

From classifiers to subordination: Nominal origins of relativizers and subordinators in Tù'un na Ñuu Sá Mátxíí Ntxè'è (Mixtec)

Albert VENTAYOL-BOADA

University of California, Santa Barbara – U.S.A.

Abstract: Grammaticalization patterns of nominal classification devices suggest that class terms tend to evolve into classifier systems, which may develop into gender and noun class systems. Additionally, classifiers are involved in the grammaticalization of other morphosyntactic categories, including applicatives, benefactives, and tense. This paper provides evidence for noun classifiers as the source of relative pronouns and subordinators in *Tù'un na Ñuu Sá Mátxíí Ntxè'è* (San Martín Durazos Mixtec, Otomanguan) and identifies a grammaticalization pathway from different generic nouns in the language to noun classifiers, thence to third person pronouns, relative pronouns, and finally subordinators. Each step on the grammaticalization cline is functionally motivated and is consistent with cross-linguistic patterns of language change, and thus this study provides new insight into the sources of relativizers.

Keywords: classifiers, grammaticalization, Mixtec, relative clauses, subordination

1. Introduction

In linguistic research, analyzing a phonological form with more than one function in a given language is a common endeavor. Epps (2008) summarizes very well the questions that the linguist is faced with at first: is the resemblance a historical accident of two or more independent forms converging into the same phonological shape? Is it a case of polysemy, that is, a single form with meanings vaguely related? Or is it perhaps the case that synchronically their functions have diverged so much that there is no longer a discernible semantic relation even though the forms are historically related? In addition, segmental and suprasegmental variation in the forms attested can make unravelling the tangle even more challenging: do differences in tone reflect distinct origins, context-dependent and thus predictable phonological alternations, or even obscured morphological material? Semantic and typological evidence can aid the analysis; when the

semantic link among the forms attested is not obvious, however, intermediate realizations can bridge and help uncover the grammatical connection. Ultimately, examining the language in its own terms can shed light into its historical development and assist the field linguist in the description process.

This paper presents an example of such an analytical challenge from *Tù'un na Ñuu Sá Mátxíí Ntxè'è*, the Mixtec variety spoken in San Martín Durazos (Oaxaca, Mexico), and it highlights the importance of considering diachronic change for language description. In this Mixtec variety, the form *ñà* is attested in a variety of morphosyntactic environments with distinct functions. First, it is attested as a dependent pronoun (1), and in combination with a demonstrative as an independent pronoun (2). Example (1) already displays some of the variation attested for this form: *àn* is a segmentally reduced form, whereas the variation in *ña* is suprasegmental and the form appears with a mid-tone. Both forms are allomorphs of *ñà* and are coreferential: they track the same referent. All the examples are given in the practical orthography¹.

- (1) *távà na kua'nu kí-àn chii xàà kúnì-ì kaxi-i-ña*
 távà na kua'nu kî=àn; chii xàà kúnì=ì
 CONJ MOD POT.grow soon=3GNR; because already IPFV.want=1SG
 kaxi=ì=ña;
 POT.eat=1SG=3GNR;
 'so that it grows right away because I want to eat it already' (SMD-0009-Jardín)
- (2) *ñàkán kíkí ntxìkuíí ña kúníñú'-àn*
 ñà=kán kíkí ntxìkuíí ña kúníñú'ú=àn
 3GNR=DEM.INV IPFV.COP all REL.GNR POT.need=3SG.F
 'that is everything that she will need' (SMD-0005-ArrozAmarillo)

¹ The orthographic conventions are in line with Guadalupe Joaquina's (2014) proposal and the Ndusu Tu'un Savi alphabet chart with the following modifications to represent the sibilants in the language: *sh* /ʃ/, *ch* /tʃ/, *x* /s/, *tx* /ts/. The glottal stop is represented with the saltillo (') and hyphens mark the boundaries between classifiers and dependent person forms, and their hosts. As for the abbreviations used in interlinear glosses, they follow the Leipzig Glossing Rules, with the addition of ADD 'additive', ANML 'animal', CONJ 'conjunction', COS 'change of state', GNR 'generic', H 'human', LNK 'linker', LQD 'liquid', MED 'medial', MOD 'modal', POT 'potential mood', RND 'round', TOPZ 'topicalizer', TREE 'tree/wood'. The glosses of the grams analyzed in this paper reflect their function in the morphosyntactic environment in which they appear.

In addition, the form *ñà* is attested as a relativizer in both headed (3) and headless (4) relative clause constructions, and as a subordinator in complement (5) and adverbial clauses (6).

- (3) *nì ìtu'un shina-ì ntxikuíí yìkì ñà xàà vèè íchì*
nì ìtu'un shina=ì ntxì yìkì ñà xàà vèè íchì
 MOD PFV.grab first=1SG all plant REL.GNR already heavy IPFV.dry
 'but first I grabbed all the herbs that were already drying' (SMD-0009-Jardín)
- (4) *ta ñà sá'a-na kíkí ña,*
ta ñà sá'a=na kíkí ña
 CONJ REL.GNR IPFV.do=3PL IPFV.COP SUB
 'and what they do is that'
ntíkí-na na-múxikà na síká'a.
ntíkí=na na=múxikà na síká'a
 IPFV.search=3PL CLF.H.PL=music REL.H.PL POT.play
 'they look for a band that will play' (SMD-0004-FiestaPatronal)
- (5) *siyàà-nà-rí xàà xínì-nà ña kuè'è-rí*
siyàà=nà=rí xàà xínì=nà ña kuè'è=rí
 let.out=3PL=3ANML already IPFV.know=3PL SUB IPFV.be.aggressive=3ANML
 'they let it out, they know that it's aggressive' (SMD-0007-AccidenteToby)
- (6) *na sá'-ó lúchá ñà kan-ó chínùùn*
na sá'a=ó lúchá ñà kana=ó Chínùùn
 MOD IPFV.do=1PL.EXCL fight SUB POT.leave=1PL.EXCL Ahead
 'we fight to come out ahead' (SMD-0011-MiHistoria)

Finally, the form *ñà/ña* can function as a general linker between nouns (7). In this context an adnominal modification reading is possible.

- (7) *ntiká'nu,*
ntiká'nu
 IPFV.enlarge
ña kíkí to'ó ña ñuu,
ña kíkí to'ó ña ñuu
 REL.GNR IPFV.COP saint LNK village
 'the saint in the village gets aggrandized' (lit. 'what it is the saint of the village gets aggrandized') (SMD-0004-FiestaPatronal)

Hollenbach (1995a) already noted the overlap in a single gram of the relativizer and subordinator in complement clauses, along with what she labelled a nominalizer. She describes this overlap not only in several Mixtec varieties, but also in Trique and Cuicatec, thus comprising the

larger Mixtecan language group to which Mixtec belongs (see Campbell 2017a for subgroupings in Otomanguean). Hollenbach (1995a) argues that these three functions are all historically related and proposes that the nominalizer and complementizer functions developed from the relativizer. Her account, however, does not take into consideration the structural resemblance with dependent and independent pronouns. In addition, the development of pronominal forms from noun classifiers is well attested in Mesoamerica (Craig 1986, de León 1988). In this paper I argue that the functions of relativizer and subordinator have also developed from the noun classifier system through a series of steps that are functionally motivated and consistent with cross-linguistic patterns of language change, thereby enriching our understanding of pathways of change from noun classifiers to more grammatical elements.

This paper is organized as follows. Section § 2 includes a discussion of noun classifier systems and their proposed grammaticalization from generic nouns, with a special focus on Mesoamerica. In Section § 3 I offer a description of all the forms in *Tù'un na Ñuu Sá Mátxíí Ntxè'è* that display overlapping functions as presented above. The analysis is based on monologic narratives collected as part of the documentation of the language. Section § 4 proposes a grammaticalization pathway from different generic nouns in the language to noun classifiers, thence to third person pronouns, to relative pronouns and subordinators in complement and adverbial clauses, supported by typological evidence. Finally, I draw some conclusions in Section § 5.

2. Noun classifiers

The study of noun categorization has a long tradition of research in the field of typology (Allan 1977, Dixon 1982, 1986, Craig 1986, Mithun 1986, Derbyshire & Payne 1990, Aikhenvald 2000, Grinevald 2000, Senft 2007, Krasnoukhova 2012, Contini-Morava & Kilarski 2013). Grinevald (2000) proposes a continuum of nominal classification devices that runs from purely lexical to purely grammatical, with measure and class terms on the lexical end and gender and noun class systems on the grammatical end. Measure terms comprise expressions that identify units of mass nouns (*e.g.*, a head of cattle, a pound of beans) or specify arrangements or

amounts of count nouns (*e.g.*, a pile of books), while class terms create taxonomies in a given semantic domain and are of lexical origin (*e.g.*, –berry in strawberry, blueberry). At the grammatical end, gender and noun class systems categorize nouns in a language by obligatorily dividing nouns into rigid subclasses, which may participate in agreement patterns. Classifier systems lie halfway along this lexico-grammatical continuum and can be characterized by their “incomplete grammaticalization” (Grinevald 2000: 61).

This proposal of nominal classification devices along a lexico-grammatical continuum allows Grinevald (2002) to generalize over attested grammaticalization patterns: class terms evolving into a classifier system, as attested in several languages of South East Asia (Bisang 1993; DeLancey 1986), and classifiers developing into a gender and noun class system, for example the Ngan’gityemerri gender system (Southern Daly, Australia; Reid 1997) and the noun classes of Bantu languages (Contini-Morava 1994, Grinevald & Seifart 2004, Meissner & Storch 2000). The grammaticalization pathway proposed by Grinevald (2002) is summarized below in Figure 1.

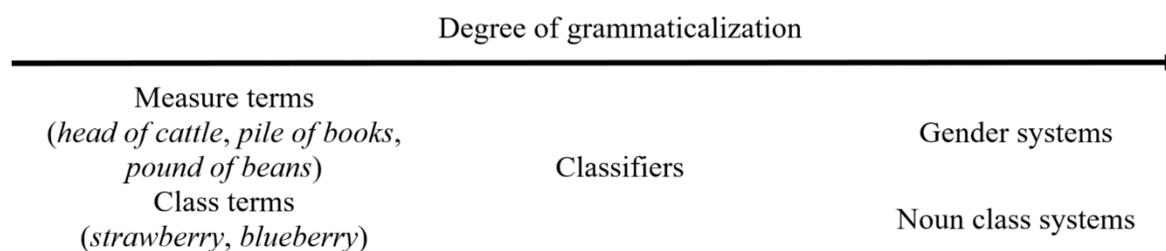


Figure 1: Lexico-grammatical continuum of nominal classification (adapted from Grinevald 2002)

This proposed cline on the lexico-grammatical continuum captures the idiosyncrasies of classifiers and shows their fluidity cross-linguistically: from full productivity to remnants of a system that is fossilized, depending on their position on the cline. Like any grammaticalization pathway, Grinevald’s proposal does not preclude the existence of mixed systems, in which forms on different stages in the grammaticalization continuum may,

and often do, coexist in a language. Concurrent systems that combine gender and classifiers, for example, have been identified in several languages of South America (Fedden & Corbett 2017, 2018).

Aikhenvald (2000) and Grinevald (2000) propose a taxonomy of classifiers into different types based on the morphosyntactic environments in which they occur: numeral, noun, verbal, possessive, and deictic classifiers. Aikhenvald (2000) notes, however, that in some languages the same set of classifiers can be used in several of the morphosyntactic environments listed above; she labels these as ‘multiple classifier systems’ and argues that classifiers may have different grammatical properties in each environment. For the remainder of this section, I will focus on noun classifiers, which have been described for several Mesoamerican languages, including the *Q’anjob’alan* branch of the Mayan language family (see Craig 1977, 1979, 1986 and Grinevald 2000, 2002 for Jakaltek Popti’, and Zavala 2000 for Akatek), some Mamean languages of the same family (see England 1983), and several Eastern Otomanguean languages (see Suárez 1983 for an overview; de León 1988, Small 1990, Macaulay 1996 and Mantenuto 2020 for different Mixtec varieties; Antonio Ramos 2013 for San Pedro Mixtepec Zapotec; Veerman-Leichsenring 2004 for the larger Popolocan group and Costaouec & Swanton 2015 for Ixcatec in particular). Hopkins (2012) argues that noun classifiers in Mesoamerica are a clear example of grammatical borrowing and were borrowed into Mayan languages that were in intense contact with speakers of Otomanguean varieties. Outside Mesoamerica, similar systems have been described for languages of South America (Epps 2007, Martins & Martins 1999, Valenzuela 2016, Vidal 1997), and Australia (Dixon 2002, Sands 1995, Wilkins 2000).

Grinevald (2000: 64) characterizes noun classifiers as “free morphemes standing in a noun phrase, next to the noun itself or within the boundaries of the noun phrase with other determiners of the noun” and independent to other constituents inside of it. It is unclear, however, why noun classifiers are defined specifically as “free morphemes” given that she considers classifiers as a whole to work either as free or bound forms. Aikhenvald (2000: 81), on the other hand, defines noun classifier systems as “a type of non-agreeing noun categorization device” that is determined by lexical

selection, and she proposes the following five core properties: i) they are chosen based on semantics and thus not every noun in a language necessarily takes one, ii) the same noun can take different classifiers with changes in meaning, iii) co-occurrence of more than one classifier within a noun phrase is allowed, iv) noun classifier inventories can vary in size from rather closed to fairly large, and v) noun classifiers are often used anaphorically.

According to Aikhenvald (2000), the ability of noun classifiers to function as anaphoric devices is the direct result of their grammaticalization. Craig (1986) notes that in Jakaltek Popti' the anaphoric properties of noun classifiers and their use as pronouns are an innovation in the Mayan language family, which generally lacks third person pronouns. Noun classifiers in Jakaltek Popti' have the ability to track referents in discourse and work as dependent pronouns when attached directly to the verb stem; when combined with demonstratives they form independent pronouns.

Diachronically one common process is for noun classifiers to originate from constructions involving two nouns, in which the first noun displays broader or generic semantics, such as 'person,' 'animal' or 'tree' (Aikhenvald 2000, Grinevald 2002). Over time these generic nouns undergo phonetic reduction as a result of their higher token frequency (Bybee 2011) and the resulting monosyllabic forms stand close to the noun they preceded, forming a single phonological and lexical unit with it (de León 1988, Grinevald 2002, Veerman-Leichsenring 2004). For the most part, the semantic relationship between the classifier and the full noun it goes with syntagmatically is synchronically transparent. In some cases, further erosion causes these reduced forms to become fossilized elements on the nouns they once classified. In Mesoamerica this process has been attested in Chalcatongo Mixtec (Macaulay 1987, 1996), Coatzacoapan Mixtec (Small 1990), and Zapotecan (Suárez 1983, Antonio Ramos 2013); in these languages classifiers are no longer productive and have often undergone unpredictable phonological changes. For these older forms, one can no longer talk of classifiers given their limited productivity and high degree of lexicalization (Aikhenvald 2000), although newer classifiers could have developed in the grammar by iterating the process outlined

above. Grinevald (2016) has documented this phenomenon in Jakalteq Popti', in which new noun classifiers have been added to the inventory for plastic and leather objects. This shows the importance of considering classifier systems both in a static and dynamic position on the grammaticalization cline.

In addition to this line of research on the history of nominal classification systems, other studies have found evidence that classifiers are involved in the grammaticalization of other morphosyntactic categories. Song (2005) argues that possessive classifiers have grammaticalized into benefactive markers in Oceanic languages. Similarly, Epps (2008) shows that in Hup (Nadahup, Brazil) the source of the future marker can be traced back to an incipient noun classifier system of nominal origin. Finally, Rose (2019) demonstrates that classifiers are an intermediate step that explains the development of nouns into applicatives in Mojeño Trinitario (Arawak, Bolivia).

This paper contributes to this growing number of studies that demonstrate the involvement of classifiers in the grammaticalization of other morphosyntactic functions. The following case study illustrates a grammaticalization pathway from nouns to classifiers, thence to third person pronouns, to relative pronouns, and to subordinators in *Tù'un na Ñuu Sá Mátxíí Ntxè'è*, where all these functions remain present in the synchronic grammar. In the next section I offer a description of all the grams attested that display functions overlapping along this cline.

3. Grams with overlapping functions in *Tù'un na Ñuu Sá Mátxíí Ntxè'è*

Tù'un na Ñuu Sá Mátxíí Ntxè'è is the Mixtec variety spoken in the community of San Martín Duraznos in Oaxaca (Mexico); it is part of the Mixtecan subgroup in the Otomanguean language family (Longacre 1957). There is a high degree of diversification across Mixtec varieties, and there is no agreement on how many varieties there may be (or if counting them is even possible), and where the boundaries among them lie (Josserand 1983: 119, Campbell 2017b). They tend to form dialect continua across the vast area they occupy (Jiménez Moreno 1962), which covers most of

western Oaxaca, parts of eastern Guerrero and some neighbouring areas in Puebla, Mexico. Varieties are often grouped by geographic area: Mixteca Alta, Mixteca Baja, and Mixteca de la Costa, but these are not linguistic groupings. From a linguistic perspective, the most extensive study that analyzes variation across Mixtec was carried out by Josserand in 1983. She surveyed lexical data from 120 villages where Mixtec is spoken and proposed 12 major dialectal clusters. *Tù'un na Ñuu Sá Mátxíí Ntxè'è* belongs to the Southern Baja subgroup in Josserand's (1983) proposal.

Some typological features of syllable structure and word order are important for this study. The canonical shape of lexical roots in Mixtec is the bimoraic couplet (Pike 1948, Longacre 1957), with the mora being the tone-bearing unit. This structure is realized as two short vowels in disyllabic roots and as a long vowel in monosyllabic roots. For nouns, prosodic words with three or more mora can usually be analyzed as diachronically or synchronically morphologically complex. In terms of phrase structure, *Tù'un na Ñuu Sá Mátxíí Ntxè'è* is strongly head-initial: nouns precede adjectives, demonstratives, and relative clauses. The basic constituent order is VS/VAO, both with lexical NPs and pronominal enclitic arguments.

The analysis presented here is based on 8 monologic narratives² that were collected in 2018 in the context of documenting Mixtec varieties spoken in diaspora communities in Southern California. All the texts were recorded, transcribed, and analyzed as part of a graduate field methods class at UC Santa Barbara, and were spoken by the same person, my collaborator and member of the community Carmen Hernández Martínez. Each monologue is about two minutes long and they include personal stories, recipes, descriptions of daily activities, and retellings of major events in the community. In addition, elicitation was used to clarify the form of grams (*e.g.*, forms displaying segmental and suprasegmental variation) and their paradigmatic relation in different morphosyntactic environments. All elicited sentences were based on utterances in the narratives and not the contact language(s) and involved creating the appropriate semantic and pragmatic context in each case.

² The data will be publicly available at the end of 2021 *through* the Endangered Languages Archive (ELAR) at <https://www.elararchive.org/dk0629>.

3.1. Noun classifiers in Tù'un na Ñuu Sá Mátxíí Ntxè'è

Tù'un na Ñuu Sá Mátxíí Ntxè'è has a set of seven noun classifiers that can occur preposed to nouns, as shown on Table 1. *Txà*, *ñá*, and *na* are used for human referents, whereas *txí*, *ntxà* and *tù* are used for non-human referents. In addition, *ñà* is used as a generic classifier, predominantly for non-human referents, but it can be used for human referents too when the gender of the referent is unknown.

Noun classifiers	Semantics
<i>txà</i>	Male human
<i>ñá</i>	Female human
<i>na</i>	Human (collective)
<i>ñà</i>	Generic (usu. non-human)
<i>txí</i>	Animals & round objects
<i>ntxá</i>	Liquids
<i>tù</i>	Trees

Table 1: Noun classifiers in Tù'un na Ñuu Sá Mátxíí Ntxè'è

The usage of noun classifiers tends to be highly conventionalized and idiosyncratic. The classifiers are chosen based on their semantics, and while some nouns can take different classifiers with changes in meaning, not every noun in the language takes one. For example, from the form *na-múxikà* (CLF.H.PL=music: ‘band’) in example (8) below, a similar form with the male classifier *txà-múxikà* (CLF.M=music: ‘male musician’) can be elicited and is deemed acceptable, but the equivalent with the female classifier *ñá-múxikà* (CLF.F=music: ‘female musician’) is not; rather *ñá síká'a* is preferred (REL.F=CAUS:IPFV.make.noise lit. ‘[the one] who plays’). When adding non-human noun classifiers, the meaning, while still transparent, is not always predictable: *tú-múxikà* (CLF.TREE=music) for ‘trumpet’.

- (8) ta ñà sá'a-na kíkí ñà,
 ta ñà sá'a=na kíkí ñà
 CONJ REL.GNR IPFV.do=3PL IPFV.COP SUB
 'and what they do is that'
 ntíkí-na **na-múxikà** na síká'a.
 ntíkí=na **na=múxikà** na síká'a
 IPFV.search=3PL **CLF.H.PL=music** REL.H.PL POT.play
 'they look for a band that will play' (SMD-0004-FiestaPatronal)

The diachronic origin of the noun classifiers is consistent with the cross-linguistic pattern outlined in Section 2 above involving constructions with two nouns, in which the first noun displays generic semantics. This structure is morphosyntactically tantamount to noun modifying constructions, in which the second noun restricts the meaning of the preceding one: *ntò'ò* 'basket (Sp: tenate)' > *ntò'ò shità* 'tortilla basket (Sp: tortillera).' Table 2 below summarizes the noun classifiers identified in *Tù'un na Ñuu Sá Mátxíí Ntxè'è*, their semantics, and their historical origins.

Noun classifiers	Semantics	Historical origin
txà	Male human	<i>txàa</i> 'man'
ñá	Female human	<i>ñá'a</i> 'woman'
na	Human (collective)	Unkown
ñà	Generic (usu. non-human)	<i>ñà'a</i> 'thing'
txí	Animals & round objects	<i>kitxì</i> 'animal' <i>ntxìvì</i> 'egg'
ntxá	Liquids	* <i>ndute</i> 'water'
tù	Trees	<i>itùn</i> 'wood'

Table 2: Noun classifiers in *Tù'un na Ñuu Sá Mátxíí Ntxè'è* and their historical origins

The proposed historical origin of the classifier *ntxá* for liquids is based on Josserand's (1983) reconstruction of the word for 'water' in Proto-Mixtec, which in this variety has been replaced by a different form: *txìkuíí*. Of all the classifier forms, *ntxá* is the least common, but it can be found in the word for 'soda': *ntxávìshì* (CLF.LQD:sweet).

The origin of the collective human classifier *na* is unclear and is potentially older than the other classifiers. Josserand (1983) documented a proposed *na/ñà* form in the word for 'people' in almost all of the varieties she surveyed and thus included the classifier in her reconstruction of the word for Proto-Mixtec: **yě yìwi?*. In *Tù'un na Ñuu Sá Mátxíí Ntxè'è* the

classifier has fossilized in the word for ‘people’ *nà-yivi*, in which the second component is not semantically transparent and never attested on its own. When singular, the word is always preposed by a noun classifier: *txà-yivi* ‘male person,’ *ñá-yivi* ‘female person’ or *ñà-yivi* (with the generic classifier) if the gender is unknown.

In addition, the classifier *txì* ‘animal, round thing’ is the result of a historical merger of the forms derived from *kitxì* ‘animal’ and *ntxìvì* ‘egg,’ a process that was already documented by Longacre (1957: 148) and attested in several Mixtec varieties (Macri 1983, de Ávila 2004). The classifier can be found fused into several words for animals (*txìna* ‘dog,’ *txìka* ‘grasshopper,’ *txìín* ‘mouse,’ *txìní’ín* ‘skunk,’ *txìkáká* ‘crow,’ *txìñúmi* ‘owl’), some fruits and vegetables that are prototypically round (*txìkuàá* ‘orange,’ *txìnana* ‘tomato,’ *txìkomi* ‘garlic’) and objects that are conceptualized as round (*txìntò’ò* CLF.RND:basket: ‘pitcher,’ *chìkàchì* CLF.RND:cotton: ‘blanket’ with *txì* becoming *chì* due to sibilant harmony).

Finally, there is an eighth form that is most likely related to the noun classifier system: *kà* presumably from *kàa* ‘metal, bell.’ This form is only attested preceding numerals to express time, as exemplified in example (9) below, and cannot be preposed to nouns. Its origin is likely related to church bells, which are metallic and are used to tell the time in small towns in Mexico, and the association between the object and material and its function. Given its limited productivity and high degree of idiosyncrasy in its usage, this form can be best understood as not participating in the noun classifier system, even if it is most likely related to it diachronically.

- (9) *nì ntxii lo'o ìntikò-ì xàà kà-ììnví ìntikò-ì*
nì ntxii lo'o ìntikò=ì xàà kà=ììn=ví ìntikò=ì
 MOD late little PFV.get.up=1SG already CLF.METAL=nine=until PFV.get.up=1SG
 ‘I got up a little late; it was almost nine already when I got up’ (SMD-0009-Jardín)

3.2. Dependent and Independent Pronouns

In *Tù'un na Ñuu Sá Mátxii Ntxè'è* there is one set of dependent, pragmatically neutral pronouns in the grammar, summarized below in Table 3. Their host and the ordering of the dependent pronouns after it

indicate their syntactic function: S on intransitive verbs, A and P (in this order) on transitive verbs, subject on adjectival predicates, possessor on nouns, and obliques on relational nouns.

		Singular	Plural	
1 st person		-ì	Exclusive:	-ntì
			Inclusive:	-ó
2 nd person	Informal	-ún	-ntó	
	Formal	-ní		
3 rd person	Male	-rà	Human:	-na
	Female	-ñá, -án		
	Tree	-dún		
	Animal	-rí		
	Liquid	-ntxá		
	Generic	-ñà, -àn		

Table 3: Dependent pronouns forms in Tù'un na Ñuu Sá Mátxíí Ntxè'è

There is an uncanny resemblance between the third person dependent forms and the noun classifiers presented above. The differences attested can be explained in terms of morphophonology. For example, /tɛ/ and /r/ in the pairs *txà* — *rà* for males and *txì* — *rì* for animals and round objects are often in complementary distribution, with the tap being restricted mostly to clitic-initial position and Spanish loanwords. Similarly, the female and generic third persons³ have a reduced form with a single nasal vowel that is attested in rapid speech and with phonological hosts that have a CV?V couplet structure. In this last case, the last vowel of the verbal stem is replaced by the reduced dependent pronoun; this process also takes place with first- and second-person dependent forms that do not have a consonant in their onset.

The ability of noun classifiers-turned-into-pronouns to mark possession on nouns is not uncommon (Grinevald 2000, 2002), and is often a result of their grammaticalization as anaphoric devices (Aikhenvald 2000). Examples (10-11) illustrate how third person pronouns have come to

³ These female and generic third persons forms also display tonal variation: the female form is attested with a low tone (*ñà, àn*) and the generic with a high tone (*ñá, án*). These are due to morphophonological processes that are not fully understood yet and are beyond the scope of this paper. Evidence from the texts and verbal paradigms, however, suggests that these forms are distinct and not homophonous suprasegmentally: when the female third person bears a low tone, the generic form in that same environment takes a high tone.

function in the same way as first and second persons through analogy. The pronouns with possessor function are shown in bold. The pronoun and the corresponding NP cannot co-occur, and such examples are deemed ungrammatical, as exemplified in (12).

- (10) ta ùnì kíf se'-ì
 ta ùnì kíf se'e=ì
 CONJ three IPFV.COP offspring=**1SG**
 'and I have three children' (lit 'and my children are three') (SMD-0003-Presentación)
- (11) ta nì ìkani-ì míí txìna kuíxín,
 ta nì ìkani-ì míí txìna kuíxín
 CONJ MOD PFV.hit=**1SG** TOPZ dog white
 'so I hit the white dog,'
 ta sààn ikán nì ìxinu-rí kuànu'ù-rì ve'e-**rí**
 ta sààn i=kán nì ìxinu=rí kuànu'ù=rì ve'e=**rí**
 CONJ then LOC=DEM.INV MOD PFV.run=**3ANML** PFV.leave=**3ANML** house=**3ANML**
 'and then it ran back to its house' (SMD-0007-AccidenteToby)
- (12) ta sààn ikán nì ìxinu-**rí** (*txìna)
 ta sààn i=kán nì ìxinu=**rí** (*txìna)
 CONJ then LOC=DEM.INV MOD PFV.run=**3ANML** (*dog)
 'and then it (the dog) ran' (elicited)

In addition to the dependent pronouns above, in *Tù'un na Ñuu Sá Mátxíí Ntxè'è* there is a set of independent pronouns that function as contrastive pronouns. These forms are summarized in Table 4 below. Morphologically, third person independent pronouns are formed by combining noun classifiers with one of the four demonstratives: *yó'o* (proximal), *jààn* (medial), *káa* (distal), and *kán* (invisible)⁴.

⁴ The system of independent pronouns displays some additional asymmetries. First person plural inclusive and second person singular formal lack an independent form comparable to the other first and second persons. Rather, the forms documented are created with the topicalizer *míí*, a strategy that is also available to all the other persons. The forms derived with *míí* also carry a topic shift reading resulting from the topicalizer. Analyzing the full range of functions and environments in which the pronominal forms with *míí* occur is beyond the scope of this paper.

		Singular	Plural	
1 st person		yì'ì	Exclusive:	ntì'ì
			Inclusive:	míí-ó
2 nd person	Informal	yó'ó	ntó'ó	
	Formal	míí-ní		
3 rd person	Male	txà=DEM	Human:	nà=DEM
	Female	ñá=DEM		
	Tree	tú=DEM		
	Animal	txí=DEM, rí=DEM		
	Liquid	ntxá=DEM		
	Generic	ñà=DEM		

Table 1: Independent pronouns forms in Tù'un na Ñuu Sá Mátxíí Ntxè'è

3.3. Relative pronouns

In *Tù'un na Ñuu Sá Mátxíí Ntxè'è* relative clause constructions can be headed or headless. In headed relative clauses, the head precedes the relative clause. The relativizing strategy attested in the data display relative pronouns based on the noun classifier system, with some different tonal realizations in some cases. Relative clauses with a general subordinator and a gap strategy as in Chalcatongo Mixtec (Macaulay 1996) or with interrogative pronouns as in Nieves and Melchor Ocampo Mixtec (Caponigro *et al.* 2013) are not attested in the texts, but they might occur in specific contexts that are not yet identified. The choice of the relativizer in *Tù'un na Ñuu Sá Mátxíí Ntxè'è* is based on the semantics of the relativized noun. In (13) the relativized noun is *txìna-ì* ‘my dog’ and thus the relative pronoun takes the form of the animal classifier *txí*, whereas in (14) the relative pronoun is the female classifier since the relativized noun is the speaker’s sister. Both examples show that the pronominal element coreferential with the head noun appears only clause-initially and does not appear again following the verb as in non-subordinate clauses (see example 11 for comparison).

- (13) vitxi na ntkani-ì nùù Júlíá,
 vitxi na ntkani-ì nùù Júlíá
 today MOD IPFV.tell=1SG face Julia
 ‘today I’m going to tell Julia,’
 ña kíl ña nì into'o txìna-ì txí naní Tóví.
 ña kíl ña nì into'o txìna-ì txí naní Tóví
 REL.GNR IPFV.COP SUB MOD PFV.happen dog=1SG REL.ANML IPFV.be.called Toby
 ‘what it is that happened to my dog Toby’ (SMD-0007-AccidenteToby)
- (14) so yoó tiki inkà kì'v-ì ñà íin Tijuáná,
 so yoó tiki inkà kì'vi-ì ñà íin Tijuáná
 but there.is also other same.sex.sibling=1SG REL.SG.F IPFV.live Tijuana
 ‘but there’s also my other sister who lives in Tijuana’ (SMD-0012-VisitaMamá)

Further evidence that the choice of relative pronoun is based on the semantics of the relativized noun comes from elicitation: by changing the referent from *yìkì* ‘plant’ in (15) to *itùn* ‘tree’ in (16) the relative pronoun changes from the generic noun classifier to the tree form.

- (15) nì itu'un shina-ì ntxikuíf yìkì ñà xàà vèe íchì
 nì itu'un shina-ì ntxikuíf yìkì ñà xàà vèe íchì
 MOD PFV.grab first=1SG all plant REL.GNR already AUX.COS IPFV.dry
 ‘but first I grabbed all the herbs that were already drying’ (SMD-0009-Jardín)
- (16) nì itu'un shina-ì ntxikuíf itùn tú xàà vèe íchì
 nì itu'un shina-ì ntxikuíf itùn tú xàà vèe íchì
 MOD PFV.grab first=1SG all tree REL.TREE already AUX.COS IPFV.dry
 ‘but first I grabbed all the trees that were already drying’ (elicited)

Examples (13) through (16) above all show relativizing on subjects. Relativizing on objects with the same strategy is also possible, exemplified in (17). The data also show an example of relativizing a possessor: in (18) the relative clause is introduced by the human plural form *na*, the possessor of the head *txìna* ‘dog’ and subject of the subordinate clause, instead of the animal classifier *txí* as in (13). More texts are necessary, however, to confirm the positions on the Accessibility Hierarchy (Keenan & Comrie 1977) that can be relativized.

- (17) vitxi na ntíkani-ì nùù txà iyàn
 vitxi na ntíkani=ì nùù txà iyàn
 today MOD POT.tell=1SG face CLF.M Yi-Yang
 ‘today I’m going to tell Yi-Yang,’
 ña kíkí ñà kuchúun-nti ta sùvà'a-nti
 ña kíkí ñà kuchúun=nti ta sùvà'a=nti
 REL.GNR IPFV.COP SUB IPFV.USE=1PL.EXCL CONJ MAKE=1PL.EXCL
 ‘what it is that we use when we make’
 síin síin nùú kò'ò komídá ñà xíxá'an=va=ó
 síin síin nùú kò'ò komídá ñà xíxá'an=va=ó
 different different type plate food REL.GNR IPFV.eat=ADD=1PL.INCL
 ‘the many different types of dishes that we eat’ (SMD-0008-Hierbas)
- (18) ìnkà txìna ká'nu na ntóo,
 ìnkà txìna ká'nu na ntóo,
 other dog big REL.H.PL IPFV.COP
 ‘another big dog whose owners live’
 yatxi míí nùù ntóo-nti ikán,
 yatxi míí nùù ntóo=nti i=kán,
 near TOPZ FACE IPFV.COP=1PL.EXCL LOC=DEM.INV
 ‘in front of where we live there,’
 nì ìxàà-rì ve'-ì,
 nì ìxàà=rì ve'e=ì,
 MOD PFV.arrive=3ANML house=1SG
 ‘it arrived at my house’ (SMD-0007-AccidenteToby)

Incidentally, these two examples also display the development of body-parts like *nùù* ‘face’ into prepositions (17) and subordinating conjunctions (18), a process that is well attested in Otomanguean languages (Brugman & Macaulay 1986, Hollenbach 1995b, Lillehaugen 2004).

Headless relative clause constructions also use the noun classifier forms as relative pronouns. In (19-20) the relative pronoun takes the form of the male noun classifier *txà* since the head (or domain nominal) is male, although headless relative clauses do not typically refer anaphorically to a noun phrase previously mentioned in the discourse. A nominal head for the headless relative clause can be elicited (21) as long as the head is not the source lexical item of the relativizer: **txàa txà* ‘man REL.M’ is deemed ungrammatical.

- (19) ta ikán kíf nùù nì ìxìni-ì **txà** kíf tátà se'-ì
 ta ikán kíf nùù nì ìxìni-ì txà kíf tátà
 CONJ LOC=DEM.INV IPFV.COP face MOD PFV.see=1SG **REL.SG.M** IPFV.COP father
 se'e=ì
 offspring=1SG
 'and there is where I met (the one) who is the father of my children' (SMD-0011-MiHistoria)
- (20) ta ikán kíf nùù nì ìxìni-ì **txà** ìxà'àn Nùùntóvá
 ta ikán kíf nùù nì ìxìni-ì txà ìxà'àn Nùùntóvá
 CONJ LOC=DEM.INV IPFV.COP face MOD PFV.see=1SG **REL.SG.M** PFV.leave Oaxaca
 'and there is where I met (the one) who left to Oaxaca' (elicited)
- (21) ta ikán kíf nùù nì ìxìni-ì **txà-múxìkà** txà ìxà'àn Nùùntóvá
 ta ikán kíf nùù nì ìxìni-ì **txà=múxìkà** txà
 CONJ LOC=DEM.INV IPFV.COP face MOD PFV.see=1SG **CLF.M=music** **REL.SG.M**
 ìxà'àn Nùùntóvá
 PFV.leave Oaxaca
 'and there is where I met the (male) musician who left to Oaxaca' (elicited)

Headless relative clauses are a productive strategy for coining new terms. Over time they can become conventionalized and lexicalize. The word *ñà-txítxìin* (REL.GNR=IPFV.paste) for 'glue' in (22) is a good example of this process. Other examples of documented cases include *txà-kuì'na* (REL.M=IPFV.steal for 'thief'), *txà-síkua'a* (REL.M=IPFV.study for 'student'), and *ñá-síká'a* (REL.F=CAUS:IPFV.make.noise for 'female musician')⁵.

- (22) sayá'a **ñà**-txítxìin jààn nù-ìn
 sa-yá'a **ñà**=txítxìin jààn nùù-ì
 CAUS-IPFV.pass REL.GNR=IPFV.paste MED.DEM face=1SG
 'pass me the glue' (elicited)

Headless relative clauses are also common with quantified expressions and numerals, as exemplified in (23). The semantics of the relativized noun are still accessible in headless relative clauses, and alternative relative pronouns can be elicited based on spoken utterances by creating the appropriate semantic context (24-26).

⁵ Hollenbach (1995a) analyzes similar examples in the varieties of Mixtec, Trique and Cuicatec that she surveyed as nominalizations. However, there does not seem to be enough evidence in Tù'un na Ñuu Sá Mátxíí Ntxè'è to suggest these examples are structurally different from headless relative clauses. Proposing a distinct function in these examples would require two identical sets of forms, one with a nominalizing function and the other with a relativizing one. Further evidence that these are best analyzed as lexicalized subordinate clauses comes from other new terms coined with the same strategy but introduced by body-parts (e.g., *nùù* 'face') turned into conjunctions: *nùù kásún páàn* (SUB=IPFV.get.toasted-bread for 'toaster'), *nùù xikutxíyá'á-na* (SUB=IPFV.swing=3PL for 'hammock').

- (23) ñàkán kíí ntxikuuí **ña** kúníñú'-àn,
 ñà=kán kíí ntxikuuí **ña** kúníñú'ú=àn
 3GNR=DEM.INV IPFV.COP all **REL.GNR** POT.need=3SG.F
 'that is everything that she will need'
 távà na koan sùvà'-án iin **ña** kíí ntxéí kuáán
 távà na koan sùvà'a=án iin **ña** kíí ntxéí kuáán
 CONJ MOD POT.be.able POT.make=3SG.F one REL.GNR IPFV.COP mole yellow
 'to be able to make one mole amarillo' (SMD-0005-ArrozAmarillo)
- (24) sántuvi-ná ntxikuuí **ña** kuchúún-na
 sá-ntuvi=ná ntxikuuí **ña** kuchúún=na
 CAUS-IPFV.clean=3PL all **REL.GNR** POT.use=3PL
 'they prepare everything that they will use' (SMD-0004-FiestaPatronal)
- (25) sántuvi-ná ntxikuuí **tù** kuchúún-na
 sá-ntuvi=ná ntxikuuí **tù** kuchúún=na
 CAUS-IPFV.clean=3PL all **REL.TREE** POT.use=3PL
 'they prepare all [trees/wood] that they will use' (elicited)
- (26) sántuvi-ná ntxikuuí **txí** kuchúún-na
 sá-ntuvi=ná ntxikuuí **txí** kuchúún=na
 CAUS-IPFV.clean=3PL all **REL.RND** POT.use=3PL
 'they prepare all [round fruits/vegetables] that they will use' (elicited)

Arguably, examples (23-26) could be analyzed as having a light head (Citko 2004). It is unclear, however, whether these quantified expressions can have a pronominal reading without the following noun phrase. In the data, they always appear preceding a lexical noun phrase (27), a classifier turned into pronoun (28), or a headless relative clause as exemplified above.

- (27) ta kixa'á ntataka **ntxikuuí na-ve'e mátómá**
 ta kixa'á ntataka **ntxikuuí na=ve'e mátómá**
 CONJ IPFV.start IPFV.gather **all CLF.H.PL=house stewardship**
 'All people from the stewardship start gathering' (SMD-0004-FiestaPatronal)
- (28) nì ìkà'-ìn jí'-án **ña** na sùvà'-án **i-an** kì'-ìn
 nì ìkà'àn=ì jí'in=án **ña** na sùvà'a=án **iin=an**
 MOD PFV.speak=1SG with=3SG.F SUB MOD POT.make=3SG.F **one=3GNR**
 kì'in=ì
 POT.grab=1SG
 'I told her that she could make one [bag] for me' (SMD-0012-VisitaMamá)

In the texts the most common relativizer attested in both relative clause constructions is the form *ñà*. This is unsurprising given that its origin can be traced back to the word *ñà'a* 'thing' and that it has developed into a generic classifier and pronoun. The generic semantics of *ñà* is most likely the reason why it is the only relative pronoun attested in pseudo-cleft

constructions: a copular construction with a headless relative clause in one slot and a noun phrase in the other, or with a headless relative preceding and following the copula (Payne 1997). None of the other relativizers is ever attested or can be elicited in pseudo-cleft constructions with the same reading. This construction is a very common discourse strategy to signal that the upcoming segments of talk are important (Hopper 2001). In the texts pseudo-clefts are used to introduce new referents into the discourse (29) and state the general theme of the next utterances (30).

- (29) ta vitxi **ñà** ntíkani-ì kíkí yìkì kinì
 ta vitxi **ñà** ntíkani=ì kíkí yìkì kinì
 CONJ today REL.GNR POT.tell=1SG IPFV.COP oregano
 ‘and today what I will talk about is the oregano’ (SMD-0008-Hierbas)
- (30) ta **ñà** sá'-án kíkí ña kotóò ntxìv'à'-án sá'-án ntò'ò
 ta **ñà** sá'a=án kíkí ña kotoó=ntxìv'à'=án sá'a=án
 CONJ REL.GNR IPFV.DO=3SG.F IPFV.COP SUB IPFV.like=very=3SG.F IPFV.do=3SG.F
 ntò'ò
 basket
 ‘what she does is that she really likes making baskets’ (SMD-0012-VisitaMamá)

3.4. The subordinator ña

The form *ña* —sometimes *ñà*— can also introduce two other types of subordinate clauses: complement clauses and adverbial clauses. Examples (31)-(32) show complement clauses introduced by *ña*. In both examples a headless reading is not possible, because the verbs in the subordinate clause are intransitive and display their S argument marked with an enclitic pronominal subject in (31) and a lexical NP in (32).

- (31) ta xàà ìxini-àn **ñà** vèè kúkuá-àn lo'o sààn chikà-àn ña kíkí ya'á ìchí
 ta xàà ìxini=àn **ñà** vèè kúkuán=àn lo'o
 CONJ already PFV.see=3SG.F SUB AUX.COS IPFV.become.yellow=3GNR little
 sààn chikàà=àn ña kíkí ya'á ìchí
 then POT.put.inside=3SG.F REL.GNR IPFV.COP chili dry
 ‘when she sees that it becomes a little yellow, then she will add the dried chili’
 (SMD-0005-ArrozAmarillo)
- (32) chii ìtxasò'-ì **ñà** nté'è txìna lo'-ì
 chii ìtxasò'o=ì **ñà** nté'è txìna lo'o=ì
 because PFV.listen=1SG SUB IPFV.bark dog small=1SG
 ‘because I heard that my little dog was barking’ (SMD-0007-AccidenteToby)

However, the form *ñā* can be potentially ambiguous. By creating the appropriate context, a subordinate clause introduced by *ñā* can have a headless relative clause reading, where *ñā* functions as a generic relative pronoun, or a complement clause reading, with *ñā* functioning as a subordinator (33). Hollenbach (1995a) has suggested that this double interpretation of some subordinate clauses illustrates that these grams are historically related and provides the bridging context for reanalysis from a headless relative clause to a complement clause.

- (33) xín-ì **ñā** ìkà'àn-rà
 xínì=i *ñā* ìkà'àn=rà
 IPFV.know=1SG REL.GNR/SUB PFV.speak=3SG.M
 'I know what he said' HEADLESS RELATIVE CLAUSE
 'I know that he spoke' COMPLEMENT CLAUSE (elicited)

Adverbial clauses introduced by *ñā* are scarce in the texts, but they are attested for reason, purpose, and temporal clauses. For example, in (34) the clause introduced by *ñā* appears to be an adverbial reason clause: 'I am very happy because she came to visit me.' In (35) the purpose reading of the subordinate clause can only be obtained with *ñà*. If the subordinator is left out in (35), an asyndetic clause coordination reading is obtained: "we fight (and) we come out ahead."

- (34) kúsì ntxìvà'a ini-ì **ñā** nì ìkixi-án ìxìtonì'ni-án yìì
 kúsì=ntxìvà'a ini-ì **ñā** nì ìkixi=án ìxìtonì'ni=án yìì
 IPFV.be.happy=very inside=1SG SUB MOD PFV.come=3SG.F PFV.visit=3SG.F 1SG
 'I am very happy because she came to visit me' (SMD-0012-VisitaMamá)
- (35) na sá'-ó lúchá **ñà** kan-ó chínùùn
 na sá'a=ó lúchá **ñà** kana=ó chínùùn
 MOD IPFV.do=1PL.EXCL fight SUB POT.leave=1PL.EXCL ahead
 'we fight to come out ahead' (SMD-0011-MiHistoria)

In adverbial clauses with a temporal meaning the subordinator *ñā* can follow an adverb or a noun, although the subordinator is not always present. The example in (36) contains sequential intonation units, but only in the first one does *ñā* follow *vitxi* 'today.' It is unclear what principles determine the use of the subordinator and whether each intonation unit in (36) are structurally equivalent.

- (36) ta **vitxi ña xàà** ìxà'a-nà titi-àn,
 ta **vitxi ña xàà** ìxa'a=nà
 CONJ **now SUB already** PFV.give=3PL
 'and now that they already gave her documents'
vitxi xàà kíxi-án nùù-ntí
vitxi xàà kíxi=án nùù=ntí
now already IPFV.come=3SG.F face=1PL.INCL
 'now she already comes to us' (SMD-0012-VisitaMamá)

The example in (37) shows a temporal adverbial clause introduced by *ña* following the noun *kuià* 'year.' This construction is structurally very similar to headed relative clauses introduced by the generic relative pronoun. This structural resemblance suggests that adverbial clauses might indeed be historically related to relative clause constructions. Additionally, temporal adjuncts have access to relativization; in (38) the relativized noun 'month' is the subject of the subordinate clause and functions as an adjunct in the matrix clause.

- (37) ta vitxi xàà yoó ìxì kuià **ña** nì ìnto'-ì saa
 ta vitxi xàà yoó ìxì kuià **ña** nì ìnto'o=ì saa
 CONJ today already there.is ten year **SUB** MOD PFV.stay=1SG like.this
 'today it has been ten years since this happened' (Lit. 'today it has been ten years that I stayed like this' (SMD-0011-MiHistoria))
- (38) sáchu-ìn yóó **ña** kumí ókó ìxì kìì
 sá-chuun=ì yóó **ña** kumí ókó ìxì kìì
 CAUS-IPFV.work=1SG month **REL.GNR** IPFV.have twenty ten day
 'I work in months that have thirty days' (elicited)

3.5. The linker *ña*

As mentioned above, in noun modifying constructions in *Tù'un na Ñuu Sá Mátxíí Ntxè'è* the second noun restricts the meaning of the preceding one: *ntò'ò* 'basket (Sp: tenate)' > *ntò'ò shità* 'tortilla basket (Sp: tortillera).' In some cases, however, a general linker *ña* is attested in between both nouns, as in (7) above, repeated in (39). Unlike the noun classifiers described above, the linker can usually be omitted without a significant difference in meaning. However, some evidence that the linkers might be historically related to classifiers comes from elicitation: on the basis of (39) an alternative linker can be elicited with a slight change of meaning (40).

- (39) ntiká'nu,
 ntiká'nu
 IPFV.enlarge
 ña kíl to'ó ña ñuu,
 ña kíl to'ó ña ñuu
 REL.GNR IPFV.COP saint LNK village
 'the saint of the village gets aggrandized' (lit. 'what is the saint of the village gets aggrandized') (SMD-0004-FiestaPatronal)
- (40) ntiká'nu,
 ntiká'nu
 IPFV.enlarge
 ña kíl to'ó na ñuu,
 ña kíl to'ó na ñuu
 REL.GNR IPFV.COP saint LNK village
 'the saint of the community gets aggrandized' (lit. 'what is the saint of the people of the village gets aggrandized') (elicited)

Arguably, in (40) the form *na* could be working as the human plural noun classifier rather than a linker: the semantic change from 'village' to 'community' would be consistent with this analysis. Presumably, a classifier reading is also possible in (39), *i.e.*, '(what is) the saint of the things of the village gets aggrandized.' However, this interpretation is not possible in (41): *úniversídád ña Sántá Bárbará* 'the university of the things of Santa Barbara'.

- (41) míi kíl vitxi ntíkàà=ì ntxáá ña kíl úniversídád ña Sántá Bárbará
 míi kíl vitxi ntíkàà=ì ntxáá ña kíl úniversídád ña
 TOPZ day today IPFV.be.inside=1SG from REL.GNR IPFV.COP university LNK
 Sántá Bárbará
 Santa Barbara
 'today I am at the University of Santa Barbara' (SMD-0011-MiHistoria)

Rather, in the constructions in which the general linker *ña* is attested the relationship between the two nouns can be best analyzed as a case of adnominal modification: in (39) *to'ó ña ñuu* can be interpreted as 'the saint in the village'. The same reading is plausible in (41): *úniversídád ña Sántá Bárbará* 'the university in Santa Barbara.' It is possible that the linker functions as an associative marker of the type described for Mandarin (Li & Thompson 1981, Chappell & Thompson 1992) which can appear between two nouns and encompasses a wide range of associations between them. More texts are necessary to identify whether the constructions in

(39-41) are equivalent and what the semantic and pragmatic factors are for the usage of the linker.

4. Grammaticalization pathway

The following summary outlines the series of steps that nouns with generic meanings may have followed in their grammaticalization to noun classifiers, pronouns, and relativizers and the subsequent grammaticalization of the generic *ñà* into a subordinator:

- Constructions involving two nouns in sequence are commonly used in the language. In these constructions, the second noun modifies or restricts the meaning of the preceding noun.
- Some of the nouns in the first position in the construction that display rather generic meanings begin to erode, presumably because of their higher token frequency. The bimoraic root is reduced to a monomoraic form, and other unpredictable phonological changes occur (*e.g.*, loss of vowel nasality from *itùn* to *tù* for the ‘tree’ root).
- These reduced forms develop an identity as classifying elements. The semantic range of each form is established (*e.g.*, *ñà* extends its meaning from ‘thing’ to a generic).
- Classifiers combine productively with nominal stems and adjectives that can get a noun interpretation (*e.g.*, *txà-vàlí* CLF.M=small ‘boy’ or *nà-xá’nu* CLF.H.PL=big ‘elders’). Over time, some forms become fossilized on the noun they precede (*e.g.*, *txikáká* CLF.ANML=crow ‘crow’).
- Classifiers begin to combine with demonstratives. This process leads to the development of independent third person pronouns. Third person dependent pronouns are formed through analogy with dependent first and second persons.
- By acquiring anaphoric properties, the classifier forms gain a more grammatical status: they begin to stand in place for a noun or domain nominal, and not only next to them.

- In addition to combining with nouns, adjectives and demonstratives, classifiers begin to combine with clauses. This process gives rise to headless relative clause constructions.
- The generic relative pronoun *ñā* becomes a subordinator by reanalysis of headless relative clauses into complement clauses in environments that allow both interpretations: I know what he said > I know that he spoke.
- Similarly, headed relative clause constructions in which the head encodes temporal meaning are reanalyzed as adverbial clauses. This process contributes to the grammaticalization of *ñā* as a subordinator, as the default relative pronoun for these head nouns is the generic *ñā*.

All the manifestations resulting from the grammaticalization pathway coexist synchronically, with minimal tonal variations. The different functions of the grams are disambiguated based on the syntactic and semantic context. While the steps outlined above may suggest a linear order, some of these processes most likely occurred in parallel: as is well-known grammaticalization pathways do not involve abrupt transitions, but grammaticalizing elements rather go through intermediate stages with different lexico-grammatical status at each point. Figure 2 summarizes a possible grammaticalization pathway that is not entirely linear and captures different grammaticalization processes occurring at the same time.

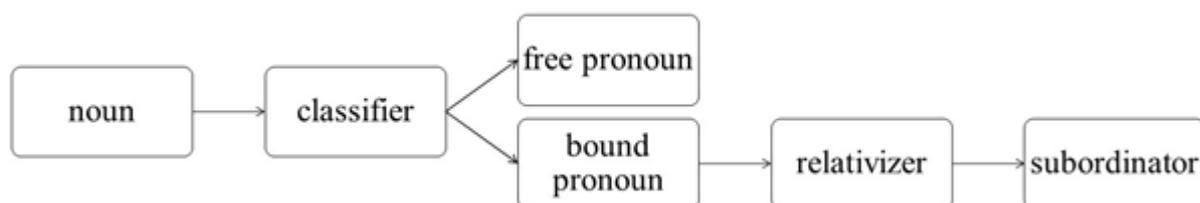


Figure 2: Grammaticalization pathway with some processes occurring in parallel

Similarly, forms participating in the same grammaticalization process need not move at the same speed or arrive at the same endpoints. This is apparent in the case of *Tù'un na Ñuu Sá Mátxíí Ntxè'è*: examples like (9) suggest that the development of the form *kà* from *kàa* ‘metal, bell’ is most likely related to the classifier system diachronically, but synchronic uses

suggest it did not participate further in the grammaticalization that the other noun classifiers underwent. On the other side of this continuum, the form *ña* moved beyond its function as a relative pronoun and grammaticalized as a subordinator in complement clauses and adverbial clauses. In addition, lexicalizations and fossilized classifier forms are well attested in *Tù'un na Ñuu Sá Mátxii Ntxè'è*. All this evidence demonstrates the idiosyncrasies of the classifiers and classifier-related grams on the lexico-grammatical continuum (Grinevald 2002).

The first part of the grammaticalization pathway described here is not new. The grammaticalization from different generic nouns to noun classifiers and thence to third person pronouns in *Tù'un na Ñuu Sá Mátxii Ntxè'è* is consistent with the processes documented in other Mixtec varieties (de León 1988, Macaulay 1987, 1996, Mantenido 2020), and in Mesoamerica in general (Craig 1977, 1979, 1986, England 1983, Hopkins 2012, Zavala, 2000).

The development of relative pronouns from classifiers might seem unusual at first, but it is rather natural from the perspective of the constructions involved: classifiers extended the range of elements they could combine with from nouns, adjectives and demonstratives to full clauses. Epps (2012) has documented a similar process in Hup (Nadahup, Brazil), in which the domain nominal in one type of relative clause construction is a bound-noun or a classifying noun. In Rawang (Sino-Tibetan, Tibeto-Burman; Myanmar) classifiers are attested as heads of relative clauses (LaPolla 2008) and in Chaozhou (Sino-Tibetan, Sinitic; China) the classifier *kai* has seemingly been reanalyzed as a relativizer after acquiring anaphoric properties (Xu & Matthews 2011). *Tù'un na Ñuu Sá Mátxii Ntxè'è* likely represents a more grammaticalized case of that same pattern, given that the grams show evidence of functioning as full relative pronouns: in headed relative clauses the relativized noun is followed by the classifier-turned-into-relativizer of the same semantic domain (male, female, tree, animal, etc.). Overall, the case of *Tù'un na Ñuu Sá Mátxii Ntxè'è* provides new insight into the origins and sources of relativizers.

Finally, the grammaticalization of relativizers into complementizers and subordinators in general is also attested cross-linguistically (Kuteva *et al.* 2019), although the reverse pattern is more common in certain parts of the world (Heine & Kuteva 2006). Hollenbach (1995a) already suggested that the complementizer developed from the relativizer in the varieties of Mixtec, Trique and Cuicatec that she surveyed. In *Tù'un na Ñuu Sá Mátxíi Ntxè'è* the form that underwent this development is the generic relative pronoun. This last development can be explained given the core semantics of the gram as a generic classifier and pronoun and its high frequency in discourse: referents that do not fall under one of the other six semantic domains are tracked with the form *ña*.

In addition to complement clauses, the same subordinator in *Tù'un na Ñuu Sá Mátxíi Ntxè'è* is attested introducing adverbial clauses. Examples like (37) above suggest that, diachronically, adverbial clauses might also have developed from relative clause constructions by reanalysis of the headed relative clause modifying a noun in the matrix clause as a modifier of the entire main clause. This is not entirely unsurprising, as cross-linguistically adverbial clauses that express time, location and manner often share structural properties with, and can resemble, relative clause constructions (Thompson *et al.* 2007). In addition, the development of adverbial clauses from relative clauses is not entirely unattested in the Americas (Epps 2009, Estrada-Fernández 2012), although structurally the constructions documented differ significantly from *Tù'un na Ñuu Sá Mátxíi Ntxè'è*. More research is needed on the range of adverbial clauses that can be introduced by the subordinator *ña* in *Tù'un na Ñuu Sá Mátxíi Ntxè'è* and their relationship to adnominal modification in order to understand their historical development.

5. Conclusion

Analyzing a gram with overlapping functions remains a well-known analytical challenge for the field linguist. This paper addresses the importance of considering semantic, typological and diachronic evidence in analyzing several grams in the language that bear a clear formal resemblance. When grams display segmental and suprasegmental variation, elicitation on the basis of attested utterances —rather than the

contact language(s)— has proven to be very useful if the appropriate semantic and pragmatic context is created. The case of *Tù'un na Ñuu Sá Mátxíí Ntxè'è* shows that grammatical categories along a cline can often coexist synchronically and reflect the continuum by which they emerged.

This study contributes to the overall understanding of noun classifiers in Mesoamerica and their grammaticalization into morphosyntactic categories beyond nominal classification. The grammaticalization pathway proposed here involves nouns with broader, generic semantics developing into classifiers, thence to third person pronouns, to relative pronouns, and finally to subordinators. This development in *Tù'un na Ñuu Sá Mátxíí Ntxè'è* deepens our understanding of the sources of relativizers and subordinators cross-linguistically and provides new evidence for the widespread resemblance of relative clause constructions and adverbial clauses.

References:

- Aikhenvald, A. Y. (2000). *Classifiers. A Typology of Noun Categorization Devices*. Oxford University Press.
- Allan, K. (1977). Classifiers. *Language*, 53, 285-311.
- Antonio Ramos, P. (2013). Los aparentes clasificadores nominales en el zapoteco de San Pedro Mixtepec (ZSPM). *La Linguistique*, 49(2), 33-50.
- Bisang, W. (1993). Classifiers, quantifiers and class nouns in Hmong. *Studies in Language*, 17, 1-51.
- Brugman, C., & Macaulay, M. (1986). Interacting Semantic Systems: Mixtec Expressions of Location. *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistic Society*, 315-327.
- Bybee, J. L. (2011). Usage-based theory and grammaticalization. *The Oxford Handbook of Grammaticalization*, 69-78.
- Campbell, E. W. (2017a). Otomanguean historical linguistics: Exploring the subgroups. *Language and Linguistics Compass*, 11(7), e12244.
- Campbell, E. W. (2017b). Otomanguean historical linguistics: Past, present, and prospects for the future. *Linguistics and Language Compass*, 11(4), 1-22.

- Caponigro, I., Torrence, H., & Cisneros, C. (2013). Free Relative Clauses in two Mixtec languages. *International Journal of American Linguistics*, 79(1), 61-96.
- Chappell, H., & Thompson, S. A. (1992). The semantics and pragmatics of associative de in Mandarin discourse. *Cahiers de Linguistique - Asie Orientale*, 21(2), 199-229.
- Citko, B. (2004). On Headed, Headless, and Light-Headed Relatives. *Natural Language & Linguistic Theory*, 22(1), 95-126.
- Contini-Morava, E. (1994). *Noun classification in Swahili*. Virginia: Publications of the Institute for Advanced Technology in the Humanities, University of Virginia.
- Contini-Morava, E., & Kilarski, M. (2013). Functions of nominal classification. *Language Sciences*, 40, 263-299.
- Costaouec, D., & Swanton, M. (2015). Classification nominale en Ixcatèque. *La Linguistique*, 51(2), 201-238.
- Craig, C. G. (1977). *The Structure of Jacaltec*. Austin: University of Texas Press.
- Craig, C. G. (1979). Jacaltec: Fieldwork in Guatemala. In T. Shopen (Ed.), *Languages and Their Speakers*, 3-57. Philadelphia: University of Pennsylvania Press.
- Craig, C. G. (1986). Jacaltec noun classifiers. *Lingua*, 70, 241-284.
- de Ávila Blomberg, A. (2004). La clasificación de la vida en las lenguas de Oaxaca. In A. J. García-Mendoza, M. de J. Ordoñez, & M. Briones-Salas (Eds.), *Biodiversidad de Oaxaca*, 481-539. Mexico City: Instituto de Biología UNAM, Fondo Oaxaqueno para la Conservación de la Naturaleza, World Wildlife Fund.
- de León Pasquel, M. de L. (1988). *Noun and numeral classifiers in Mixtec and Tzotzil*. PhD dissertation, University of Sussex.
- DeLancey, S. (1986). Toward a history of Tai classifier systems. In C. G. Craig (Ed.), *Noun Classes and Categorization*, 437-452. Amsterdam: John Benjamins Publishing Company.
- Derbyshire, D., & Payne, D. (1990). Noun classification systems of Amazonian languages. In D. Payne (Ed.), *Amazonian linguistics. Studies*

- in lowland South American languages*, 243-271. Austin: University of Texas Press.
- Dixon, R. M. W. (1982). *Where have all the adjectives gone? and other essays in semantics and syntax*. Berlin: Mouton de Gruyter.
- Dixon, R. M. W. (1986). Noun class and noun classification. In *Noun Classes and Categorization*, 105-112. Amsterdam: John Benjamins Publishing Company.
- Dixon, R. M. W. (2002). *Australian Languages*. Cambridge: Cambridge University Press.
- England, N. C. (1983). *A grammar of Mam, a Mayan language*. Austin: University of Texas Press.
- Epps, P. (2007). Birth of a noun classification system: the case of Hup. In W. L. 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*, 107-127. Leiden: Research School CNWS, Universiteit Leiden.
- Epps, P. (2008). From “wood” to future tense: Nominal origins of the future construction in Hup. *Studies in Language*, 32(2), 383-404.
- Epps, P. (2009). Escape from the noun phrase: From relative clause to converb and beyond in an Amazonian language. *Diachronica*, 26(3), 287-318.
- Epps, P. (2012). Between headed and headless relative clauses. In B. Comrie & Z. Estrada-Fernández (Eds.), *Relative Clauses in Languages of the Americas: A typological overview*, 191-211. Amsterdam: John Benjamins Publishing Company.
- Estrada-Fernández, Z. (2012). From demonstrative to relative marker to clause linker. In B. Comrie & Z. Estrada-Fernández (Eds.), *Relative Clauses in Languages of the Americas: A typological overview*, 127-146. Amsterdam: John Benjamins Publishing Company.
- Fedden, S., & Corbett, G. G. (2017). Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: A Journal of General Linguistics*, 2(1), 34.

- Fedden, S., & Corbett, G. G. (2018). *Extreme classification*. *Cognitive Linguistics*, 29(4), 633-675.
- Grinevald, C. (2000). A morphosyntactic typology of classifiers. In G. Senft (Ed.), *Systems of nominal classification*, 50-92. Cambridge: Cambridge University Press.
- Grinevald, C. (2002). Making sense of nominal classification systems. Noun classifiers and the grammaticalization variable. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization*, 259-275. Amsterdam: John Benjamins Publishing Company.
- Grinevald, C. (2016). The Jakalteq Popti' noun classifier system Changes due to Spanish contact. In A. L. Berez-Kroeker, D. M. Hintz, & C. Jany (Eds.), *Language Contact and Change in the Americas. Studies in honor of Marianne Mithun*, 273-295. Amsterdam: John Benjamins Publishing Company.
- Grinevald, C., & Seifart, F. (2004). Noun classes in African and Amazonian languages: Towards a comparison. *Linguistic Typology*, 8(2), 243-285.
- Guadalupe Joaquina, A. (2014). *Ndusu Tu'un Savi. Ve'e Tu'un Savi A.C Academia de la Lengua Mixteca*.
- Heine, B., & Kuteva, T. (2006). *The Changing Languages of Europe*. Oxford: Oxford University Press.
- Hollenbach, B. E. (1995a). Cuatro morfemas funcionales en las lenguas mixtecas. In R. Arzápalo Marín & Y. Lastra (Eds.), *Vitalidad e influencia de las lenguas indígenas en Latinoamérica: II Coloquio Mauricio Swadesh*, 284-293. Mexico City: Universidad Nacional Autónoma de México.
- Hollenbach, B. E. (1995b). Semantic and Syntactic Extensions of Body-Part Terms in Mextecan: The Case of "Face" and "Foot." *International Journal of American Linguistics*, 61(2), 168-190.
- Hopkins, N. A. (2012). The noun classifiers of Cuchumatán Mayan languages: A case of diffusion from Otomanguean. *International Journal of American Linguistics*, 78(3), 411-427.

- Hopper, P. J. (2001). Grammatical Constructions and their Discourse Origins: Prototype or Family Resemblance? In M. Pütz, S. Niemeier, & R. Dirven (Eds.), *Applied Cognitive Linguistics I: Theory and Language Acquisition*, 109-129. Berlin: Mouton de Gruyter.
- Jiménez Moreno, W. (1962). Estudios mixtecos. In F. F. de Alvarado (Ed.), *Vocabulario en lengua mixteca*, 9-105. Mexico City: Instituto Nacional Indigenista e Instituto Nacional de Antropología e Historia S.E.P.
- Josserand, J. K. (1983). *Mixtec dialect history*. PhD dissertation, Tulane University.
- Keenan, E. L., & Comrie, B. (1977). Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*, 8(1), 63-99.
- Krasnoukhova, O. (2012). The Noun Phrase in the *Languages of South America*. Utrecht: LOT.
- Kuteva, T., Heine, B., Hong, B., Long, H., & Narrog, H. (2019). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- LaPolla, R. J. (2008). Relative Clause Structures in the Rawang language. *Language and Linguistics*, 9(4), 797-812.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- Lillehaugen, B. D. (2004). The Syntactic and Semantic Development of Body Part Prepositions in Valley Zapotec Languages. *Proceedings from the Sixth Workshop on American Indigenous Languages*, Santa Barbara Papers in Linguistics, 14, 69-83.
- Longacre, R. E. (1957). Proto-Mixtecan. Bloomington: Indiana University Research Center in Anthropology, *Folklore and Linguistics*.
- Macaulay, M. (1987). Cliticization and the Morphosyntax of Mixtec. *International Journal of American Linguistics*, 53(2), 119-135.
- Macaulay, M. (1996). *A grammar of Chalcatongo Mixtec*. Berkeley: University of California Press.
- Macri, M. J. (1983). Two noun class systems in Mixtec. In A. Schlichter, W. L. Chafe, & L. Hinton (Eds.), *Studies in Mesoamerican Linguistics*, 291-306. Berkeley: University of California Press.

- Mantenuto, I. (2020). *'Other' and 'More' in San Sebastián del Monte Mixtec and Beyond*. PhD dissertation, University of California, Los Angeles.
- Martins, S., & Martins, V. (1999). Makú. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *The Amazonian Languages*, 251-267. Cambridge: Cambridge University Press.
- Meissner, A., & Storch, A. (Eds.). (2000). *Nominal Classification in African Languages*. Cologne: Rüdiger Köppe Verlag.
- Mithun, M. (1986). The Convergence of Noun Classification Systems. In *Noun Classes and Categorization*, 379-397. Amsterdam: John Benjamins Publishing Company.
- Payne, T. E. 1997. *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge: Cambridge University Press.
- Pike, K. L. (1948). *Tone Languages: A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*. Ann Arbor: University of Michigan Press.
- Reid, N. (1997). Class and classifier in Ngan'gityemerri. In M. Harvey & N. Reid (Eds.), *Nominal classification in aboriginal Australia*, 165-228. Amsterdam: John Benjamins Publishing Company.
- Rose, F. (2019). From classifiers to applicatives in Mojeño Trinitario: A new source for applicative markers. *Linguistic Typology*, 23(3), 435-466.
- Sands, K. (1995). Nominal Classification in Australia. *Anthropological Linguistics*, 37(3), 247-346.
- Senft, G. (2007). Nominal Classification. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics*, 676-696. Oxford: Oxford University Press.
- Small, P. C. (1990). A Syntactic Sketch of Coatzospan Mixtec. In H. C. Bradley & B. E. Hollenbach (Eds.), *Studies in the Syntax of Mixtecan Languages*, vol. 2, 261-479. Dallas: Summer Institute of Linguistics & University of Texas at Arlington.

- Song, J. J. (2005). Grammaticalization and structural scope increase: Possessive-classifier-based benefactive marking in Oceanic languages. *Linguistics*, 43(4), 795-838.
- Suárez, J. A. (1983). *The Mesoamerican Indian Languages*. Cambridge: Cambridge University Press.
- Thompson, S. A., Longacre, R. E., & Hwang, S. J. J. (2007). Adverbial clauses. In T. Shopen (Ed.), *Language Typology and Syntactic Description. Volume II: Complex Constructions* (2nd ed., 237-300). Cambridge: Cambridge University Press.
- Valenzuela, P. M. (2016). Classifiers in Shiwilu (Kawapanan) in Northwestern Amazonian Perspective. *Anthropological Linguistics*, 58(4), 333-380.
- Veerman-Leichsenring, A. (2004). Popolocan noun classifiers: A reconstruction. *International Journal of American Linguistics*, 70(4), 416-451.
- Vidal, A. (1997). Noun classification in Pilagá (Guaykuran). *Journal of Amazonian Languages*, 1(1), 58-111.
- Wilkins, D. P. (2000). Ants, Ancestors and Medicine: a semantic and pragmatic account of classifier constructions in Arernte (Central Australia). In Senf (Ed.), *Systems of nominal classification*, 147-216. Cambridge: Cambridge University Press.
- Xu, H. L., & Matthews, S. (2011). On the polyfunctionality and grammaticalization of the morpheme kai in the Chaozhou dialect. In F. H. Yap, K. Grunow-Hårsta, & J. Wrona (Eds.), *Nominalization in Asian Languages: Diachronic and typological perspectives*, 109-214. Amsterdam: John Benjamins Publishing Company.
- Zavala, R. (2000). Multiple Classifiers Systems in Akatek (Mayan). In G. Senft (Ed.), *Systems of nominal classification*, 114-146. Cambridge: Cambridge University Press.