hindi does not have classifier, but its Southern neighbor, namely Marathi has acquired one classifier, such as Konkani, Assamese (eastern India), on the other hand, has lost gender altogether, but has acquired lot of classifiers which can be used on all categories of Shibatani's hierarchy. These distributions suggest that the loss of gender as well as acquisition of classifier, starts from the left-most category on Shibatani's hierarchy and then proceed rightward.

On gender neutralization in the Kozhikode variety, our preliminary observations suggest that there is a tendency to neutralize the gender in the Mod use in contrast to NP use where the preference is to mark the gender, in the case of inanimate referent. The gender neutralization in the Kozhikode variety needs further investigation.

#### 5. CONCLUSION

Konkani is an ideal language for studying nominalizations since it has both devices of nominalization, namely gender marking and classifiers. The limited distribution of classifier marking helps to understand the path of acquisition and spread of classifier marking. Loss of gender distinction (gender neutralization) offers a glimpse into the process of loss of gender marking.

As mentioned earlier, the observations made in this report are based on a limited set of data and there are gaps in the data as well. We need to cast our data net further and refine our description as well as analysis. We also need to do research on the diachronic development of the use of classifiers and/or gender markers with grammatical nominalizations and compare Konkani grammatical nominalizations with their counterparts in neighboring languages such as Marathi, Gujarati and Hindi.

## ABBREVIATIONS

|       |                        |
|-------|------------------------|
| 1     | first person           |
| 2     | second person          |
| 3     | third person           |
| ACC   | accusative             |
| ADJ   | adjective              |
| CLF   | classifier             |
| COMP  | comparative            |
| COREL | correlative marker     |
| DAT   | dative                 |
| DEM   | demonstrative          |
| EMPH  | emphatic particle      |
| EPC   | epicene                |
| F     | feminine               |
| GEN   | genitive noun phrase   |
| IMPF  | imperfective           |
| LOC   | locative               |
| M     | masculine              |
| N     | neuter                 |
| NOM   | nominative             |
| NUM   | numeral                |
| OBL   | oblique                |
| PF    | perfective             |
| PL    | plural                 |
| PRS   | present                |
| PST   | past                   |
| REL   | relative clause marker |
| SG    | singular               |

## ACKNOWLEDGMENTS

First, we would like to thank our Konkani consultants: Divya Pingulkar, Sakshi Thakur, Gaurav Satyawar Naik (Pernem), Sejal Pagi, Samarth Komarpant (Canacona), Janardhana Gowda, Narayana Gowda, Shivarama Gowda, Santhosh Kumar Gowda (Mangalore), Pradeep Kumar, Asha D. (Kozhikode), for sharing their invaluable intuition and for their cooperation and patience in answering to our questions. This report would not have been possible without the wholehearted cooperation from these consultants. We would also like to thank our local coordinators: (i) The officials of HPSM's Ganpat Parsekar College of Education, Harmal-Pernem, for providing space to conduct fieldwork and also for the hospitality extended to us, (ii) Dr. Ruta Paradkar for her help and cooperation in collecting data in Pernem and her insights on transcribing the data, (iii) Advocate Hrishikesh Kadam for providing space to conduct our fieldwork and also for the hospitality extended to us in Canacona, (iv) Dr. Sayeegeetha Hegade, Head (In-Charge), Department of Humanities, NITTE (Deemed To Be) University, Deralakatte, Mangaluru, Karnataka for her cooperation and support in finding and recruiting consultants of

Kudubi-Konkani, Koraga, Arebhashe and Tulu, which itself is a very difficult task, and also providing us a space to conduct fieldwork in an ideal set up, (v) Arun Shyam and Geetha Shyam for introducing us to Pradeep Kumar, Asha D. and also for providing much needed local hospitality and support, and (vi) Pradeep Kumar, Asha D. allowed us to use their home as a field station.

The present study has been supported by grants from the Japan Society for the Promotion of Science (KAKENHI Project IDs: 22H00659 and 23K21931, ‘Cross-linguistic studies on grammatical nominalizations: with a focus on classifiers and gender markers’, PI: Kazuhiro Kawachi) and from the National Institute for Japanese Language and Linguistics (‘Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective’, PI: Prashant Pardeshi).

## REFERENCES

- Corbett, Greville G. 2013. Number of Genders. In Matthew S. Dryer and Martin Haspelmath (eds.) WALS Online (v2020.4) [Data set]. Zenodo.  
<https://doi.org/10.5281/zenodo.13950591>
- Dhongade, Ramesh V. 2022. *Konkani*. (London Oriental and African Language Library, 21). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Dixon, Robert M. W. 1982. *‘Where have all the adjectives gone?’ and other essays in semantics and syntax*. Berlin: Mouton de Gruyter.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
[https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/\\_pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/_pdf/-char/en)

# Gender and classifier marking in Dravidian languages

Niranjan Uppoor<sup>1</sup>

## ABSTRACT

This study examines gender and classifier marking in select Dravidian languages from a nominalization perspective. The analysis is grounded in dialectal variation, fieldwork data, and historical sources. Dravidian languages generally exhibit gender marking on numerals, demonstratives, genitives, adjectives, and V-based forms in noun-phrase use. Malayalam, however, stands out as an exception to this general tendency. In the context of modification use, gender marking is more commonly found on numerals and less commonly on demonstratives. Notably, genitives, adjectives, and V-based forms do not display gender marking in modification use. Furthermore, classifier marking appears to be restricted exclusively to the numeral context. Overall, the patterns of gender and classifier marking observed in these Dravidian languages provide empirical support for Shibatani's (2023) hierarchy.

## 1. INTRODUCTION

This study investigates gender and classifier marking in the following Dravidian languages—Arebase (/areba:se/), Beary (/bya:ri/), Kannada (/kannaɖa/), Koraga (/koraga/), Malayalam (/malaya:lam/), Tamil (/tamil/), Tulu (/tuɭu/), and Telugu (/telugu/)—from the perspective of nominalization theory, following Shibatani (2023). The patterns of gender and classifier marking in these languages largely align with Shibatani's proposed hierarchy. In noun-phrase (NP) use, Dravidian languages typically exhibit gender marking on categories including numerals, demonstratives, genitives, adjectives, and V-based (~ verb-based) forms. However, in modification (Mod) use, gender marking is restricted to numerals and demonstratives, while genitives, adjectives, and V-based forms do not show gender distinctions. Classifier marking is explored in the context of numerals, although its occurrence on other structures requires further investigation.

Shibatani (2023) proposes a framework known as 'dynamic functional typology', designed to enhance the description and analysis of classifiers and gender marking across the world's languages. By treating classifiers and gender markers as nominalization devices, Shibatani introduces an implicational hierarchy—NUM > DEM > GEN > ADJ > V-based—that incorporates both structural and functional dimensions. This two-dimensional framework enables fine-grained cross-linguistic comparisons and offers constraints on synchronic distribution patterns, while also predicting historical developments in classifier and gender marking across constructions, functions, dialects, and languages. This framework brings to light various classifier/gender-marked constructions that do not accommodate a head noun

---

<sup>1</sup> Department of Linguistics and Tribal Languages, Faculty of Literature, Tripura University (A Central University), Suryamaninagar, West Tripura District, Tripura State, India, PIN 799022. Email: niranjan.uppoor@gmail.com

(popularly known as ‘headless constructions’), which Shibatani calls NP use and takes this as a starting point from where modification use, accommodating head noun of a structure on the above hierarchy, diachronically emerges. The present study applies Shibatani’s hierarchy to examine the use of classifiers and gender markers across the five structural types in the hierarchy—numeral (NUM), demonstrative (DEM), genitive (GEN), adjectival (ADJ), and verb-based (V-based)—in both NP-use and modification-use constructions. Focusing on various Dravidian languages, the study evaluates the utility of the dynamic functional typology in accounting for synchronic distribution patterns and in predicting pathways of historical development in classifier and gender marking.

The Dravidian languages form a major language family of South Asia, spoken across Pakistan (Brahui, a North Dravidian language), India (all other Dravidian languages), and Sri Lanka (Tamil). The long-distance genetic affiliation of the Dravidian family remains unresolved. Various hypotheses have been proposed linking Dravidian to other language families, including Elamite, Mongolic, and Japanese, though none have been conclusively established. The Dravidian family is typically classified into four subgroups: South Dravidian, South-Central Dravidian, Central Dravidian, and North Dravidian. Among the languages examined in this study, all except Telugu belong to the South Dravidian subgroup; Telugu is classified under the South-Central subgroup. Of the languages considered here, Kannada, Malayalam, Tamil, and Telugu are recognized as major Dravidian languages. In contrast, Arebase, Beary, Koraga, and Tulu are considered minor Dravidian languages, which have yet to receive scholarly attention on par with their major counterparts. Besides, Arebase is argued to be a dialectal variety of Kannada (Gowda 2019). Kannada is widely spoken in Karnataka, Malayalam in Kerala, Tamil in Tamil Nadu, and Telugu in the states of Andhra Pradesh and Telangana. Arebase is spoken in coastal Karnataka and the Kodagu region, while Beary, Koraga, and Tulu are spoken primarily along the coastal areas of Karnataka. Typologically, Dravidian languages are head-final and exhibit the following linear word orders: subject–object–verb (SOV), modifier–noun, and subordinate clause–main clause. They are agglutinative and primarily suffixing in nature. Nominalization is marked by suffixes, as illustrated in Kannada: *ha:ru* jump.IMP ‘jump’ (imperative), *ha:ruvudu* jump.NMLZ ‘the act of jumping’ (nominalized form).

This report is based on data collected from a range of primary and secondary sources since 2021. The research began with an initial investigation into nominalization patterns in major Dravidian languages, including Kannada, Malayalam, Tamil, and Telugu. This was followed by a series of fieldwork projects conducted in 2023 and 2024, focusing on lesser-studied Dravidian languages such as Arebase, Beary, Koraga, and Tulu. Preliminary findings from these investigations have been previously reported in Pardeshi et al. (2023), Shibatani (2023), Uppoor (2022), Uppoor (2024), and Uppoor (2025). The present paper offers a comprehensive account of the current understanding of gender and classifier marking in Dravidian languages, with reference to Shibatani’s (2023) hierarchy.

The data are presented using broad transcription. In line with the general conventions (Asher & Kumari 1997 or Krishnamurthi 2003, for example) in Dravidian linguistics, the symbol *j* represents a voiced palatal stop (i.e., *j* = IPA /j/) and the symbol *y* represents a voiced palatal approximant (i.e., *y* = IPA /j/). The final /a/ in the transcriptions may correspond to /ə/,

or, in the case of Tulu, to /u/. These variations do not affect the analysis. The symbol ‘\*’ indicates either ungrammaticality or reconstructed forms, while ‘!’ denotes marked or unusual constructions.

## 2. OVERVIEW OF GENDER AND CLASSIFIER MARKING AND GRAMMATICAL NOMINALIZATIONS IN DRAVIDIAN LANGUAGES

### 2.1 Gender and classifier marking in Dravidian languages

Dravidian languages are generally not rich in classifiers. Relevant to our discussion, here, are the human classifiers *a:l* (Beary), *a:l* (Malayalam), *jana* (Arebase, Kannada, Koraga, and Tulu), *mandi* (Kannada and Telugu), and *pe:r* ~ *per* (Tamil and Malayalam). These languages feature a natural or semantic gender system rather than a grammatical gender system. Human entities are typically classified as masculine or feminine, while all other entities—including animals and natural elements—are grouped under the neuter gender. The gender system varies across Dravidian languages, reflecting its evolution at different historical stages. Some languages distinguish three genders—masculine, feminine, and neuter—whereas others maintain a two-way distinction between masculine and non-masculine.

### 2.2 Grammatical nominalizations in Dravidian languages

A brief overview of some relevant features of nominalization is presented in this section.

#### [1] Numerals

##### [1a] Cardinal numerals

This study focuses on numeral ‘one’ and, therefore, numerals up to three are provided for each of the Dravidian languages under study here. These are the citation forms.

Table 1. Numerals in various Dravidian languages, citation forms

|           | One   | Two    | three |
|-----------|-------|--------|-------|
| Arebase   | ondu  | eraḍu  | mu:ru |
| Beary     | onnu  | jendə  | mu:nu |
| Kannada   | ondu  | eraḍu  | mu:ru |
| Koraga    | oŋji  | radḍə  | mu:ji |
| Malayalam | onnu  | raṇḍu  | mu:nu |
| Tamil     | onru  | iraṇṭu | munru |
| Tulu      | oŋji  | radḍa  | mu:ji |
| Telugu    | okaṭi | reṇḍu  | mu:ḍu |

##### [1b] Ordinal numerals

Ordinal numerals are not explored in this study.

## [2] Demonstratives

Given below are demonstratives in the Dravidian languages under study here. There are no plural forms of demonstratives in these languages. Likewise, the demonstratives do not indicate gender.

Table 2. Demonstratives in Dravidian languages

|           | Proximal | Distal |
|-----------|----------|--------|
| Arebase   | i:       | a:     |
| Beary     | i:       | a:     |
| Kannada   | i:       | a:     |
| Koraga    | i:       | a:     |
| Malayalam | i:       | a:     |
| Tamil     | inda     | anda   |
| Tulu      | i:       | a:     |
| Telugu    | i:       | a:     |

Pertinent to our discussion are the personal pronouns in Dravidian languages. These languages distinguish only two numbers—singular and plural—with no dual or trial forms. This two-way number distinction is consistently reflected in the personal pronouns. Gender distinctions of masculine, feminine, and neuter occur only in the third-person singular; no gender distinctions exist in the first and second persons or in the plural forms of the third person. In the plural, a distinction is made between human (H) and non-human (NH) referents. The table below presents the personal pronouns of the Dravidian languages examined so far. Some languages further distinguish between inclusive and exclusive first-person plural pronouns. Third-person pronouns also encode deictic reference: A-type pronouns denote distal persons, while I-type pronouns indicate proximal persons. As shown in the table, some Dravidian languages mark the third person with a masculine versus non-masculine distinction, whereas others maintain a three-way distinction among masculine, feminine, and neuter genders.



Table 3. Personal pronouns in Dravidian languages

|           | 1SG    | 1PL                      | 2SG   | 2PL     | 3SGM           | 3SGF           | 3SGN          | 3HPL              | 3NHPL             |
|-----------|--------|--------------------------|-------|---------|----------------|----------------|---------------|-------------------|-------------------|
| Arebase   | na:    | na:vu                    | ni:   | ni:vu   | avā<br>ivā     | ad<br>id       |               | avu<br>ivu        |                   |
| Beary     | nya:n  | naŋa                     | ni:n  | ni:ŋa   | avonu<br>onu   | avolu<br>olu   | ad<br>id      | aŋa<br>oŋa        | adaŋa<br>idaŋa    |
| Kannada   | na:nu  | na:vu                    | ni:nu | ni:vu   | avanu<br>ivanu | ava u<br>iva u | adu<br>idu    | avaru<br>ivaru    | avu<br>ivu        |
| Koraga    | ya:n   | eŋkulu                   | ijji  | nikulu  | a:yi<br>ikke   | addu<br>uttu   |               | akul<br>ikkel     | addu<br>uttu      |
| Malayalam | ja:n   | nammal<br>na:m<br>paŋŋal | ni:   | niŋŋal  | avan<br>ivan   | ava <br>iva    | atə<br>itə    | avar<br>ivar      | ava<br>iva        |
| Tamil     | na:n   | na:ŋgal<br>na:m          | ni:   | ni:ŋgal | avan<br>ivan   | ava <br>iva    | adu<br>idu    | avarga <br>ivarga | avayga <br>ivayga |
| Tulu      | ya:nuu | ŋkulu<br>nama            | i:    | nuguluu | a:ye<br>imbe   | a:lu<br>imbolu | avuu<br>induu | akulu<br>moku     | aykulu<br>ndekulu |
| Telugu    | ne:nu  | me:mu<br>manamu          | ni:wu | mi:ru   | wa:du<br>wi:du | adi<br>idi     | adi<br>idi    | wa:ru<br>wi:ru    | awi<br>iwi        |

### [3] Genitive noun phrases

Examples of nominal and pronominal genitive noun phrases in Kannada are provided below. As may be seen from the below paradigms, Kannada genitives carry gender marking in NP use with some exceptions. On the other hand, Kannada genitives do not carry any gender marking in Mod use. The Kannada patterns presented here more or less represent the Dravidian scenario of genitives.

- (1) Genitives in Kannada (South Dravidian), NP use, 1SG
 

|              |              |              |
|--------------|--------------|--------------|
| nannavanu    | nannava u    | nannaddu     |
| 1SG.GEN.3SGM | 1SG.GEN.3SGF | 1SG.GEN.3SGN |
| ‘mine (M)’   | ‘mine (F)’   | ‘mine (N)’   |
- (2) Genitives in Kannada (South Dravidian), NP use, 1PL
 

|              |              |              |
|--------------|--------------|--------------|
| nammavanu    | nammava u    | nammaddu     |
| 1PL.GEN.3SGM | 1PL.GEN.3SGF | 1PL.GEN.3SGN |
| ‘ours (M)’   | ‘ours (F)’   | ‘ours (N)’   |
- (3) Genitives in Kannada (South Dravidian), NP use, 2SG
 

|              |              |              |
|--------------|--------------|--------------|
| ninnavanu    | ninnava u    | ninnaddu     |
| 2SG.GEN.3SGM | 2SG.GEN.3SGF | 2SG.GEN.3SGN |
| ‘yours (M)’  | ‘yours (F)’  | ‘yours (N)’  |
- (4) Genitives in Kannada (South Dravidian), NP use, 2PL
 

|              |              |              |
|--------------|--------------|--------------|
| nimmavanu    | nimmava u    | nimmaddu     |
| 2PL.GEN.3SGM | 2PL.GEN.3SGF | 2PL.GEN.3SGN |
| ‘yours (M)’  | ‘yours (F)’  | ‘yours (N)’  |

- (5) Genitives in Kannada (South Dravidian), NP use, 3SGM  
!avanavanu !avanava[u] avanaddu  
3SGM.GEN.3SGM 3SGM.GEN.3SGF 3SGM.GEN.3SGN  
‘his (M)’ ‘his (F)’ ‘his (N)’
- (6) Genitives in Kannada (South Dravidian), NP use, 3SGF  
!ava[avanu] !ava[ava[u] ava[addu  
3SGF.GEN.3SGM 3SGF.GEN.3SGF 3SGF.GEN.3SGN  
‘hers (M)’ ‘hers (F)’ ‘hers (N)’
- (7) Genitives in Kannada (South Dravidian), NP use, 3SGN  
!adaravanu !adarava[u] adaraddu  
3SGN.GEN.3SGM 3SGN.GEN.3SGF 3SGN.GEN.3SGN  
‘its (M)’ ‘its (F)’ ‘its (N)’
- (8) Genitives in Kannada (South Dravidian), NP use, 3HPL  
!avaravanu !avarava[u] avaraddu  
3HPL.GEN.3SGM 3HPL.GEN.3SGF 3HPL.GEN.3SGN  
‘theirs (M)’ ‘theirs (F)’ ‘theirs (N)’
- (9) Genitives in Kannada (South Dravidian), NP use, 3NHPL  
!avuga[avanu] !avuga[ava[u] avuga[addu  
3NHPL.GEN.3SGM 3NHPL.GEN.3SGF 3NHPL.GEN.3SGN  
‘theirs (M)’ ‘theirs (F)’ ‘theirs (N)’
- (10) Genitives in Kannada (South Dravidian), NP use, 3SGM, nominal  
!tandeyavanu !tandeyava[u] tandeyaddu  
father.GEN.3SGM father.GEN.3SGF father.GEN.3SGN  
‘the father’s (M)’ ‘the father’s (F)’ ‘the father’s (N)’
- (11) Genitives in Kannada (South Dravidian), NP use, 3SGF, nominal  
!ta:yyivanu !ta:yyiva[u] ta:yyiaddu  
mother.GEN.3SGM mother.GEN.3SGF mother.GEN.3SGN  
‘the mother’s (M)’ ‘the mother’s (F)’ ‘the mother’s (N)’
- (12) Genitives in Kannada (South Dravidian), NP use, 3SGN, nominal  
maneyavanu maneyava[u] maneyaddu  
house.GEN.3SGM house.GEN.3SGF house.GEN.3SGN  
‘the house’s (M)’ ‘the house’s (F)’ ‘the house’s (N)’

- (13) Genitives in Kannada (South Dravidian), NP use, 3HPL, nominal  
 nenṭaravanu                      nenṭaravaḷu                      nenṭaraddu  
 relatives.GEN.3SGM                      relatives.GEN.3SGF                      relatives.GEN.3SGN  
 ‘the relatives’ (M)’                      ‘the relatives’ (F)’                      ‘the relatives’ (N)’
- (14) Genitives in Kannada (South Dravidian), NP use, 3NHPL, nominal  
 haḷḷigaḷavanu                      haḷḷigaḷavaḷu                      haḷḷigaḷaddu  
 villages.GEN.3SGM                      villages.GEN.3SGF                      villages.GEN.3SGN  
 ‘the villages’ (M)’                      ‘the villages’ (F)’                      ‘the villages’ (N)’

In the above patterns, the examples in (5), (6), (7), (8), (9), (10), and (11) seem to be marked or unusual constructions. In these instances, the masculine, feminine, and neuter pronominals are the 3SGM *avanu*, 3SGF *avaḷu*, 3SGN *adu*, 3HPL *avaru*, 3NHPL *avugaḷu*, and regular nominals, *tande* ‘father’ and *ta:yi* ‘mother’. In the same examples, the neuter suffix may appear with the same items (see the last columns) and they are perfectly grammatical. The genitives do not carry gender marking in Mod use. Given below are examples of Kannada genitives in Mod use.

- (15) Genitives in Kannada (South Dravidian), Mod use, 1SG  
 nanna                      huḍuga                      nanna                      huḍugi                      nanna                      mane  
 1SG.GEN                      boy                      1SG.GEN                      girl                      1SG.GEN                      house  
 ‘my boy’                                           ‘my girl’                                           ‘my house’
- (16) Genitives in Kannada (South Dravidian), Mod use, 1PL  
 namma                      huḍuga                      namma                      huḍugi                      namma                      mane  
 1PL.GEN                      boy                      1PL.GEN                      girl                      1PL.GEN                      house  
 ‘our boy’                                           ‘our girl’                                           ‘our house’
- (17) Genitives in Kannada (South Dravidian), Mod use, 2SG  
 ninna                      huḍuga                      ninna                      huḍugi                      ninna                      mane  
 2SG.GEN                      boy                      2SG.GEN                      girl                      2SG.GEN                      house  
 ‘your boy’                                           ‘your girl’                                           ‘your house’
- (18) Genitives in Kannada (South Dravidian), Mod use, 2PL  
 nimma                      huḍuga                      nimma                      huḍugi                      nimma                      mane  
 2PL.GEN                      boy                      2PL.GEN                      girl                      2PL.GEN                      house  
 ‘your boy’                                           ‘your girl’                                           ‘your house’
- (19) Genitives in Kannada (South Dravidian), Mod use, 3SGM  
 avana                      huḍuga                      avana                      huḍugi                      avana                      mane  
 3SGM.GEN                      boy                      3SGM.GEN                      girl                      3SGM.GEN                      house  
 ‘his boy’                                           ‘his girl’                                           ‘his house’

- (20) Genitives in Kannada (South Dravidian), Mod use, 3SGF  
 ava|a huḍuga ava|a huḍugi ava|a mane  
 3SGF.GEN boy 3SGF.GEN girl 3SGF.GEN house  
 ‘her boy’ ‘her girl’ ‘her house’
- (21) Genitives in Kannada (South Dravidian), Mod use, 3SGN  
 adara huḍuga adara huḍugi adara mane  
 3SGN.GEN boy 3SGN.GEN girl 3SGN.GEN house  
 ‘its boy’ ‘its girl’ ‘its house’
- (22) Genitives in Kannada (South Dravidian), Mod use, 3HPL  
 avara huḍuga avara huḍugi avara mane  
 3HPL.GEN boy 3HPL.GEN girl 3HPL.GEN house  
 ‘their boy’ ‘their girl’ ‘their house’
- (23) Genitives in Kannada (South Dravidian), Mod use, 3NHPL  
 !avuga|a huḍuga !avuga|a huḍugi avuga|a mane  
 3NHPL.GEN boy 3NHPL.GEN girl 3NHPL.GEN house  
 ‘their boy’ ‘their girl’ ‘their house’
- (24) Genitives in Kannada (South Dravidian), Mod use, 3SGM, nominal  
 tandeya huḍuga tandeya huḍugi tandeya mane  
 father.GEN boy father.GEN3SGM girl father.GEN house  
 ‘the father’s boy’ ‘the father’s girl’ ‘the father’s house’
- (25) Genitives in Kannada (South Dravidian), Mod use, 3SGF, nominal  
 ta:yiya huḍuga ta:yiya huḍugi ta:yiya mane  
 mother.GEN boy mother.GEN girl mother.GEN house  
 ‘the father’s boy’ ‘the father’s girl’ ‘the father’s house’
- (26) Genitives in Kannada (South Dravidian), Mod use, 3SGN, nominal  
 maneya huḍuga maneya huḍugi maneya pustaka  
 house.GEN boy house.GEN girl house.GEN book  
 ‘the house’s boy’ ‘the house’s girl’ ‘the house’s book’
- (27) Genitives in Kannada (South Dravidian), Mod use, 3HPL, nominal  
 nenṭara huḍuga nenṭara huḍugi nenṭara mane  
 relatives.GEN boy relatives.GEN girl relatives.GEN house  
 ‘the relatives’ boy’ ‘the relatives’ girl’ ‘the relatives’ book’

- (28) Genitives in Kannada (South Dravidian), Mod use, 3NHPL, nominal
- |                    |        |                     |        |                     |       |
|--------------------|--------|---------------------|--------|---------------------|-------|
| ha  iga a          | huḍuga | ha  iga a           | huḍugi | ha  iga a           | da:ri |
| villages.GEN       | boy    | villages.GEN        | girl   | villages.GEN        | path  |
| ‘the villages’ boy |        | ‘the villages’ girl |        | ‘the villages’ path |       |

In the above examples, (23) seems to be marked or unusual for pragmatic reasons. However, in the same example, the neuter suffix may appear with the 3NHPL and it is perfectly grammatical.

#### [4] Adjectives

Whether Dravidian adjectives constitute a distinct grammatical category remains an open question in the literature (see Krishnamurti 2003: 388–406 for a detailed discussion). Of particular relevance here is the fact that adjectives in Dravidian languages do not inflect for number, gender, or case when used as modifiers. Additionally, adjectives consistently precede the nouns they modify. The following examples from Kannada illustrate these features.

- (29) Adjectives in Kannada (South Dravidian), the Mod use
- |             |         |                 |          |                |             |
|-------------|---------|-----------------|----------|----------------|-------------|
| dodḍa       | huḍuga  | dodḍa           | huḍugige | dodḍa          | maraga a    |
| big         | boy.NOM | big             | girl.DAT | big            | tree.PL.GEN |
| ‘a big boy’ |         | ‘to a big girl’ |          | ‘of big trees’ |             |

In the above example, the adjective *dodḍa* ‘big’ modifies the nouns ‘boy’, ‘girl’, and ‘tree’. As mentioned previously, Dravidian gender is natural and not grammatical. Further, the masculine and feminine genders are typically assigned to humans and all other non-human entities take the neuter gender. Therefore, in the Kannada example in (29) above, the boy is masculine, the girl is feminine, and the tree is neuter. Besides, in Dravidian languages, there are only two numbers, the singular and the plural; they lack dual or trial number. In the Kannada example above, the boy and girl are in the singular number and the tree is in the plural number. In addition, the boy is in the nominative case, the girl is in the dative case, and the tree is in the genitive case. It may be seen that, in all the instances, the adjective *dodḍa* ‘big’ does not inflect for number, gender, and case in Mod use. On the other hand, when the adjectives appear in NP use, they carry number, gender, and case inflections as may be seen below.

- (30) Adjectives in Kannada (South Dravidian), NP use
- |                 |                    |                   |
|-----------------|--------------------|-------------------|
| dodḍavanu       | dodḍava ige        | dodḍavuga a       |
| big.M.NOM       | big.F.DAT          | big.NH.PL.GEN     |
| ‘a big one (M)’ | ‘to a big one (F)’ | ‘of big ones (N)’ |

As will be seen in the following sections, in our fieldwork, no exceptions to these features were found in other Dravidian languages.

## [5] V-based nominalizations

### [5a] Relative clauses

Dravidian languages employ a participial strategy to relativize clauses. The Relative clause precedes the head noun as may be seen in (32) below. The V-based forms carry gender marking only in NP use. There is no gender marking on the V-based forms in Mod use. The below examples from Kannada illustrate this aspect.

- (31) Relatives in Kannada (South Dravidian), NP use  
manege      bandavanu  
house.ALL come.PST.PTCP.3SGM  
‘the one who came (M) home’
- (32) Relatives in Kannada (South Dravidian), Mod use  
manege      banda                      maga  
house.ALL come.PSTPTCP   son  
‘the son who came to the house’

As will be shown in the following sections, our fieldwork found no exceptions to these features in other Dravidian languages. Another type of relative clause attested in Dravidian languages is the relative-corerelative construction. There is ongoing debate in the literature regarding its origin: some linguists argue that this construction is borrowed from Indo-Aryan languages, while others maintain that it is a native Dravidian strategy. This relative-corerelative type is not under study here. Furthermore, during our fieldwork, this construction did not appear in responses across the various Dravidian languages examined.

### [5b] Complement clauses

Complement clauses are not under study here.

## 3. DATA

This section addresses the following topics: (3.1) general patterns of gender marking in Dravidian languages; (3.2) the absence of gender marking on Malayalam genitives in NP use; (3.3) gender marking on demonstratives in Mod use; (3.4) a preliminary account of classifier marking in Dravidian languages; (3.5) the diachronic development of gender marking in Kannada; and (3.6) the absence of gender marking on the numeral ‘one’ in Tamil and Telugu in Mod use.

### 3.1 Gender marking in Dravidian languages: The story so far

The following section presents nominalization patterns in eight Dravidian languages, in the order: Arebase, Beary, Kannada, Koraga, Malayalam, Tamil, Tulu, and Telugu. Telugu is presented last, as it belongs to the South-Central Dravidian subgroup, whereas all the other languages examined here belong to the South Dravidian subgroup. The order of discussion follows Shibatani’s hierarchy: [1] numerals, [2] demonstratives, [3] genitives, [4] adjectives,

and [5] V-based forms. For each category, examples illustrating NP use are presented first, followed by examples of the Mod use.

### **Arebase (South Dravidian)**

#### **[1] Numerals**

In Arebase, numerals carry gender marking in NP use, as shown in the examples below.

- (33) Numeral in NP use, Arebase (South Dravidian)

obba        bat  
one.M      came  
'One (M) came.'

- (34) Numeral in NP use, Arebase (South Dravidian)

onda        bat  
one.F/N    came  
'One (F/N) came.'

In the context of Mod-use, Arebase numerals carry gender marking as well, as shown in the examples below.

- (35) Numeral in Mod use, Arebase (South Dravidian)

obba        gaṇḍasə    bat  
one.M      man        came  
'One (M) man came.'

- (36) Numeral in Mod use, Arebase (South Dravidian)

onda        hengasa    bat  
one.F       woman      came  
'One (F) woman came.'

- (37) Numeral in Mod use, Arebase (South Dravidian)

onda        na:yi bat  
one.N       dog        came  
'One (N) dog came.'

#### **[2] Demonstratives**

In the context of NP use, Arebase demonstratives exhibit gender marking, as shown in the examples below.

- (38) Demonstrative in NP use, Arebase (South Dravidian)

ivn            nannavn  
this.3SGM mine.3SGM  
'This (M) is mine (M).'

- (39) Demonstrative in NP use, Arebase (South Dravidian)

ida            nannada  
this.3SGF/N    mine.3SGF/N  
'This (F/N) is mine (F/N).'

Gender marking is optional on Arebase demonstratives in Mod use as may be seen below.

- (40) Demonstrative in Mod use, Arebase (South Dravidian)

i: /ivn            hayda            nannavan  
this/this.3SGM    boy            mine.3SGM  
'This/this (M) boy is mine (M).'

- (41) Demonstrative in Mod use, Arebase (South Dravidian)

i:/ida            guḍe nannada  
this/this.3SGF/N    girl    mine.3SGF/N  
'This/this (F/N) girl is mine (F/N).'

- (42) Demonstrative in Mod use, Arebase (South Dravidian)

i:/ida            pustaka            nannada  
this/this.3SGF/N    book            mine.3SGF/N  
'This/that (F/N) book is mine (F/N).'

### [3] Genitives

In the context of NP use, Arebase genitives carry gender marking, as shown in the examples below.

- (43) Genitive in NP use, Arebase (South Dravidian)

ivn            nannavn            avn            ninnavn  
this.3SGM mine.3SGM    that.3SGM yours.3SGM  
'This (M) is mine (M), that (M) is yours (M).'

- (44) Genitive in NP use, Arebase (South Dravidian)

ida            nannada            ada            ninnada  
this.3SGF/N    mine.3SGF/N    that.3SGF/N    yours.3SGF/N  
'This (F/N) is mine (F/N), that (F/N) is yours (F/N).'

In the context of Mod-use, on the other hand, genitives do not bear any gender marking, as shown in the examples below.



- (45) Genitive in Mod use, Arebase (South Dravidian)

nanna/ninna/avnana    appa  
my/your/his                father  
'my/your/his father'

- (46) Genitive in Mod use, Arebase (South Dravidian)

nanna/ninna/avnana    amma  
my/your/his                mother  
'my/your/his mother'

- (47) Genitive in Mod use, Arebase (South Dravidian)

nanna/ninna/avnana    na:yi  
My/your/his                dog  
'my/your/his dog'

#### [4] Adjectives

Arebase adjectives exhibit gender marking in NP use, as shown in the examples below.

- (48) Adjective in NP use, Arebase (South Dravidian)

doḍḍavn    nannavn                saṇṇavn                ninnava  
big.3SHM   mine.3SGM                small.3SGM                yours.3SGM  
'The big one (M) is mine (M), the small one (M) is yours (M).'

- (49) Adjective in NP use, Arebase (South Dravidian)

doḍḍada                nannada                saṇṇada                ninnada  
big.3SGF/N                mine.3SGF/N                small.3SGF/N                yours.3SGF/N  
'The big one (F/N) is mine (F/N), the small (F/N) one is yours (F/N).'

On the other hand, Arebase adjectives do not carry gender marking in Mod use.

- (50) Adjective in Mod use, Arebase (South Dravidian)

doḍḍa                hayda/guḍe/na:yi  
big                boy/girl/dog  
'a big boy/girl/dog'

- (51) Adjective in Mod use, Arebase (South Dravidian)

saṇṇa                hayda/guḍe/na:yi  
small                boy/girl/dog  
'a small boy/girl/dog'

#### [5] V-based

Arebase V-based forms carry gender marking in NP use.

- (52) V-based in NP use, Arebase (South Dravidian)  
 ninne bandavan nanna maṇa  
 yesterday came.PTCP.3SGM my son  
 ‘The one (M) who came yesterday is my son.’
- (53) V-based in NP use, Arebase (South Dravidian)  
 ninne bandada nanna maga/na:yi  
 yesterday came.PTCP.3SGF/N my daughter/dog  
 ‘The one (F/N) who came yesterday is my daughter/dog.’

On the other hand, Arebase V-based forms do not carry gender marking in Mod use.

- (54) V-based in Mod use, Arebase (South Dravidian)  
 ninne banda hayda/guḍe/na:yi nanna maṇa/maga/na:yi  
 yesterday came.PTCPboy/girl/dog my son/daughter/dog  
 ‘The boy/girl/dog who came yesterday is my son/daughter/dog.’

### **Beary (South Dravidian)**

#### [1] Numerals

Beary numerals carry gender marking in NP use.

- (55) Numeral in NP use, Beary (South Dravidian)  
 orta banna  
 one.3SGM came.3SGM  
 ‘One (M) came.’
- (56) Numeral in NP use, Beary (South Dravidian)  
 orti banna:la  
 one.3SGF came.3SGF  
 ‘One (F) came.’
- (57) Numeral in NP use, Beary (South Dravidian)  
 onnu banta  
 one.3SGN came.3SGN  
 ‘One (N) came.’

In the context of Mod use, Beary numerals may or may not carry gender marking.

- (58) Numeral in Mod use, Beary (South Dravidian)

orta/oru            a:ninga      banna  
 one.3SGM/one   boy            came.3SGM  
 ‘One (M)/one boy came.’

- (59) Numeral in Mod use, Beary (South Dravidian)

orti/oru            peninga      banna:la  
 one.3SGF/one   girl            came.3SGF  
 ‘One (F)/one female came.’

- (60) Numeral in Mod use, Beary (South Dravidian)

onnu/oru            na:yi banta  
 one.3SGN/one   dog   came.3SGN  
 ‘One (N)/one dog came.’

## [2] Demonstratives

Beary demonstratives carry gender marking in NP use.

- (61) Demonstrative in NP use, Beary (South Dravidian)

avonu      beliyavonu      onu      ceriyavonu  
 that.3SGM fair.3SGM      this.3SGM small.3SGM  
 ‘That (M) is a fair one (M), this (M) is a small one (M).’

- (62) Demonstrative in NP use, Beary (South Dravidian)

avolu      beliyavolu      olu      ceriyavolu  
 that.3SGF fair.3SGF      this.3SGF      small.3SGF  
 ‘That (F) is a fair one (F), this (F) is a small one (F).’

- (63) Demonstrative in NP use, Beary (South Dravidian)

ada      beliyada      ida      ceriyada  
 that.N      fair.N      this.N      small.N  
 ‘That (N) is a fair one (N), this (N) is a small one (N).’

On the other hand, Beary demonstratives do not carry gender marking in Mod use.

- (64) Demonstrative in Mod use, Beary (South Dravidian)

i:/a:      carka/pen/bu:ku  
 this/that   boy/girl/book  
 ‘this/that boy/girl/book’

## [3] Genitives

In the context of NP use, Beary genitives carry gender marking.

- (65) Genitive in NP use, Beary (South Dravidian)

onu            nandonu  
 this.3SGM mine.3SGM  
 ‘This (M) is mine (M).’

- (66) Genitive in NP use, Beary (South Dravidian)

olu            nandolu  
 this.3SGF mine.3SGF  
 ‘This (F) is mine (F).’

- (67) Genitive in NP use, Beary (South Dravidian)

ida            nandada  
 this.3SGN mine.3SGN  
 ‘This (N) is mine (N).’

In the context of Mod use, Beary demonstratives do not carry gender marking.

- (68) Genitive in Mod use, Beary (South Dravidian)

iya:lu            nando            abba  
 this.3HON.M my father  
 ‘This (M) is my father.’

- (69) Genitive in Mod use, Beary (South Dravidian)

ima            nando            umma  
 this.3HON.F my mother  
 ‘This (F) is my mother.’

- (70) Genitive in Mod use, Beary (South Dravidian)

ida            nando            na:yi  
 this.3SGN my dog  
 ‘This (N) is my dog.’

#### [4] Adjectives

Beary adjectives carry gender markings in NP use.

- (71) Adjective in NP use, Beary (South Dravidian)

uddaullonu            ningalo            monu  
 tall.be.3SGM your son  
 ‘The tall one (M) is your son.’

- (72) Adjective in NP use, Beary (South Dravidian)

uddaullolu      ningalo      molu  
tall.be.3SGF      your      daughter  
'The tall one (F) is your daughter.'

- (73) Adjective in NP use, Beary (South Dravidian)

vayaraullada      ningala      na:yi  
big.be.3SGN      your      dog  
'The tall one (N) is your dog.'

On the other hand, Beary demonstratives do not carry gender marking in Mod use.

- (74) Adjective in Mod use, Beary (South Dravidian)

avanu      beliyo      carka  
that.3SGM      fair      boy  
'That (M) is a fair boy.'

- (75) Adjective in Mod use, Beary (South Dravidian)

avalu      beliyo      peṇṇu  
that.3SGF      fair      girl  
'That (F) is a fair girl.'

- (76) Adjective in Mod use, Beary (South Dravidian)

ada      beliyo      na:yi  
that.3SGN      white      dog  
'That (N) is a white dog.'

[5] V-based

Beary V-based forms carry gender marking in NP use.

- (77) V-based in NP use, Beary (South Dravidian)

innale      bannonu      nando      mo:nu  
yesterday      came.PTCP.3SGM      my      son  
'The one (M) who came yesterday is my son.'

- (78) V-based in NP use, Beary (South Dravidian)

innale      bannolu      nando      mo:lu  
yesterday      came.PTCP.3SGF      my      daughter  
'The one (F) who came yesterday is my daughter.'

- (79) V-based in NP use, Beary (South Dravidian)  
innale bannada nando na:yi  
yesterday came.PTCP.3SGN my dog  
‘The one (N) who came yesterday is my dog.’

Beary V-based forms do not carry gender marking in Mod use.

- (80) V-based in Mod use, Beary (South Dravidian)  
innale banno carka/peṇṇa/na:yi nando monu/molun/na:yi  
yesterday came.PTCPboy/girl/dog my son/daughter/dog  
‘The boy/girl/dog who came yesterday is my son/daughter/dog.’

### Kannada (South Dravidian)

#### [1] Numerals

Kannada numerals carry gender marking in NP use.

- (81) Numeral in NP use, Kannada (South Dravidian)  
namage obba be:ku  
we.DAT one.3SGM want.NPST  
‘We need one (M).’
- (82) Numeral in NP use, Kannada (South Dravidian)  
namage obba[u] be:ku  
we.DAT one.3SGF want.NPST  
‘We need one (F).’
- (83) Numeral in NP use, Kannada (South Dravidian)  
namage ondu be:ku  
we.DAT one.3SGN want.NPST  
‘We need one (N).’

Kannada numerals carry gender marking in Mod use.

- (84) Numeral in Mod use, Kannada (South Dravidian)  
obba huḍuga  
one.3SGM boy  
‘one (M) boy’
- (85) Numeral in Mod use, Kannada (South Dravidian)  
obba[u] huḍugi  
one.3SGF girl  
‘one (F) girl’

- (86) Numeral in Mod use, Kannada (South Dravidian)  
 ondu mara  
 one.3SGN tree  
 ‘one (N) tree’

[2] Demonstratives

Kannada demonstratives carry gender marking in NP use.

- (87) Demonstrative in NP use, Kannada (South Dravidian)  
 avanu nammavanu  
 that.3SGM ours.3SGM  
 ‘That (M) is ours (M).’
- (88) Demonstrative in NP use, Kannada (South Dravidian)  
 ava[u] nammava[u]  
 that.3SGF ours.3SGF  
 ‘That (F) is ours (F).’
- (89) Demonstrative in NP use, Kannada (South Dravidian)  
 adu nammadu  
 that.3SGN ours.3SGN  
 ‘That (N) is ours (N).’

On the other hand, Kannada demonstratives do not carry gender marking in Mod use.

- (90) Demonstrative in Mod use, Kannada (South Dravidian)  
 a: huḍuga  
 that boy  
 ‘that boy’
- (91) Demonstrative in Mod use, Kannada (South Dravidian)  
 a: huḍugi  
 that girl  
 ‘that girl’
- (92) Demonstrative in Mod use, Kannada (South Dravidian)  
 a: ga:ḍi  
 that vehicle  
 ‘that vehicle’

[3] Genitives

Kannada genitives carry gender marking in NP use.

- (93) Genitive in NP use, Kannada (South Dravidian)

a: huḍuga nammavanu  
that boy ours.3SGM  
'That boy is ours (M).'

- (94) Genitive in NP use, Kannada (South Dravidian)

a: huḍugi nammava[u]  
that girl ours.3SGF  
'That girl is ours (F).'

- (95) Genitive in NP use, Kannada (South Dravidian)

a: pustaka nammadu  
that book ours.3SGN  
'That book is ours (N).'

Kannada genitives do not carry gender marking in Mod use.

- (96) Genitive in Mod use, Kannada (South Dravidian)

namma huḍuga  
our boy  
'our boy'

- (97) Genitive in Mod use, Kannada (South Dravidian)

namma huḍugi  
our girl  
'our girl'

- (98) Genitive in Mod use, Kannada (South Dravidian)

namma vishaya  
our matter  
'our matter'

#### [4] Adjectives

Kannada adjectives carry gender marking in NP use.

- (99) Adjective in NP use, Kannada (South Dravidian)

dodḍavanu namma geḷeya  
big.3SGM our friend  
'The big one (M) is our friend.'



- (100) Adjective in NP use, Kannada (South Dravidian)

dodḍavaḷu      namma      geḷati  
big.3SGF      our      friend  
'The big one (F) is our friend.'

- (101) Adjective in NP use, Kannada (South Dravidian)

dodḍadu      namma      mane  
big.3SGN      our      house  
'The big one is our house.'

Kannada adjectives do not carry gender marking in Mod use.

- (102) Adjective in Mod use, Kannada (South Dravidian)

dodḍa      huḍuga  
big      boy  
'big boy'

- (103) Adjective in Mod use, Kannada (South Dravidian)

dodḍa      huḍugi  
big      girl  
'big girl'

- (104) Adjective in Mod use, Kannada (South Dravidian)

dodḍa      mane  
big      house  
'big house'

[5] V-based

Kannada V-based forms carry gender marking in NP use.

- (105) V-based in NP use, Kannada (South Dravidian)

manege      bandavanu      nanna      geḷeya  
house.ALL      come.PST.PTCP.3SGM      my      friend  
'The one who came to my house is my friend.'

- (106) V-based in NP use, Kannada (South Dravidian)

manege      bandavaḷu      nanna      geḷati  
house.ALL      come.PST.PTCP.3SGF      my      friend  
'The one who came to my house is my friend.'

- (107) V-based in NP use, Kannada (South Dravidian)  
 manege                      badaddu                      nanna                      ga:dʒi  
 house.ALL                      come.PST.PTCP.3SGN my                      vehicle  
 ‘The one who came to my house is my vehicle.’

On the other hand, Kannada V-based forms do not carry gender marking in Mod use.

- (108) V-based in Mod use, Kannada (South Dravidian)  
 manege                      banda                      maga/maga[u/ga:dʒi  
 house.ALL                      come.PST.PTCP                      son/daughter/vehicle  
 ‘The son/daughter/ (vehicle) who (which) came to the house.’

### **Koraga (South Dravidian)**

#### **[1] Numerals**

Koraga numerals carry gender marking in NP use.

- (109) Numeral in NP use, Koraga (South Dravidian)  
 ori                      bakki  
 one.3SGM came.3SGM  
 ‘One (M) came.’

- (110) Numeral in NP use, Koraga (South Dravidian)  
 orti                      bakka  
 one.3SGF came.3SGF  
 ‘One (F) came.’

- (111) Numeral in NP use, Koraga (South Dravidian)  
 onni                      bakka  
 one.3SGN came.3SGN  
 ‘One (N) came.’

Koraga numerals carry gender marking in Mod use.

- (112) Numeral in Mod use, Koraga (South Dravidian)  
 ori                      gaŋde                      bakki  
 one.3SGM man                      came.3SGM  
 ‘One (M) man came.’

- (113) Numeral in Mod use, Koraga (South Dravidian)  
 onji    poŋŋu                      bakka  
 one woman                      came.3SGF  
 ‘One woman came.’

- (114) Numeral in Mod use, Koraga (South Dravidian)

onji na:yi bakka  
 one dog came.3SGN  
 'One dog came.'

## [2] Demonstratives

Koraga demonstratives carry gender marking in NP use.

- (115) Demonstrative in NP use, Koraga (South Dravidian)

a:yi/ikke anna student  
 that.3SGM/this.3SGM my student  
 'That (M)/this (M) is my student.'

- (116) Demonstrative in NP use, Koraga (South Dravidian)

adda/uttu anna student  
 that.3SGF/N/this.3SGF/N my student  
 'That (F/N)/this (F/N) is my student.'

- (117) Demonstrative in NP use, Koraga (South Dravidian)

adda/uttu anna na:yi  
 that.3SGF/N/this.3SGF/N my dog  
 'That (F/N)/this (F/N) is my dog.'

Koraga demonstratives may or may not carry gender marking in Mod use.

- (118) Demonstrative in Mod use, Koraga (South Dravidian)

a:/a:yi ku:je anna student  
 that/that.3SGM boy my student  
 'That/that (M) boy is my student.'

- (119) Demonstrative in Mod use, Koraga (South Dravidian)

a:/adda ku:ji anna student  
 that/that.3SGF/N girl my student  
 'That/ that (F/N) girl is my student.'

- (120) Demonstrative in Mod use, Koraga (South Dravidian)

a:/adda na:yi anna na:yi  
 that/that.3SGF/N dog my dog  
 'That/that (F/N) dog is my dog.'

## [3] Genitives

Koraga genitives carry gender marking in NP use.

- (121) Genitive in NP use, Koraga (South Dravidian)

akla                      annakla  
that.3M.HON      mine.M.HON  
'That (M) is mine (husband).'

- (122) Genitive in NP use, Koraga (South Dravidian)

adda                      annada  
that.3SGF/N      mine.3SGF  
'That (F/N) is mine (wife).'

- (123) Genitive in NP use, Koraga (South Dravidian)

adda                      anna  
that.3SGF/N      mine  
'That (F/N) is mine.'

Koraga genitives do not carry gender marking in Mod use.

- (124) Genitive in Mod use, Koraga (South Dravidian)

akla                      anna                      kora:yi  
that.3M.HON      my                      husband  
'That (M) is my husband.'

- (125) Genitive in Mod use, Koraga (South Dravidian)

adda                      anna                      korti  
that.3SGF/N      my                      wife  
'That (F/N) is my wife.'

- (126) Genitive in Mod use, Koraga (South Dravidian)

adda                      anna                      bu:ku  
that.3SGF/N      my                      book  
'That (F/N) is my book.'

#### [4] Adjectives

Koraga adjectives carry gender marking in NP use.

- (127) Adjective in NP use, Koraga (South Dravidian)

malla:yi      bakki  
big.3SGM      came.3SGM  
'The big one (M) came.'

- (128) Adjective in NP use, Koraga (South Dravidian)

mallada      bakka  
big.3SGF    came.3SGF/N  
'The big one (F) came.'

- (129) Adjective in NP use, Koraga (South Dravidian)

mallavu      bakka  
big.3SGN    came.3SGF/N  
'The big one (N) came.'

On the other hand, Koraga adjectives do not carry gender marking in Mod use.

- (130) Adjective in Mod use, Koraga (South Dravidian)

mallā          ku:je          bakki  
big            boy            came.3SGM  
'The big boy came.'

- (131) Adjective in Mod use, Koraga (South Dravidian)

mallā          ku:ji          bakka  
big            girl            came.3SGF/N  
'The big girl came.'

- (132) Adjective in Mod use, Koraga (South Dravidian)

mallā          na:yi          bakka  
big            dog            came.3SGF/N  
'The big dog came.'

#### [5] V-based

Koraga V-based forms carry gender marking in NP use.

- (133) V-based in NP use, Koraga (South Dravidian)

eru                      bareyi                      anna mage  
yesterday              came.PTCP.3SGM      my    son  
'The one (M) who came yesterday is my son.'

- (134) V-based in NP use, Koraga (South Dravidian)

eru                      baredda                      anna magal  
Yesterday              came.PTCP.3SGF/N      my    daughter  
'The one (F) who came yesterday is my daughter.'

- (135) V-based in NP use, Koraga (South Dravidian)  
 eru                      baredda                      anna na:yi  
 yesterday              came.PTCP.3SGF/N      my      dog  
 ‘The one (N) who came yesterday is my dog.’

On the other hand, Koraga V-based forms do not carry gender marking in Mod use.

- (136) V-based in Mod use, Koraga (South Dravidian)  
 eru              bare                      kuje/kuji/na:yi              anna  
 yesterday      came.PTCP              boy/daughter/dog              my  
 mage/magal/na:yi  
 son/daughter/dog  
 ‘The boy/girl/dog who/which came yesterday is my son/daughter/dog.’

### Malayalam (South Dravidian)

#### [1] Numerals

Malayalam numerals carry gender marking in NP use.

- (137) Numeral in NP use, Malayalam (South Dravidian)  
 oruvan/oruttan              vannu  
 one.3SGM/one.3SGM      came  
 ‘One (M) came.’

- (138) Numeral in NP use, Malayalam (South Dravidian)  
 oruva/orutti              vannu  
 one.3SGF/one.3SGF      came  
 ‘One (F) came.’

- (139) Numeral in NP use, Malayalam (South Dravidian)  
 onnu              vannu  
 one.3SGN      came  
 ‘One (N) came.’

Malayalam numerals do not carry gender marking in Mod use.

- (140) Numeral in Mod use, Malayalam (South Dravidian)  
 oru      a:ŋkutti  
 one      boy  
 ‘one boy’

- (141) Numeral in Mod use, Malayalam (South Dravidian)

oru penkutti

one girl

‘one girl’

- (142) Numeral in Mod use, Malayalam (South Dravidian)

oru vi:da

one house

‘one house’

## [2] Demonstratives

Malayalam demonstratives do carry gender marking in NP use.

- (143) Demonstrative in NP use, Malayalam (South Dravidian)

avan nammude kuttuka:ran

that.3SGM our friend.3SGM

‘That (M) is our friend.’

- (144) Demonstrative in NP use, Malayalam (South Dravidian)

aval nammude kuttuka:ri

that.3SGF our friend.3SGF

‘That (F) is our friend.’

- (145) Demonstrative in NP use, Malayalam (South Dravidian)

ada nammude ka:ryam

that.3SGN our matter

‘That (N) is our matter.’

On the other hand, Malayalam demonstratives do not carry gender marking in Mod use.

- (146) Demonstrative in Mod use, Malayalam (South Dravidian)

a: payyan

that man

‘that man’

- (147) Demonstrative in Mod use, Malayalam (South Dravidian)

a: sthri:

that woman

‘that woman’

- (148) Demonstrative in Mod use, Malayalam (South Dravidian)

a:     vaṇḍi  
that   vehicle  
‘that vehicle’

### [3] Genitives

Malayalam genitives do not carry gender marking in NP use. This is an exception to the Shibatani’s hierarchy since the Malayalam adjectives and V-based forms carry gender marking in NP use as will be seen later. This issue is discussed in the later parts of the paper.

- (149) Genitive in NP use, Malayalam (South Dravidian)

a:             manushyan             nammuḍe   a:ṇa  
that         man                   ours         be  
‘That man is ours.’

- (150) Genitive in NP use, Malayalam (South Dravidian)

a:             sthree                 nammuḍe   a:ṇa  
that         woman               ours         be  
‘That woman is ours.’

- (151) Genitive in NP use, Malayalam (South Dravidian)

a:             vaṇḍi                 nammuḍe   a:ṇa  
that         vehicle               ours         be  
‘That vehicle is ours.’

Malayalam genitives do not carry gender marking in Mod use.

- (152) Genitive in Mod use, Malayalam (South Dravidian)

eṇḍe     payyan  
our     man  
‘our man’

- (153) Genitive in Mod use, Malayalam (South Dravidian)

eṇḍe     sthri:  
our     woman  
‘our woman’

- (154) Genitive in Mod use, Malayalam (South Dravidian)

eṇḍe     vaṇḍi  
our     vehicle  
‘our vehicle’



#### [4] Adjectives

Malayalam adjectives carry gender marking in NP use.

- (155) Adjective in NP use, Malayalam (South Dravidian)

valiyavan      nammoḍe      a:ṇa  
big.3SGM      ours              be  
'The big one is ours.'

- (156) Adjective in NP use, Malayalam (South Dravidian)

valiyavaḷ      nammoḍe      a:ṇa  
big.3SGF      ours              be  
'The big one is ours.'

- (157) Adjective in NP use, Malayalam (South Dravidian)

valida              nammoḍe      a:ṇa  
big.3SGN      our              be  
'The big one is ours.'

On the other hand, Malayalam adjectives do not carry gender marking in Mod use.

- (158) Adjective in Mod use, Malayalam (South Dravidian)

nalla      payyan  
nice      man  
'a nice man'

- (159) Adjective in Mod use, Malayalam (South Dravidian)

nalla      sthri:  
nice      woman  
'a nice woman'

- (160) Adjective in Mod use, Malayalam (South Dravidian)

nalla      vaṇḍi  
nice      vehicle  
'a nice vehicle'

#### [5] V-based

Malayalam V-based forms carry gender marking in NP use.

- (161) V-based in NP use, Malayalam (South Dravidian)

viṭṭil              vannavan              eṇḍe      makan  
house.ALL      came.PST.PTCP.3SGM      my      son  
'The one (M) who came to the house is my son.'

- (162) V-based in NP use, Malayalam (South Dravidian)  
 vittil vannaval eṇḍe makal  
 house.ALL came.PSTPTCP.3SGF my daughter  
 ‘The one (F) who came to the house is my daughter.’

- (163) V-based in NP use, Malayalam (South Dravidian)  
 vittil vannada eṇḍe vaṇḍi  
 house.ALL came.PSTPTCP.3SGN my vehicle  
 ‘The one (N) that came to the house is my vehicle.’

On the other hand, Malayalam V-based forms do not carry gender marking in Mod use.

- (164) V-based in Mod use, Malayalam (South Dravidian)  
 vittil vanna makan/makal/vaṇḍi  
 house.ALL come.PSTPTCP son/daughter/vehicle  
 ‘the son/daughter who came to the house/ the vehicle that came to the house’

### **Tamil (South Dravidian)**

#### **[1] Numerals**

Tamil numerals carry gender marking in NP use.

- (165) Numeral in NP use, Tamil (South Dravidian)  
 oruvan vanda:n  
 one.3SGM came.3SGM  
 ‘One (M) came.’

- (166) Numeral in NP use, Tamil (South Dravidian)  
 oruval vanda  
 one.3SGF came.3SGF  
 ‘One (F) came.’

- (167) Numeral in NP use, Tamil (South Dravidian)  
 onna vantata  
 one.3SGN came.3SGN  
 ‘One (N) came.’

Tamil numerals do not carry gender marking in Mod use.

- (168) Numeral in Mod use, Tamil (South Dravidian)  
 oru payyan  
 one boy  
 ‘one boy’

- (169) Numeral in Mod use, Tamil (South Dravidian)

oru      poṇṇu  
one      girl  
'one girl'

- (170) Numeral in Mod use, Tamil (South Dravidian)

oru      vi:ḍu  
one      house  
'one house'

## [2] Demonstratives

Tamil demonstratives carry gender marking in NP use.

- (171) Demonstrative in NP use, Tamil (South Dravidian)

avan              ennoḍeya      nanban  
that.3SGM      our              friend  
'That (M) is our friend.'

- (172) Demonstrative in NP use, Tamil (South Dravidian)

ava[              ennoḍeya      nanbi  
that.3SGF      our              friend  
'That (F) is our friend.'

- (173) Demonstrative in NP use, Tamil (South Dravidian)

ada              ennoḍeya      vishayam  
that.3SN      our              matter  
'That (N) is our matter.'

Tamil demonstratives do not carry gender marking in Mod use.

- (174) Demonstrative in Mod use, Tamil (South Dravidian)

anda      payyan  
that      man  
'that man'

- (175) Demonstrative in Mod use, Tamil (South Dravidian)

anda      ponnu  
that      woman  
'that woman'

- (176) Demonstrative in Mod use, Tamil (South Dravidian)  
 anda      vaṇḍi  
 that      vehicle  
 ‘that vehicle’

### [3] Genitives

Tamil genitives carry gender marking in NP use.

- (177) Genitive in NP use, Tamil (South Dravidian)  
 anda              payyan      ennoḍeyavan  
 that              man              ours.3SGM  
 ‘That man is ours (M).’

- (178) Genitive in NP use, Tamil (South Dravidian)  
 anda              peṇ              enneḍeyavaḷ  
 that              woman              ours.3SGF  
 ‘That woman is ours (F).’

- (179) Genitive in NP use, Tamil (South Dravidian)  
 anda              puttakam      ennoḍeyada  
 that              book              ours.3SGN  
 ‘That book is ours (N).’

On the other hand, Tamil genitives do not carry gender marking in Mod use.

- (180) Genitive in Mod use, Tamil (South Dravidian)  
 ennoḍeya              nanban  
 our                      friend.3SGM  
 ‘our friend (M)’

- (181) Genitive in Mod use, Tamil (South Dravidian)  
 ennoḍeya              nanbi  
 our                      friend.3SGF  
 ‘our friend (F)’

- (182) Genitive in Mod use, Tamil (South Dravidian)  
 ennoḍeya              vishayam  
 our                      matter  
 ‘our matter’

### [4] Adjectives

Tamil adjectives carry gender marking in NP use.

- (183) Adjective in NP use, Tamil (South Dravidian)

avan                      nallavan  
that.3SGM              nice.3SGM  
'That (M) is nice (M).'

- (184) Adjective in NP use, Tamil (South Dravidian)

ava[                      na][ava]  
that.3SGF              nice.3SGF  
'That (F) is nice (F).'

- (185) Adjective in NP use, Tamil (South Dravidian)

ada                      nallada  
that.3SGN              nice.3SGN  
'That (N) is nice.'

On the other hand, Tamil adjectives do not carry gender marking in Mod use.

- (186) Adjective in Mod use, Tamil (South Dravidian)

nalla      payyan  
nice      man  
'a nice man'

- (187) Adjective in Mod use, Tamil (South Dravidian)

nalla      sthri:  
nice      woman  
'a nice woman'

- (188) Adjective in Mod use, Tamil (South Dravidian)

nalla      vaṇḍi  
nice      vehicle  
'a nice vehicle'

[5] V-based

Tamil V-based forms carry gender marking in NP use.

- (189) V-based in NP use, Tamil (South Dravidian)

vittiku      vandavan                      ennuḍaya      magan  
house.ALL      come.PST.PTCP.3SGM      my      son  
'The one (M) who came to the house is my son.'

- (190) V-based in NP use, Tamil (South Dravidian)  
 vittiku vandaval ennuḍaya magal  
 house.ALL come.PST.PTCP.3SGF my daughter  
 ‘The one (F) who came to the house is my daughter.’

- (191) V-based in NP use, Tamil (South Dravidian)  
 vittiku vantatu ennuḍaya vaṇḍi  
 house.ALL come.PST.PTCP.3SGN my vehicle  
 ‘The one (N) who came to the house is my vehicle.’

Tamil V-based forms do not carry gender marking in Mod use.

- (192) V-based in Mod use, Tamil (South Dravidian)  
 vittiku vanda magan/magal/vaṇḍi  
 house.ALL come.PST.PTCP son/daughter/vehicle  
 ‘the son/daughter who came to the house/the vehicle that came to the house’

### **Tulu (South Dravidian)**

#### **[1] Numerals**

Tulu numerals carry gender marking in NP use.

- (193) Numeral in NP use, Tulu (South Dravidian)  
 ko:ḍe ori batte  
 yesterday one.3SGM came.3SGM  
 ‘One (M) came yesterday.’

- (194) Numeral in NP use, Tulu (South Dravidian)  
 ko:ḍe orti battal  
 yesterday one.3SGF came.3SGF  
 ‘One (F) came yesterday.’

- (195) Numeral in NP use, Tulu (South Dravidian)  
 ru:muḍa oṇḍi uṇḍu  
 room.LOC one.N be.PRE.N  
 ‘There is one (N) in the room.’

Tulu numerals carry gender marking in Mod use.

- (196) Numeral in Mod use, Tulu (South Dravidian)  
 ko:ḍe ori a:njovu batte  
 yesterday one.3SGM man came.3SGM  
 ‘One (M) man came yesterday.’

- (197) Numeral in Mod use, Tulu (South Dravidian)  
 ko:ɖe                      orti                      ponjovu                      battal  
 yesterday                      one.3SGF                      woman                      came. F  
 ‘One (F) woman came yesterday.’

- (198) Numeral in Mod use, Tulu (South Dravidian)  
 ru:muɖa                      oŋji                      na:yi                      uŋɖu  
 room.LOC                      one.N                      dog                      be.PRE.N  
 ‘There is one (N) dog in the room.’

## [2] Demonstratives

Tulu demonstratives carry gender marking in NP use.

- (199) Demonstrative in NP use, Tulu (South Dravidian)  
 a:ye/imbe                      husa:r  
 that.3SGM/this.3SGM                      intelligent  
 ‘That (M)/this (M) is intelligent.’

- (200) Demonstrative in NP use, Tulu (South Dravidian)  
 a:l/imba:l                      husa:r  
 that.3SGF/this.3SGF                      intelligent  
 ‘That (F)/this (F) is intelligent.’

- (201) Demonstrative in NP use, Tulu (South Dravidian)  
 avu/undu                      hecca                      krayada  
 that.3SGN/this.3SGN                      high                      price.N  
 ‘That (N)/this (N) is expensive.’

Tulu demonstratives may or may not carry gender marking in Mod use.

- (202) Demonstrative in Mod use, Tulu (South Dravidian)  
 a:/a:ye                      a:ŋ                      husa:r  
 that/that.3SGM                      boy                      intelligent  
 ‘That/this (M) boy is intelligent.’

- (203) Demonstrative in Mod use, Tulu (South Dravidian)  
 a:/a:l                      poŋŋu                      husa:r  
 that/that.3SGF                      girl                      intelligent  
 ‘That/this (F) girl is intelligent.’

- (204) Demonstrative in Mod use, Tulu (South Dravidian)  
a:/avu                      mobaylu    hecca            krayada  
that/that.3SGN    mobile            high            price.N  
‘That/this (N) mobile is expensive.’

### [3] Genitives

Tulu genitives carry gender marking in NP use.

- (205) Genitive in NP use, Tulu (South Dravidian)  
a:r                              enna:r  
that.M.HON            mine.M.HON  
‘That (M) is mine (M).’

- (206) Genitive in NP use, Tulu (South Dravidian)  
a:l                              enna:l  
that.3SGF            mine.3SGF  
‘That (F) is mine (F).’

- (207) Genitive in NP use, Tulu (South Dravidian)  
avu                              ennavu  
that.3SGN            mine.3SGN  
‘That (N) is mine.’

Tulu genitives do not carry gender marking in Mod use.

- (208) Genitive in Mod use, Tulu (South Dravidian)  
a:ye                              enna            kaṇḍani  
that.3SGM            my            husband  
‘That (M) is my husband.’

- (209) Genitive in Mod use, Tulu (South Dravidian)  
a:l                              enna            boḍedi  
that.3SGF            my            wife  
‘That (F) is my wife.’

- (210) Genitive in Mod use, Tulu (South Dravidian)  
avu                              enna            pustaka  
that.3SGN            my            book  
‘That (N) is my book.’

### [4] Adjectives

Tulu adjectives carry gender marking in NP use.



- (211) Adjective in NP use, Tulu (South Dravidian)  
 torada:ye          sapurada:yedda          mast          husa:r  
 fat.3SGM          lean.3SGM.PP          very          intelligent  
 ‘The fat one (M) is intelligent than the lean one (M).’

- (212) Adjective in NP use, Tulu (South Dravidian)  
 torada:l          sapurada:ledda          mast          husa:r  
 fat.3SGF          lean.3SGF.PP          very          intelligent  
 ‘The fat one (F) is intelligent than the lean one (F).’

- (213) Adjective in NP use, Tulu (South Dravidian)  
 bolduvu          kappudedda          heccakrayada  
 white.3SGN          black.3SGN.PP          high price.N  
 ‘The white one (N) is expensive than the black one (N).’

Tulu adjectives do not carry gender marking in Mod use.

- (214) Adjective in Mod use, Tulu (South Dravidian)  
 torada          a:ŋ          sapurada          a:ŋadda          mast          husa:r  
 fat          boy          lean          boy.PP          high          intelligent  
 ‘The fat boy is intelligent than the lean boy.’

- (215) Adjective in Mod use, Tulu (South Dravidian)  
 torada          poŋŋu          sapurada          poŋŋadda          mast          husa:r  
 fat          girl          lean          girl.PP          very          intelligent  
 ‘The fat girl is intelligent than the lean girl.’

- (216) Adjective in Mod use, Tulu (South Dravidian)  
 boldu          mobaylukappu          mobayliqda          hecca          krayada  
 white          mobile          black          mobile.PP          high          price.N  
 ‘The white mobile is expensive than the black mobile.’

[5] V-based

Tulu V-based forms carry gender marking in NP use.

- (217) V-based in NP use, Tulu (South Dravidian)  
 bencida          mitta          kulludunna:ye          enna          dosti  
 bench.GEN          on          sitting.3SGM          my          friend  
 ‘The one (M) sitting on the bench is my friend.’

- (218) V-based in NP use, Tulu (South Dravidian)  
 bencida mitta kulludunna:l enna dosti  
 bench.GEN on sitting.3SGF my friend  
 ‘The one (F) sitting on the bench is my friend.’

- (219) V-based in NP use, Tulu (South Dravidian)  
 bencida mitta kulludunnavu enna na:yi  
 bench.GEN on sitting.3SGN my dog  
 ‘The one (N) sitting on the bench is my dog.’

Tulu V-based forms do not carry gender marking in Mod use.

- (220) V-based in Mod use, Tulu (South Dravidian)  
 bencida mitta kulluduna a:ŋ/pon enna dosti  
 bench.GEN on sitting boy/girl my friend  
 ‘The boy/girl who is sitting on the bench is my friend.’

- (221) V-based in Mod use, Tulu (South Dravidian)  
 bencida mitta kulluduna na:yi enna na:yi  
 bench.GEN on sitting dog my dog  
 ‘The dog which is sitting on the bench is my dog.’

### **Telugu (South-Central Dravidian)**

#### **[1] Numerals**

Telugu numerals carry gender marking in NP use.

- (222) Numeral in NP use, Telugu (South-Central Dravidian)  
 okkaḍu vacca:ḍu  
 one.3SGM came.3SGM  
 ‘One (M) came.’

- (223) Numeral in NP use, Telugu (South-Central Dravidian)  
 okkarti vaccindi  
 one.3SGF came.3SGF  
 ‘One (F) came.’

- (224) Numeral in NP use, Telugu (South-Central Dravidian)  
 okkaṭi vaccindi  
 one.3SGN came.3SGN  
 ‘One (N) came.’

Telugu numerals do not carry gender marking in Mod use.

(225) Numeral in Mod use, Telugu (South-Central Dravidian)  
 okka abba:yi  
 one boy  
 ‘one boy’

(226) Numeral in Mod use, Telugu (South-Central Dravidian)  
 okka amma:yi  
 one girl  
 ‘one girl’

(227) Numeral in Mod use, Telugu (South-Central Dravidian)  
 okka illu  
 one house  
 ‘one house’

## [2] Demonstratives

Telugu demonstratives carry gender marking in NP use.

(228) Demonstrative in NP use, Telugu (South-Central Dravidian)  
 va:ɖu ma: snehituɖu  
 that.3SGM our friend  
 ‘That (M) is our friend.’

(229) Demonstrative in NP use, Telugu (South-Central Dravidian)  
 a:me ma: snehitura:lu  
 that.3SGF our friend  
 ‘That (F) is our friend.’

(230) Demonstrative in NP use, Telugu (South-Central Dravidian)  
 adi ma: vishayamu  
 that.3SGN our matter  
 ‘That (N) is our matter.’

Telugu demonstratives do not carry gender marking in Mod use.

(231) Demonstrative in Mod use, Telugu (South-Central Dravidian)  
 a: abba:yi  
 that boy  
 ‘that boy’

- (232) Demonstrative in Mod use, Telugu (South-Central Dravidian)

a:            amma:yi  
that           girl  
'that girl'

- (233) Demonstrative in Mod use, Telugu (South-Central Dravidian)

a:            i||u  
that           house  
'that house'

[3] Genitives

Telugu genitives carry gender marking in NP use.

- (234) Genitive in NP use, Telugu (South-Central Dravidian)

a:    abba:yi        manava:du  
that   boy            ours.3SGM  
'That boy is ours (M).'

- (235) Genitive in NP use, Telugu (South-Central Dravidian)

a:    abba:yi        manadi  
that   boy            ours.3SGF/N  
'That girl is ours (F/N).'

- (236) Genitive in NP use, Telugu (South-Central Dravidian)

a:    pustakamu    manadi  
that   book            ours.3SGF/N  
'That book is ours (F/N).'

Telugu genitives do not carry gender marking in Mod use.

- (237) Genitive in Mod use, Telugu (South-Central Dravidian)

ma:            snehitudu  
our            friend.3SGM  
'our friend (M)'

- (238) Genitive in Mod use, Telugu (South-Central Dravidian)

ma:            snehigura:lu  
our            friend.F  
'our friend (F)'

- (239) Genitive in Mod use, Telugu (South-Central Dravidian)  
 ma: vishayamu  
 our matter  
 ‘our matter’

#### [4] Adjectives

Telugu adjectives carry gender marking in NP use.

- (240) Adjective in NP use, Telugu (South-Central Dravidian)  
 peddava:ḍu ma: snehituḍu  
 big.3SGM our friend.3SGM  
 ‘The big one (M) is our friend (M).’

- (241) Adjective in NP use, Telugu (South-Central Dravidian)  
 peddadi ma: snehitura:lu  
 big.3SGF/N our friend.SGF  
 ‘The big one (F/N) is our friend.’

- (242) Adjective in NP use, Telugu (South-Central Dravidian)  
 peddadi ma: illu  
 big.3SGF/N our friend  
 ‘The big one (F/N) is our house.’

On the other hand, Telugu adjectives do not carry gender marking in Mod use.

- (243) Adjective in Mod use, Telugu (South-Central Dravidian)  
 mapci abba:yi  
 nice boy  
 ‘a nice boy’

- (244) Adjective in Mod use, Telugu (South-Central Dravidian)  
 mapci amma:yi  
 nice girl  
 ‘a nice girl’

- (245) Adjective in Mod use, Telugu (South-Central Dravidian)  
 mapci illu  
 nice house  
 ‘a nice house’

#### [5] V-based

Telugu V-based forms carry gender marking in NP use.

- (246) V-based in NP use, Telugu (South-Central Dravidian)  
 inʃiki            vaccinava:ɖu            ma: koɖuku  
 house.ALL    come.PST.PTCP.3SGM    my son  
 ‘The one (M) who came to the house is my son.’
- (247) V-based in NP use, Telugu (South-Central Dravidian)  
 inʃiki            vaccinadi            ma: ku:tur  
 house.ALL    come.PST.PTCP.3SGF/N    my daughter  
 ‘The one (F) who came to the house is my daughter.’
- (248) V-based in NP use, Telugu (South-Central Dravidian)  
 inʃiki            vaccinadi            ma: baɳɖi  
 house.ALL    come.PST.PTCP.3SGF/N    my vehicle  
 ‘The one (N) which came to the house is my vehicle.’

Telugu V-based forms do not carry gender marking in Mod use.

- (249) V-based in Mod use, Telugu (South-Central Dravidian)  
 inʃiki            vaccina            koɖuku/ku:tur/    baɳɖi  
 house.ALL    come.PST.PTCP    son/daughter/vehicle  
 ‘the son/daughter who came to the house/ the vehicle which came to the house.’

The table below summarizes the patterns of gender marking in the Dravidian languages under study. As observed, all of these languages exhibit gender marking on structures such as numerals and demonstratives in NP use, with the exception of Malayalam. In Malayalam, genitives do not show gender marking in NP use, although adjectives and V-based forms do. Consequently, Malayalam deviates from the Shibatani’s (2023) hierarchy. Gender marking on numerals in Mod use is attested in six of the eight languages. Additionally, three languages show gender marking on demonstratives in Mod use. However, none of the Dravidian languages examined here display gender marking on genitives, adjectives, or V-based forms in Mod use.

Table 4 Gender marking in various Dravidian languages

| NP use    |                                      |               |                                       |           |         |
|-----------|--------------------------------------|---------------|---------------------------------------|-----------|---------|
|           | Numeral                              | Demonstrative | Genitive                              | Adjective | V-based |
| Arebase   | ✓                                    | ✓             | ✓                                     | ✓         | ✓       |
| Beary     | ✓                                    | ✓             | ✓                                     | ✓         | ✓       |
| Kannada   | ✓                                    | ✓             | ✓                                     | ✓         | ✓       |
| Koraga    | ✓                                    | ✓             | ✓                                     | ✓         | ✓       |
| Malayalam | ✓                                    | ✓             |                                       | ✓         | ✓       |
| Tamil     | ✓                                    | ✓             | ✓                                     | ✓         | ✓       |
| Tulu      | ✓                                    | ✓             | ✓                                     | ✓         | ✓       |
| Telugu    | ✓                                    | ✓             | ✓                                     | ✓         | ✓       |
| Mod use   |                                      |               |                                       |           |         |
|           | Numeral                              | Demonstrative | Genitive                              | Adjective | V-based |
| Arebase   | ✓                                    | ✓             |                                       |           |         |
| Beary     | ✓                                    |               |                                       |           |         |
| Kannada   | ✓                                    |               |                                       |           |         |
| Koraga    | ✓                                    | ✓             |                                       |           |         |
| Malayalam |                                      |               |                                       |           |         |
| Tamil     |                                      |               |                                       |           |         |
| Tulu      | ✓                                    | ✓             |                                       |           |         |
| Telugu    |                                      |               |                                       |           |         |
|           | [✓] means presence of gender marking |               | Blank means absence of gender marking |           |         |

### 3.2 Absence of gender marking on Malayalam genitives in NP use

In the previous section, an exception to Shibatani's hierarchy was identified in the nominalization patterns of Malayalam. Specifically, while adjectives and V-based forms in Malayalam carry gender marking in NP contexts, genitives do not. This absence of gender marking on genitives in NP use was confirmed by four native speakers from different regions of Kerala—Kannur, Malappuram, Thrissur, and Kollam. The consistency of this pattern across speakers from diverse dialectal backgrounds suggests that it is not attributable to dialectal variation. Furthermore, the use of ungendered genitives in Malayalam NP contexts is also reported in Asher and Kumari (1997).

- (250) Genitive in NP use, Malayalam (South Dravidian), (Asher and Kumari 1997: 175, e.g. no.852, literary gloss supplied), absence of gender marking

enre kayyilu||a paṇam enre alla  
I-GEN hand-LOC-be-RP money I-GEN NEG

‘The money I have with me is not mine.’ (Original gloss)

‘The money that I have in my hands is not mine.’ (Literary gloss supplied)

In the above Malayalam example, the genitive, *enre* I-GEN ‘my’, in Mod use (*enre kayyilu||a paṇam*/ I-GEN hand-LOC-be-RP money/ ‘the money that I have in my hands’/) is not marked by gender. Likewise, in the same example, the genitive, *enre* I-GEN ‘mine’, in NP use (*enre alla*/ I-GEN NEG/ ‘not mine’/) is not marked for any gender as well. In contrast, Tamil—the closest sister language of Malayalam—does exhibit gender marking on genitives,

adjectives, and V-based forms in NP use, as discussed in Section 3.1. An investigation into earlier stages of Malayalam is currently underway to determine whether gender marking on genitives in NP use was present historically.

### 3.3 Gender marking on demonstratives in Mod use

Section 3.1 presented the general tendencies of nominalization in several Dravidian languages. These languages consistently exhibit gender marking in NP use, with the exception of Malayalam. In Mod use, gender marking is typically restricted to numerals. Other categories—such as demonstratives, genitives, adjectives, and V-based forms—generally do not show gender marking in Mod use, as discussed in the previous section. However, our fieldwork on Arebase, Koraga, and Tulu revealed exceptions to this pattern: gender marking was observed on demonstratives in Mod use. Some representative examples of this phenomenon from Arebase, Koraga, and Tulu are presented below.

- (251) Demonstrative in Mod use, Arebase (South Dravidian), presence of gender marking on demonstrative

|                |       |           |
|----------------|-------|-----------|
| i: /ivn        | hayda | nannavan  |
| this/this.3SGM | boy   | mine.3SGM |

‘This/this (M) boy is mine.’

- (252) Demonstrative in Mod use, Koraga (South Dravidian), presence of gender marking on demonstrative

|                |       |      |         |
|----------------|-------|------|---------|
| a:/a:yi        | ku:je | anna | student |
| that/that.3SGM | boy   | my   | student |

‘That/that (M) boy is my student.’

- (253) Demonstrative in Mod use, Tulu (South Dravidian), presence of gender marking on demonstrative

|                |     |             |
|----------------|-----|-------------|
| a:/a:ye        | a:ŋ | husa:r      |
| that/that.3SGM | boy | intelligent |

‘That/that (M) boy is intelligent.’

This pattern of gender marking on demonstratives in Mod use was not observed in the preliminary investigations of Kannada, Malayalam, Tamil, and Telugu conducted prior to our fieldwork in 2023 and 2024. To confirm this, a renewed investigation was carried out in these four languages, specifically focusing on gender marking on demonstratives in Mod contexts. However, as shown below, gender marking on demonstratives in Mod use is not attested in Kannada, Malayalam, Tamil, or Telugu.



- (254) Gendered demonstrative in Mod use, Kannada (South Dravidian)

|           |         |       |        |
|-----------|---------|-------|--------|
| *avanu    | gaṇḍasu | nanna | geḷeya |
| that.3SGM | man     | my    | friend |

Cf.

|      |         |       |        |
|------|---------|-------|--------|
| a:   | gaṇḍasu | nanna | geḷeya |
| that | man     | my    | friend |

‘That man is my friend.’

- (255) Gendered demonstrative in Mod use, Malayalam (South Dravidian)

|           |           |         |             |      |
|-----------|-----------|---------|-------------|------|
| *avan     | manushyan | nammude | kuttaka:ran | a:ṇa |
| that.3SGM | man       | our     | friend      | be   |

Cf.

|      |           |         |             |      |
|------|-----------|---------|-------------|------|
| a:   | manushyan | nammude | kuttaka:ran | a:ṇa |
| That | man       | our     | friend      | be   |

‘That man is our friend.’

- (256) Gendered demonstrative in Mod use, Tamil (South Dravidian)

|           |        |          |        |
|-----------|--------|----------|--------|
| *avan     | payyan | ennodeya | nanban |
| that.3SGM | man    | my       | friend |

Cf.

|       |        |          |        |
|-------|--------|----------|--------|
| anada | payyan | ennodeya | nanban |
| that  | man    | my       | friend |

‘That man is my friend.’

- (257) Gendered demonstrative in Mod use, Telugu (South-Central Dravidian)

|           |         |     |           |
|-----------|---------|-----|-----------|
| *atanu    | abba:yi | na: | snehituḍu |
| that.3SGM | man     | my  | friend    |

Cf.

|      |         |     |           |
|------|---------|-----|-----------|
| a:   | abba:yi | na: | snehituḍu |
| that | man     | my  | friend    |

‘That man is my friend.’

Given the observed patterns of gendered and non-gendered demonstratives in Mod use, the question of whether gender marking has extended to demonstratives in Mod contexts across Dravidian languages warrants detailed investigation in future research.

### 3.4 Classifier marking in Dravidian languages: A preliminary account

Dravidian languages are generally not rich in classifiers and do not typically categorize nouns based on attributes such as shape, substance, or animacy. However, one classifier of

particular interest in the context of nominalization theory is the human classifier, attested in several Dravidian languages including Arebase, Beary, Kannada, Malayalam, Tamil, Tulu, and Telugu, as illustrated in the examples below. Notably, these classifiers do not distinguish between masculine, feminine, and neuter genders; instead, they primarily differentiate between human and non-human entities.

- (258) Arebase (South Dravidian), human classifier *jana*

NP use:            muuru        jana            bandiddo  
                       three        H.CLF        came.3PL  
                       ‘Three (people) came.’

Mod use:            muuru        jana            hengasaga        bandiddo  
                       three        H.CLF        women            came.3PL  
                       ‘Three (H.CLF) women came.’

- (259) Beary (South Dravidian), human classifier *a:l*

NP use:            jendɿ        a:l            banna:ra  
                       two        H.CLF        came.3PL  
                       ‘Two (people) came.’

Mod use:            jendɿ        a:l            tɿcarma:r        banna:ra  
                       two        H.CLF        teachers            came.3PL  
                       ‘Two (H.CLF) teachers came.’

- (260) Kannada (South Dravidian), human classifier *mandi/jana*

NP use:            hattu        jana/mandi        bandaru  
                       ten        H.CLF        came.3PL  
                       ‘Ten (people) came.’

Mod use:            hattu        jana/mandi        adhya:pakaru        bandaru  
                       ten        H.CLF        teachers            came.3PL  
                       ‘Ten (H.CLF) teachers came.’

- (261) Koraga (South Dravidian), human classifier *jana*

NP use:            na:lu        jana            barakeru  
                       four        H.CLF        came.3PL  
                       ‘Four (people) came.’

Mod use:            na:lu        jana            poŋŋulu            barakeru  
                       four        H.CLF        women            came  
                       ‘Four (H.CLF) women came.’

- (262) Malayalam (South Dravidian), human classifier *a:l/pe:r*
- |          |                      |          |           |       |
|----------|----------------------|----------|-----------|-------|
| NP use:  | pattu                | a:l/pe:r | vannu     |       |
|          | ten                  | H.CLF    | came      |       |
|          | ‘Ten (people) came.’ |          |           |       |
| Mod use: | *pattu               | a:l/pe:r | tɪcarma:r | vannu |
|          | ten                  | H.CLF    | teachers  | came  |
- (263) Tamil (South Dravidian), human classifier *pe:r*
- |          |                              |       |          |          |
|----------|------------------------------|-------|----------|----------|
| NP use:  | pattu                        | pe:r  | vanda:r  |          |
|          | ten                          | H.CLF | came.3PL |          |
|          | ‘Ten (people) came.’         |       |          |          |
| Mod use: | pattu                        | pe:r  | tɪcarkaɭ | vanda:r  |
|          | ten                          | H.CLF | teachers | came.3PL |
|          | ‘Ten (H.CLF) teachers came.’ |       |          |          |
- (264) Tulu (South Dravidian), human classifier *jana*
- |          |                                |       |          |          |
|----------|--------------------------------|-------|----------|----------|
| NP use:  | mu:ji                          | jana  | battera  |          |
|          | three                          | H.CLF | came.3PL |          |
|          | ‘Three (people) came.’         |       |          |          |
| Mod use: | muuji                          | jana  | tɪcars   | battera  |
|          | three                          | H.CLF | teachers | came.3PL |
|          | ‘Three (H.CLF) teachers came.’ |       |          |          |
- (265) Telugu (South Central Dravidian), human classifier *mandi*
- |          |                              |       |          |          |
|----------|------------------------------|-------|----------|----------|
| NP use:  | padi                         | mandi | vacca:ru |          |
|          | ten                          | H.CLF | came.3PL |          |
|          | ‘Ten (people) came.’         |       |          |          |
| Mod use: | padi                         | mandi | ticarlu  | vacca:ru |
|          | ten                          | H.CLF | teachers | came.3PL |
|          | ‘Ten (H.CLF) teachers came.’ |       |          |          |

It is worth noting that the classifiers *a:l* in Beary, *a:l* in Malayalam, *mandi* in Kannada and Telugu, *pe:r* in Tamil and Malayalam, and *jana* in Kannada, Koraga, and Tulu occur exclusively in the context of numerals. These classifiers are not found with other categories such as demonstratives, genitives, adjectives, or V-based forms. The following data from Kannada illustrate this pattern.

- (266) Kannada (South Dravidian), classifier with the demonstratives, ungrammatical

|      |            |                 |
|------|------------|-----------------|
| *a:  | mandi/jana | vidya:rthiga[u] |
| That | H.CLF      | students        |

Cf.

|                  |                 |
|------------------|-----------------|
| a:               | vidya:rthiga[u] |
| that             | students        |
| 'those students' |                 |

- (267) Kannada (South Dravidian), classifier with the genitives, ungrammatical

|        |            |                 |
|--------|------------|-----------------|
| *namma | mandi/jana | vidya:rthiga[u] |
| our    | H.CLF      | students        |

Cf.

|                |                 |
|----------------|-----------------|
| namma          | vidya:rthiga[u] |
| our            | students        |
| 'our students' |                 |

- (268) Kannada (South Dravidian), classifier with the adjectives, ungrammatical

|             |            |                 |
|-------------|------------|-----------------|
| *ja:ṇa      | mandi/jana | vidya:rthiga[u] |
| intelligent | H.CLF      | students        |

Cf.

|                        |                 |
|------------------------|-----------------|
| *ja:ṇa                 | vidya:rthiga[u] |
| intelligent            | students        |
| 'intelligent students' |                 |

- (269) Kannada (South Dravidian), classifier with the V-based forms, ungrammatical

|           |               |            |                 |
|-----------|---------------|------------|-----------------|
| *ninne    | banda         | mandi/jana | vidya:rthiga[u] |
| yesterday | come.PST.PTCP | H.CLF      | students        |

Cf.

|                                   |               |                 |
|-----------------------------------|---------------|-----------------|
| ninne                             | banda         | vidya:rthiga[u] |
| yesterday                         | come.PST.PTCP | students        |
| 'the students who came yesterday' |               |                 |

The distribution of these various human classifiers in different Dravidian languages is presented below. It may be noted that, in Malayalam, the classifiers *a:l* and *pe:r* are not attested in Mod use.

Table 5 Human classifiers in Dravidian languages

|           | mandi | Jana | a:l/a: | pe:r |
|-----------|-------|------|--------|------|
| Arebase   |       | ✓    |        |      |
| Beary     |       |      | ✓      |      |
| Kannada   | ✓     | ✓    |        |      |
| Koraga    |       | ✓    |        |      |
| Malayalam |       |      | ✓      | ✓    |
| Tamil     |       |      |        | ✓    |
| Tulu      |       | ✓    |        |      |
| Telugu    | ✓     |      |        |      |

Some observations on the diachrony of these classifiers may be made considering their distribution. First, no classifier is attested across all these Dravidian languages. Second, there is no sound correlation among the classifiers *a:l*, *jana*, *pe:r*, and *mandi*. These two issues indicate that the development of these human classifiers in these languages is not shared and took place independently in the course of time. The classifiers *a:l*, *mandi*, and *pe:r* are native Dravidian items while the classifier *jana* is a borrowing from the Indo-Aryan languages. In Kannada, the classifier *jana* tend to interchange with *mandi*. Both classifiers, *mandi* and *jana*, co-exist in Modern Kannada and the exact interplay between them is not clear at present. However, considering that *mandi* is of Dravidian origin and *jana* is borrowed from the IA stock, there is scope to assume that, diachronically, *mandi* first developed as a classifier in Kannada and *jana* subsequently started to interchange with *mandi* as a human classifier.

### 3.5 Diachrony of gender marking: Kannada numerals in modification use

It may be observed that Kannada carries gender marking on numerals in Mod use. However, Kannada, diachronically, did not have gender marking on numerals in Mod use, as shown in the examples below.

- (270) Old Kannada (South Dravidian), 7<sup>th</sup> C AD, no gender marking on the numeral ‘one’ in Mod use, human referent, (Epigraphia Carnatica, 1973: 78, inscription number 133)

...oru      muniyimda|      tapaccale      nontu...  
 ...one      sage.CASE      firm in penance      having observed  
 ‘...having observed the vow in firm penance like one/a sage...’ (Gloss uncertain as the case on *muni* ‘sage’ is not clear.)

- (271) Old Kannada, no gender marking on the numeral ‘one’ in Mod use, non-human referent, (Narayana 2007: 105)

or   pada      or   ko:l      or   cara      or   nuḍi  
 one step      one stick      one chain      one word  
 ‘a step’      ‘a stick’      ‘a chain’      ‘a word’

In (270), in the sequence *oru muni* ‘one sage’, there is no gender marking on *oru* ‘one’. Likewise, in (271), the numeral *or* ‘one’ does not carry any gender. The above examples of Old

Kannada in (270) and (271) are similar to Malayalam and Tamil patterns wherein there is no gender marking on the numeral ‘one’ in Mod use as may be seen below. In the below Malayalam and Tamil examples, the numeral *oru* ‘one’ is not marked for any gender.

- (272) Malayalam, no gender marking on the numeral ‘one’, human and non-human referents

|               |            |             |
|---------------|------------|-------------|
| oru manushyan | oru sthree | oru vanḍi   |
| one man       | one woman  | one vehicle |
| ‘a man’       | ‘a woman’  | ‘a vehicle’ |

- (273) Tamil, no gender marking on the numeral ‘one’, human and non-human referents

|            |           |             |
|------------|-----------|-------------|
| oru payyan | oru poṇḷu | oru vanḍi   |
| one man    | one woman | one vehicle |
| ‘a man’    | ‘a woman’ | ‘a vehicle’ |

Even in Old Kannada, attestations like (270) and (271) are very rare to come across. In Old Kannada as well, gender/classifier marking of numerals is attested in Mod use as may be seen below.

- (274) Old Kannada (South Dravidian), 8<sup>th</sup> C AD, gendered numeral in Mod use, human referent, *sa:sirvar* ‘thousand.CLF’, (Epigraphia Carnatica, 1974: 54, inscription number 67)

|                                                                                                                                  |              |           |             |
|----------------------------------------------------------------------------------------------------------------------------------|--------------|-----------|-------------|
| ...inda:na itto:n                                                                                                                | sa:sirvar    | pa:rvaru  | sa:sira     |
| ...it.destroy.PTCP.3SGM                                                                                                          | thousand.CLF | Brahmins  | thousand    |
| kavile                                                                                                                           | sasirvar     | rishiyaru | ko:pamekkdu |
| tawny cow                                                                                                                        | thousand.CLF | sages     | sin.attain  |
| ‘The one who destroys it will attain the sin of killing one thousand Brahmins, one thousand tawny cows, and one thousand sages.’ |              |           |             |

- (275) Old Kannada (South Dravidian), 7<sup>th</sup> C AD, gendered numeral in Mod use, non-human referent, *ondu* ‘one’ is the citation form, (Epigraphia Carnatica, 1973: 11-12, inscription number 28)

|                                                                                                                                          |               |            |                  |
|------------------------------------------------------------------------------------------------------------------------------------------|---------------|------------|------------------|
| ...ugrase:na                                                                                                                             | guvadiga      | ondu tinga | sanyasanam       |
| ...Ugrasena                                                                                                                              | Guruvadigal   | one month  | renunciation.ACC |
| no:ntu                                                                                                                                   | muḍipidda:r   |            |                  |
| having observed                                                                                                                          | caused to end |            |                  |
| ‘Ugrasena-guruvadigal, having observed the vow of <i>sanyasa</i> (complete renunciation of worldly ties) for one month, ended his life.’ |               |            |                  |

Further, in Old Kannada as well, the gendered numerals in Mod use outnumber the non-gendered numerals. But, the Old Kannada examples in (270) and (271) above make it clear that gender marking on numerals in Mod use is a historical development. This is further

substantiated by the variation in the forms of numeral ‘one’ attested in various Dravidian languages.

Table 6 Gender marking on the numeral ‘one’ in NP use

|           | Masculine | Feminine         | Neuter             |
|-----------|-----------|------------------|--------------------|
| Arebase   | obba      | onda (bare form) | onda (bare form)   |
| Beary     | orta      | orti             | onnu               |
| Kannada   | obba      | obba <u>u</u>    | ondu (bare form)   |
| Koraga    | ori       | orti             | onni (bare form)   |
| Malayalam | oruvan    | oruva <u>l</u>   | onnu (bare form)   |
| Tamil     | oruvan    | oruva <u>l</u>   | onnu (bare form)   |
| Tulu      | ori       | orti             | on̄ji (bare form)  |
| Telugu    | okkaḍu    | okkarti          | okkaṭi (bare form) |

Table 7 Gender marking on the numeral ‘one’ in Mod use

|           | Masculine | Feminine         | Neuter            |
|-----------|-----------|------------------|-------------------|
| Arebase   | obba      | onda (bare form) | onda (bare form)  |
| Beary     | oru       | oru              | oru               |
| Kannada   | obba      | obba <u>u</u>    | ondu (bare form)  |
| Koraga    | ori       | onni (bare form) | onni (bare form)  |
| Malayalam | oru       | oru              | oru               |
| Tamil     | oru       | oru              | oru               |
| Tulu      | ori       | orti             | on̄ji (bare form) |
| Telugu    | okka      | okka             | okka              |

The diachronic development of Kannada numerals from the older Dravidian form *or(u)* is presented below.

- (276) The diachrony of Modern Kannada bare numeral *ondu* ‘one’  
 oru > \*onru (euphonic nunnation) > \*ondru (insertion of /d/) > ondu (loss of medial /r/)
- (277) The diachronic development of Modern Kannada masculine *obbanu* ‘one.M’  
 oru > orvanu ~ orva (-an is M suffix, /v/ is addition, haplology of /nu/) > orbanu (fortion of /v/) > obbanu ~ obba (regressive assimilation of /r/, haplology of /nu/)
- (278) The diachronic development of Modern Kannada feminine *obbau* ‘one.F’  
 oru > orvau (-al is F suffix, /v/ is addition) > orbau (fortition of /v/ > obbau ~ obba (regressive assimilation of /r/, haplology of /lu/)

### 3.6 Absence of gender marking: The Telugu and Tamil patterns

Telugu presents a situation wherein the numeral ‘one’ does not exhibit any gender marking in Mod use while other higher numerals such as ‘two’ and ‘three’ present classifier markings as may be seen below.

- (279) Telugu (South-Central Dravidian), no gender marking on the numeral ‘one’

|              |              |           |
|--------------|--------------|-----------|
| okka abba:yi | okka amma:yi | okka illu |
| one man      | one woman    | one house |
| ‘a man’      | ‘a woman’    | ‘a house’ |

- (280) Telugu (South-Central Dravidian), classifier marking on higher numerals

|           |           |             |          |              |        |
|-----------|-----------|-------------|----------|--------------|--------|
| iddaru    | abba:yilu | iddaru      | ammayilu | raṇḍu        | illulu |
| two.CLF   | men       | two.CLF     | women    | two          | houses |
| ‘two men’ |           | ‘two women’ |          | ‘two houses’ |        |

Tamil presents a similar case wherein the numeral *oru* ‘one’ does not carry any gender marking in Mod use, but the language makes use of the human classifier *pe:r* with higher numerals as may be seen below.

- (281) Tamil (South Dravidian), no gender marking on the numeral ‘one’

|            |           |             |
|------------|-----------|-------------|
| oru payyan | oru poṇṇu | oru vaṇḍi   |
| one man    | one woman | one vehicle |
| ‘a man’    | ‘a woman’ | ‘a vehicle’ |

- (282) Tamil (South Dravidian), classifier marking on the higher numerals

|                |       |              |                |
|----------------|-------|--------------|----------------|
| pattu          | pe:r  | aysiriyarkaḷ | pattu vaṇḍikaḷ |
| ten            | H.CLF | teachers     | ten vehicles   |
| ‘ten teachers’ |       |              | ‘ten vehicles’ |

In the case of Telugu, the numeral *okka* ‘one’ does not bear any gender marking while higher numerals such as *iddaru* ‘two.H.CLF’, *mugguru* ‘three.H.CLF’, *nalguru* ‘four.H.CLF’ etc are marked for human reference. Previously, it was observed, in (265), that Telugu uses the human classifier *mandi* as well. Therefore, what may serve as an explanation for the absence of gender marking on the Telugu numeral *okka* in Mod use? Similarly, in Tamil, the numeral *oru* ‘one’ in Mod use is not marked for gender, whereas higher numerals may optionally take the human classifier *pe:r*. This raises the question: why does the Tamil numeral *oru* ‘one’ lack gender marking, even though the classifier *pe:r* is optionally used with higher numerals? The historical erosion of gender marking on the Tamil numeral *oru* ‘one’ and the Telugu numeral *okka* ‘one’ appears to be a plausible explanation. At present, an investigation is underway to identify historical attestations of gender marking on the numeral ‘one’ in both Tamil and Telugu.



#### 4. ANALYSIS AND DISCUSSION

The patterns of gender marking observed across the Dravidian languages under study allow us to draw several generalizations. As evident from the data, Dravidian languages typically mark gender on numerals, demonstratives, genitives, adjectives, and V-based forms in NP use. Malayalam stands out as an exception to this overall tendency. In Mod contexts, many Dravidian languages exhibit gender marking on numerals, while only a few languages display gender marking on demonstratives. Notably, no Dravidian language shows gender marking on genitives, adjectives, or V-based forms in Mod use. Figure 1 summarizes the patterns of gender marking in these Dravidian languages.

Figure 1. Patterns of gender marking in the Dravidian languages under study, displayed on Shibatani's hierarchy

##### Arebase (South Dravidian)

|          |      |                     |       |       |               |
|----------|------|---------------------|-------|-------|---------------|
| NP use:  | NUM> | DEM>                | GEN>  | ADJ>  | V-based NMLZ  |
| Mod use: | NUM> | <sup>OPT</sup> DEM> | *GEN> | *ADJ> | *V-based NMLZ |

##### Beary (South Dravidian)

|          |                     |       |       |       |               |
|----------|---------------------|-------|-------|-------|---------------|
| NP use:  | NUM>                | DEM>  | GEN>  | ADJ>  | V-based NMLZ  |
| Mod use: | <sup>OPT</sup> NUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

##### Kannada (South Dravidian)

|          |      |       |       |       |               |
|----------|------|-------|-------|-------|---------------|
| NP use:  | NUM> | DEM>  | GEN>  | ADJ>  | V-based NMLZ  |
| Mod use: | NUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

##### Koraga (South Dravidian)

|          |      |                     |       |       |               |
|----------|------|---------------------|-------|-------|---------------|
| NP use:  | NUM> | DEM>                | GEN>  | ADJ>  | V-based NMLZ  |
| Mod use: | NUM> | <sup>OPT</sup> DEM> | *GEN> | *ADJ> | *V-based NMLZ |

##### Malayalam (South Dravidian)

|          |       |       |       |       |               |
|----------|-------|-------|-------|-------|---------------|
| NP use:  | NUM>  | DEM>  | *GEN> | ADJ>  | V-based NMLZ  |
| Mod use: | *NUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

##### Tamil (South Dravidian)

|          |       |       |       |       |               |
|----------|-------|-------|-------|-------|---------------|
| NP use:  | NUM>  | DEM>  | GEN>  | ADJ>  | V-based NMLZ  |
| Mod use: | *NUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

##### Tulu (South Dravidian)

|          |      |                     |       |       |               |
|----------|------|---------------------|-------|-------|---------------|
| NP use:  | NUM> | DEM>                | GEN>  | ADJ>  | V-based NMLZ  |
| Mod use: | NUM> | <sup>OPT</sup> DEM> | *GEN> | *ADJ> | *V-based NMLZ |

#### Telugu (South Central Dravidian)

|          |       |       |       |       |               |
|----------|-------|-------|-------|-------|---------------|
| NP use:  | NUM>  | DEM>  | GEN>  | ADJ>  | V-based NMLZ  |
| Mod use: | *NUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

unmarked: obligatorily marked for gender

OPT: optional

\*: impossible

Further, generalizations can also be made regarding classifier marking in Dravidian languages. Figure 2 shows the patterns. Currently, little is known about the use of classifiers with demonstratives, genitives, adjectives, and V-based forms in NP contexts, so these are not included in Figure 2 for NP use. In Mod contexts, classifier marking appears to be exclusively associated with numerals, with no evidence of its use on other categories such as demonstratives and genitives. Besides, while classifiers routinely occur with numerals in both NP and Mod uses in these Dravidian languages, given the current understanding, the occurrence of classifiers in these instances is characterized as optional. Further research is needed to gain a more comprehensive understanding of classifier marking across the Dravidian language family.

Figure 2. Patterns of classifier marking in the Dravidian languages under study, displayed on Shibatani's hierarchy

#### Arebase (South Dravidian)

|          |         |       |       |       |               |
|----------|---------|-------|-------|-------|---------------|
| NP use:  | OPTNUM  |       |       |       |               |
| Mod use: | OPTNUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

#### Beary (South Dravidian)

|          |         |       |       |       |               |
|----------|---------|-------|-------|-------|---------------|
| NP use:  | OPTNUM  |       |       |       |               |
| Mod use: | OPTNUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

#### Kannada (South Dravidian)

|          |         |       |       |       |               |
|----------|---------|-------|-------|-------|---------------|
| NP use:  | OPTNUM  |       |       |       |               |
| Mod use: | OPTNUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

#### Koraga (South Dravidian)

|          |         |       |       |       |               |
|----------|---------|-------|-------|-------|---------------|
| NP use:  | OPTNUM  |       |       |       |               |
| Mod use: | OPTNUM> | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

#### Malayalam (South Dravidian)

|          |        |       |       |       |               |
|----------|--------|-------|-------|-------|---------------|
| NP use:  | OPTNUM |       |       |       |               |
| Mod use: | *NUM>  | *DEM> | *GEN> | *ADJ> | *V-based NMLZ |

Tamil (South Dravidian)

NP use: <sup>OPT</sup>NUM

Mod use: <sup>OPT</sup>NUM>      \*DEM>      \*GEN>      \*ADJ>      \*V-based NMLZ

Tulu (South Dravidian)

NP use: <sup>OPT</sup>NUM

Mod use: <sup>OPT</sup>NUM>      \*DEM>      \*GEN>      \*ADJ>      \*V-based NMLZ

Telugu (South Central Dravidian)

NP use: <sup>OPT</sup>NUM

Mod use: <sup>OPT</sup>NUM>      \*DEM>      \*GEN>      \*ADJ>      \*V-based NMLZ

\*:                      impossible

OPT:                  optional

unmarked:           obligatorily marked for human classifier

## 5. CONCLUSION

The patterns of gender and classifier marking observed in the Dravidian languages examined in this study largely provide support for Shibatani's (2023) hierarchy. In NP contexts, these languages generally exhibit gender marking on numerals, demonstratives, genitives, adjectives, and V-based forms. An exception is found in Malayalam, which does not mark gender on genitives in NP use, although adjectives and V-based forms do show gender marking. In Mod contexts, many Dravidian languages mark gender on numerals, and a subset also marks gender on demonstratives. However, none of the languages studied thus far display gender marking on genitives, adjectives, or V-based forms in Mod use. Classifier marking remains a less understood aspect of Dravidian grammar, though current evidence suggests that, at least in the case of numerals, classifiers are used in both NP and Mod contexts. These findings highlight several promising avenues for future research. In particular, gender and classifier marking in the Central and North Dravidian languages remain largely unexplored. Additionally, historical records may help address unresolved questions, such as the absence of gender marking on Malayalam genitives in NP use and the ungendered form of the numeral 'one' in Tamil and Telugu.

## ABBREVIATIONS

|     |               |
|-----|---------------|
| 1   | first person  |
| 2   | second person |
| 3   | third person  |
| ACC | accusative    |
| ALL | allative      |
| CLF | classifier    |
| DAT | dative        |
| DEM | demonstrative |
| F   | feminine      |

|       |                     |
|-------|---------------------|
| GEN   | genitive            |
| H     | human               |
| HON   | honorific           |
| M     | masculine           |
| LOC   | locative            |
| N     | neuter              |
| NEG   | negative            |
| NH    | non-human           |
| NMLZ  | nominalization      |
| NMLZR | nominalizer         |
| NOM   | nominative          |
| NUM   | numeral             |
| PL    | plural              |
| PP    | post position       |
| PRE   | present             |
| PST   | past tense          |
| PTCP  | participle          |
| RP    | relative participle |
| SG    | singular            |

## ACKNOWLEDGMENTS

I thank the National Institute for Japanese Language and Linguistics (Project: ‘Empirical study of the typology of nominalization—from theoretical, fieldwork, historical and dialectal perspective’, PI: Prashant Pardeshi) for the financial assistance in carrying out this research. I am grateful to Masayoshi Shibatani and Prashant Pardeshi for their invaluable support and guidance in carrying out this research. Many thanks are due to the entire editorial team of this volume including Kazuhiro Kawachi and Maho Naito for their kind support and assistance in finalizing this paper. I thank Arun Shyam for his comments and observations on the Beary and Malayalam data. My heartfelt gratitude goes to all the language consultants who generously contributed their time and knowledge: Arebase (Kaushik P. K., Niveditha K. K., Prapti Gowda, and Shravik N. Adka), Beary (Abubakkar Siddiq ~ Abbokar Siddiq, Abdul Rahiman Kuthethoor, Abdul Razak Ananthady, A Haidarali, and Shahla Abdul Rahman), Kannada (Basavaraja Kodagunit and Naveen Kumar D. V.), Koraga (Kudupa, Pangala Babu Koraga, Ramesh Manchakal (Makkalamane), and Shyam Koraga), Malayalam (Muneer Babu M., Niveditha K., Sanjoy C. K., Shaji M. L., and Shukkoor T.), Tamil (Seenuvasan Vedachalam and Thiru Selvan), Telugu (Bapuji Mendem, P. Karuna Purnapu Rupa, and Swathi Chintala), and Tulu (Bharati Chandrashekar Kunder, Chandrashekar Kunder, K. Gururaj Bhat, Mahesh Moolya, Padmanabha Kekunnaya, Sayeegeetha, Shakunthala, Shivani, Shrikara Shetty, and Vishwanatha Badikana). I thank the following centers and persons for their generous support in conducting various fieldworks: R. G. Pai Research Centre (Udupi), Beary Adhyayana Peeta, University College (Mangaluru), and Dr. Sayeegeetha of Dr. K. R. Shetty Centre for Tulu

Studies, NITTE (Deemed to be) University (Mangaluru). All errors and shortcomings are my own.

## REFERENCES

- Asher, R E and T C Kumari. 1997. *Malayalam*. Routledge.
- Gowda, Kushalappa K. 2019. *Arebase (Gowda Kannada) Vyakarana*. Karnataka Arebbase Samskrithi Mattu Sahitya Academy.  
(<https://archive.org/details/arebasegowdakann0000prof>)
- Epigraphia Carnatica*, Vol 2. 1973. University of Mysore/Universitat Tubingen.  
([https://opendigi.ub.uni-tuebingen.de/opendigi/EC\\_02\\_1973#p=2](https://opendigi.ub.uni-tuebingen.de/opendigi/EC_02_1973#p=2))
- Epigraphia Carnatica*, Vol 3. 1974. University of Mysore/Universitat Tubingen.  
([https://opendigi.ub.uni-tuebingen.de/opendigi/EC\\_03\\_1974#p=1](https://opendigi.ub.uni-tuebingen.de/opendigi/EC_03_1974#p=1))
- Krishnamurthi, Bh. 2003. *The Dravidian languages*. Cambridge University Press.
- Narayana, P V. 2007. *Halagannada Padasampada*. Kannada Sahitya Parishat.
- Pardeshi, Prashant, Gitanjali Bez, Niranjana Uppoor and Masayoshi Shibatani. 2023. Gender and classifier marking in South Asian languages: A nominalization perspective. Online talk given at the SALA-37 during October 4-7, 2023.
- Shibatani, Masayoshi. 2023. Towards dynamic functional typology: Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1–52.  
([https://www.jstage.jst.go.jp/article/hlj/12/0/12\\_1\\_3/pdf/-char/en](https://www.jstage.jst.go.jp/article/hlj/12/0/12_1_3/pdf/-char/en))
- Uppoor, Niranjana. 2022. Gender marking in Kannada and other Dravidian languages: A nominalization perspective. Talk given at the University of Tokyo on September 9, 2022.
- Uppoor, Niranjana. 2024. Nominalization hierarchy and gender marking: Evidence from minor Dravidian languages. Online talk given at the NINJAL project meeting on September 14, 2024.
- Uppoor, Niranjana. 2025. *Gender marking in the Dravidian languages: A nominalization perspective*. Talk given at the 52nd All India Conference of Dravidian Linguists during June 19-21, 2025.

『慶應義塾 外国語教育研究』編集委員

|      |           |      |     |
|------|-----------|------|-----|
| 経済学部 | 吉川龍生（委員長） | 事務局  | 八代誠 |
|      | 前島和也      | 加藤祐一 | 子緒  |
| 法学部  | 大野真澄      | 杉田本  | 美緒  |
|      | 永嶋宏亮      | 松木亜  | 月   |
| 理工学部 | 山下一夫      |      |     |
| 高等学校 | 大竹優志      |      |     |

慶應義塾 外国語教育研究 別冊（論集）第1号

2026年3月31日 発行

発行人 山下一夫

印刷所 有限会社 梅沢印刷所

発行所 慶應義塾大学外国語教育研究センター

神奈川県横浜市港北区日吉4-1-1

電話 045-566-1031