The Development of Classifiers in the Tukanoan Family

Wiegertjes, Jurriaan

The Development of Classifiers in the Tukanoan Family

Jurriaan Wiegertjes

Leiden

31-12-2021

MA THESIS

"The Development of Classifiers in the Tukanoan Family"

By

Jurriaan Wiegertjes, BA

Submitted for the partial satisfaction of the requirements for the degree of

MASTER OF ARTS

at

LEIDEN UNIVERSITY

in the program

LINGUISTICS (RESEARCH)

Written under the supervision of:

Martine Bruil

Thiago Costa Chacon

Rik van Gijn

Second corrector:

Konrad Rybka

Abstract

All Tukanoan languages have gender markers and classifiers, and both can be reconstructed to the proto-language to some degree. In this thesis, I provide a reconstruction of the development of the classifier system in the Tukanoan family, where I argue that it developed out of the older gender system, morpho-syntactically, but that many of the synchronically found classifiers can morphologically be analysed as grammaticalized nouns.

My arguments for this, as elaborated in this thesis, are as follows: i) all Tukanoan languages have similar gender markers which can probably be reconstructed for Proto-Tukanoan (Chacon 2021; in prep.); ii) the gender markers seem to have undergone grammaticalization at an early stage in the family; iii) many classifiers in the family are language-internal developments or can only be reconstructed for a sub-branch; iv) a few classifiers are widely found in the family and can be reconstructed for the proto-language, but these seem to be developments of either Proto-Tukanoan gender markers or originally complex forms. I analyse these complex forms as consisting of a gender marker in combination with some other marker.

I furthermore provide an analysis of the development of the Proto-Tukanoan gender system, based on Chacon (2021, in prep), where I suggest that some of its morphological material may indicate borrowings from an Arawakan source.

Lastly, by contributing to the reconstruction of the Proto-Tukanoan classifier system, this thesis may contribute to a reconstruction of classifiers in the wider area, as classifiers are a pervasive feature in non-Tukanoan languages as well, where the relatively gender-like morpho-syntactic characteristics of classifiers have been the subject of much discussion (e.g. Payne 1987; Aikhenvald 2000a, 10; Grinevald 2000, 81-82, 87).

Contents

Acknowledgements

This thesis could not have been written without the guidance of my three great supervisors Martine Bruil, Thiago Costa Chacon, and Rik van Gijn. Their feedback, suggestions, and support over the last year have been of monumental value to me, my thesis grew on the fruits of their intellectual labours, and I am grateful for the fact that they have made time to comment on my many early drafts. I feel honoured and am grateful to have had the chance to work with them. Especially Thiago Chacon, who decided to support me even though he is not directly connected to Leiden University deserves a tremendous acknowledgement for his many helpful suggestions about the development of the Tukanoan classifiers, and for sharing his unfinished work with me, without which this thesis would not have been nearly as interesting.

I would also like to thank Leonardo Arias Alvis, Nick Emlen, Nora Julmi, Sietze Norder, and all the other people of the SAPPHIRE project for their inspiring conversations and presentations. I have learned much from them.

In the early days of this thesis, when I still thought it would be about Ayoreo, I have received a lot of advice from Mily Crevels and Paz Gonzalez Gonzalez, and I am grateful to Luca Ciucci, Amadeus Benz, and Santiago Durante for making time to speak with me and answer my questions.

I am grateful to Ivo Boers, Clemens Mayer, Rogier te Paske, and Cid Swanenvleugel, to my parents Suzanne Krijgsveld and Jeroen Wiegertjes, to Max Wiegertjes, to Julia Carvalho dos Santos, and to all my teachers at Leiden University who helped me along countless times, and with whom I have had stimulating discussions about my thesis and related topics.

This, and all my future work benefits from the tireless contributions of all the Indigenous people who have taught and continue to teach about their languages to a wider audience. Especially the Ecuadorian Siona people of Sototsiaya are a continuous source of inspiration, and their hospitality, patience, and knowledge are incredibly valuable to me.

I wish to extend my thanks to Konrad Rybka for fulfilling the function of second corrector on this thesis, and for being so understanding when I requested a later deadline than the one I had originally suggested.

List of figures and tables

Figures

Tables

A note on examples

The examples in this thesis follow a five-line or four-line format. The first line states the language from which the example hails, and to which linguistic family it belongs, or in the case of the Tukanoan family, whether it belongs to the Western, or to the Eastern branch. The second line gives the utterance using the conventional orthography of a language in italics. Some sources do not provide a line using a conventional orthography, in which case I have not provided one either. In the third line, words are divided into morphemes and written in an orthography that is closer to a phonemic transcription. The fourth line provides the morpheme-by-morpheme glosses of the forms in the second line, and the fifth line provides a free translation into English (i).

(i) ECUADORIAN SIONA (WESTERN TUKANOAN)
yë'ë baquë huaja'ë se̲'se
jɨ'ɨ ba-kɨ wa-ha'ɨ sẽ'se
1SG spouse-CL:ANIM.M kill-ASS.3M.PST wild.boar
"My husband killed a pig." [FNSJ00101]

The free translation is followed by a source reference in the right bottom corner. In (i), the reference is a code referring to the author's own fieldnotes. In the remainder of this thesis, wherever, a reference is given in straight brackets […], I refer to an audio file in Bruil (2012). All other examples come from glossed sources with page numbers to which I refer in rounded brackets (…).

Proto-Tukanoan is reconstructed with a number of laryngealized obstruents (Chacon 2014), which are conventionally written with an apostrophe (') following the laryngealized consonant, and I employ the same convention here. In other languages, I denote laryngealization with the symbol ˷ underneath the laryngealized consonant, or with a voiced consonant, such as in (i) above, where <b> refers to a laryngealized voiceless bilabial stop (see also Bruil 2014, 93–95).

Abbreviations

1, 2, 3	First, second, third person
1D, 2D, 3D	One, two, three dimensional
A	Agent
ADV	Adverbial
ANA	Anaphoric
ANIM	Animate
APL	Applicative
ART	Article
ASS	Assertive mood
B	Bare form
C	Consonant
CAUS	Causative
CL	Classifier
COP	Copula
DE	Direct evidential
DECL	Declarative mood
DEIC	Deictic
DEM	Demonstrative
DEP	Dependend
DIM	Diminutive
DIST	Distal
DR	Derivational
DV	Dummy vowel
EV	Evidential
EX	Existential verb
F	Feminine
FUT	Future tense
GEN	Generic
HON	Honorific
HUM	Human
IMP	Imperative mood
IND	Independent
IRR	Irrealis mood
LINK	Linker
LOC	Locative case
M	Masculine
MASS	Mass noun
MD	Discourse marker
MISC	Miscelaneous
N	Neuter
N2/3SG	Non-first/second person singular
NASS	Non-assertive mood
NF	Non-feminine
NLZ	Nominalizer
NPL	Non-plural
NSG	Non-singular
O	Object
PET	Proto-Eastern-Tukanoan
PFV	Perfective aspect
PL	Plural number
POSS	Possessive marker
PRO	Pronominal
PRS	Present tense
PRX	Proximal
PST	Past tense
PT	Proto-Tukanoan
PWT	Proto-Western-Tukanoan
QM	Question marker
Ref.	Reference
REL	Relativizer
RFM	Relational form marker
REP	Reportative evidential
S	Subject
SG	Singular number
SIM	Simultaneous
SS	Same subject
TAM	Tense-aspect-mood
TOP	Topic
V	Vowel
VIS	Visual evidential

-	Morpheme-boundary
=	Clitic-boundary
:	Specification
*	Reconstructed form
#	Ungrammatical or unattested form
[	Word-initial
]	Word-final
>	Results in…
<	Is the result of…
<…>	Orthographically
[…]	Phonetically
/…/	Phonemically

1 Introduction

We conceptualise and classify our knowledge in different relevant categories, which allows us to make generalisations about overarching features of things and beings. Linguistically, such overarching features may be highlighted by a number of means, one of which is the use of nominal classifiers.

Nominal classifiers are a pervasive feature in many language families in Western Amazonia, such as Arawakan, Tukanoan, Witotoan, Boran, and Peba-Yaguan (e.g. Payne 1987; Aikhenvald 2000a, 10; 2012, 286–303; Farmer 2015, 125-130; Krasnoukhova 2012, 193–218). These languages have been shown to challenge proposed typologies of noun categorization systems (e.g. Payne 1987; Aikhenvald 2000a, 10; Grinevald 2000, 81-82, 87). Moreover, Payne (ibid.) argues that Tukanoan languages differ from proposed typologies of classifier systems in a similar way as some other Western Amazonian languages (viz. Zaparoan, Paba-Yaguan, Chahuita, Chahuapanan, Boran, some Witotoan and some Arawakan languages). Their most striking features are the co-occurrence of multiple systems of noun categorization in a single language and the agreement function of classifiers (*inter alia* Payne 1987; Grinevald and Seifart 2004, 259–61; Seifart and Payne 2007, 383–4; Aikhenvald 2012, 279–80). The pervasiveness of *noun classifiers* –i.e. classifiers marked on nominal heads– is another striking feature of the area (see Payne 1987; Derbyshire and Payne 1990; Aikhenvald 2012, 288–90), as this is a rarity worldwide (Kilarski 2013, 295–97). According to Payne (ibid.), the fact that the Northwest Amazonian classification systems share a number of typologically uncommon features suggests past contact between these families. In order to be able to test this claim, it is prerequisite to establish to what depth the currently attested classifier systems in the individual families can be reconstructed.

In this thesis, I contribute to that goal by comparing nominal classifiers in Tukanoan languages in order to provide an answer to the question: To what degree can proto-forms of noun categorization markers be reconstructed in the Tukanoan family, and how did the systems of noun categorization in the family develop?

I argue that some classifiers can be reconstructed to Proto-Tukanoan (henceforth *PT*), but that these have developed after the PT gender system was in place, which in turn seems to have developed under contact with Arawakan languages (Chacon 2017; 2021; in prep.). Based on the abovementioned typologically problematic features, Northwest Amazonian systems have been analysed as a "mixture" of noun class/gender and classifier systems (e.g. Payne 1987; Derbyshire and Payne 1990), or as incipient noun class/gender systems, at an earlier stage of grammaticalization than typical noun class/gender systems such as those found in the Niger-Congo family (e.g. Grinevald and Seifart 2004).

Although it is plausible and often repeated that nominal classifiers can develop into noun class/gender systems (e.g. Dixon 1982, 171–73; Corbett 1991, 139–41; Aikhenvald 2000a, 372–73), in this thesis I argue that the Tukanoan classifier systems are an expansion on an older gender system. My arguments for this claim are as follows:

i) all Tukanoan languages have similar gender markers which can probably be reconstructed for PT;
ii) the gender markers demonstrate characteristics of early grammaticalization;
iii) many classifiers in the family are language-internal developments or can only be reconstructed for a subbranch;

iv) a few classifiers are widely found in the family and can be reconstructed for the proto-language, but these seem to be developments of either PT gender markers or originally complex morphemes consisting of a gender marker in combination with some other marker.

The hypothesis that the Tukanoan classifiers are syntactically an expansion on the original gender system may furthermore be a step toward understanding the challenges they pose for typologies of classifiers, since nominal agreement is a characteristic typically associated with gender markers.

Section 2 discusses noun categorization systems, terminological and typological issues thereof, and their diachronic grammaticalization. Section 3 discusses noun categorization systems in the Tukanoan family, and Section 4 introduces the methodology used in this thesis. Section 5 and 6 discuss the development of respectively gender and classifiers in the Tukanoan family. Section 7 draws conclusions and contains some final remarks. The remainder of this section discusses the internal classification of the Tukanoan family (Section 1.1), and draws attention to the relevance of the Tukanoan family in the context of the wider area, and the abovementioned issues with noun categorization systems found there.

1.1 The Tukanoan family

The Tukanoan family comprises roughly 27 living languages (Ramirez 1997, 15; Barnes 1999), eight of which are now probably extinct (Chacon 2014). The family can be subdivided into two main branches: A Western Tukanoan (henceforth *WT*) and an Eastern Tukanoan (henceforth *ET*) branch (Mason 1950, 258–60; Chacon 2014; Chacon and List 2015). Figure 1 presents a simplified Tukanoan family tree in order to indicate the relative distance between languages, based on figures found in Vallejos (2013, 72), Bruil (2014, 8), and Chacon (2014, 282), and the analyses found in Barnes (1999, 209) and Chacon and Michael (2018, 64).

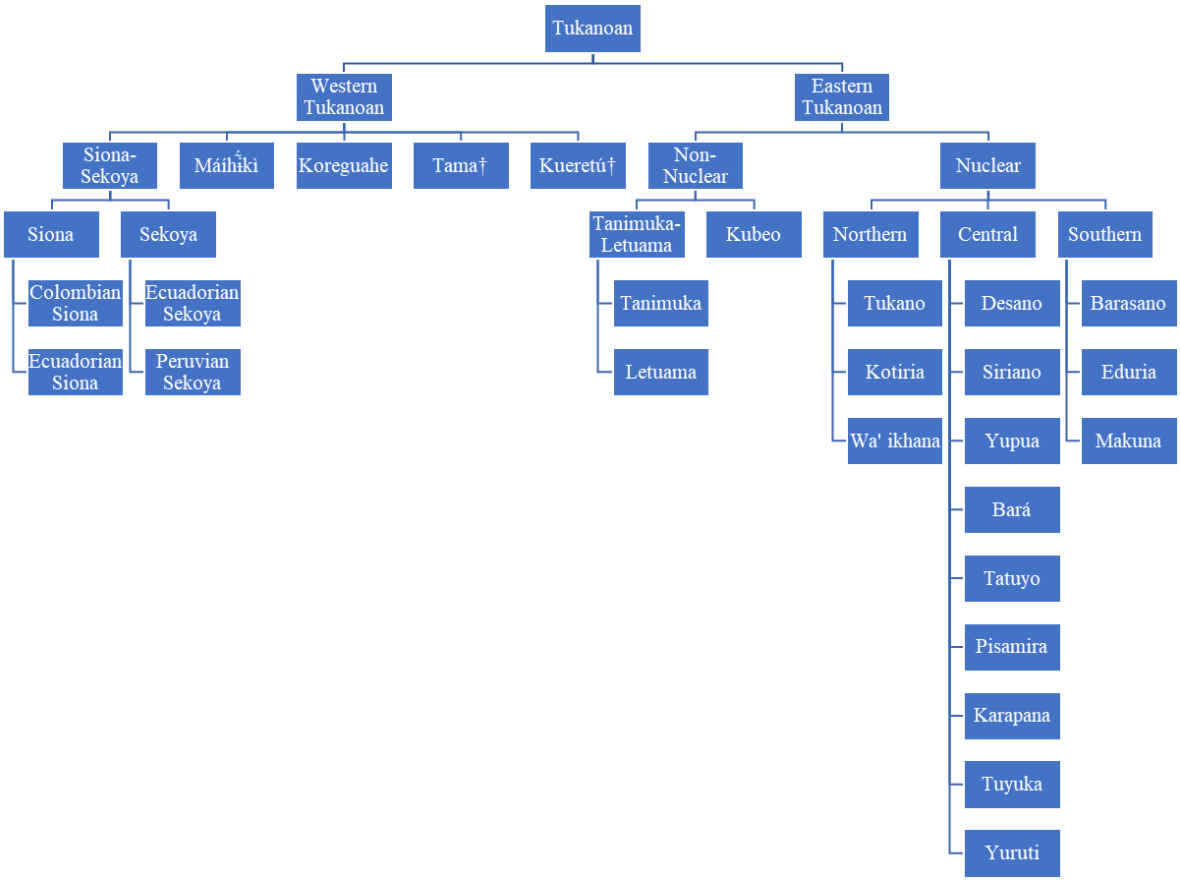


Figure 1: The Tukanoan family

Languages pertaining to the Western branch are Colombian Siona, Ecuadorian Siona, Ecuadorian Sekoya, Peruvian Sekoya, Máíhíkì (also known as Orejón), Koreguaje, as well as now extinct Tama, and Kueretú. These languages and are spoken in Colombia, Ecuador and Peru. The two Siona and two Sekoya varieties listed above are quite similar, demonstrate relatively much mutual intelligibility, and are said to form a dialect continuum (Vallejos 2013; Bruil 2014, 11–12).

ET languages are spoken in the Vaupés area in Colombia and Brazil. The Eastern branch can be subdivided into a Nuclear ET branch and a Non-nuclear ET branch, where the Nuclear ET branch consists of Piratapuyo, Tukano, Kotiria (also known as Wanano), Bará (also known as Waimajã or Northern Barasano), Karapana, Desano, Siriano, Tatuyo, Tuyuka, Yurutí, Barasana (also known as Taiwano), and Makuna, and the Non-nuclear ET branch consists of Tanimuka, Letuama (also known as Retuarã), and Kubeo (Barnes 1999, 209; Chacon and Michael 2018, 64).[1] Tanimuka and Letuama are very similar, but do demonstrate a number of differences including some pertaining to the noun categorization system (Strom 1992; Eraso 2015).

In the remainder of this thesis, I refer to languages by their commonly used abbreviations, which consist of the first three letters of their names, with the exception of Barasana (BSA), in order to avoid confusion with Bará (BAR). The different Tukanoan languages are summarized

[1] The Non-nuclear ET branch has also been analysed as a separate "Middle," or "Central Tukanoan" branch, on par with the WT and ET branches (e.g. Waltz and Wheeler 1972; Barnes 1999; Ramirez 1997). I will follow the Non-nuclear analysis here.

in Table 1 below, where I list their branch and the sources I have used for this thesis. Regarding the sub-branches listed in Table 1, I follow Barnes (1999), Chacon (2014), and Chacon and Michael (2018).

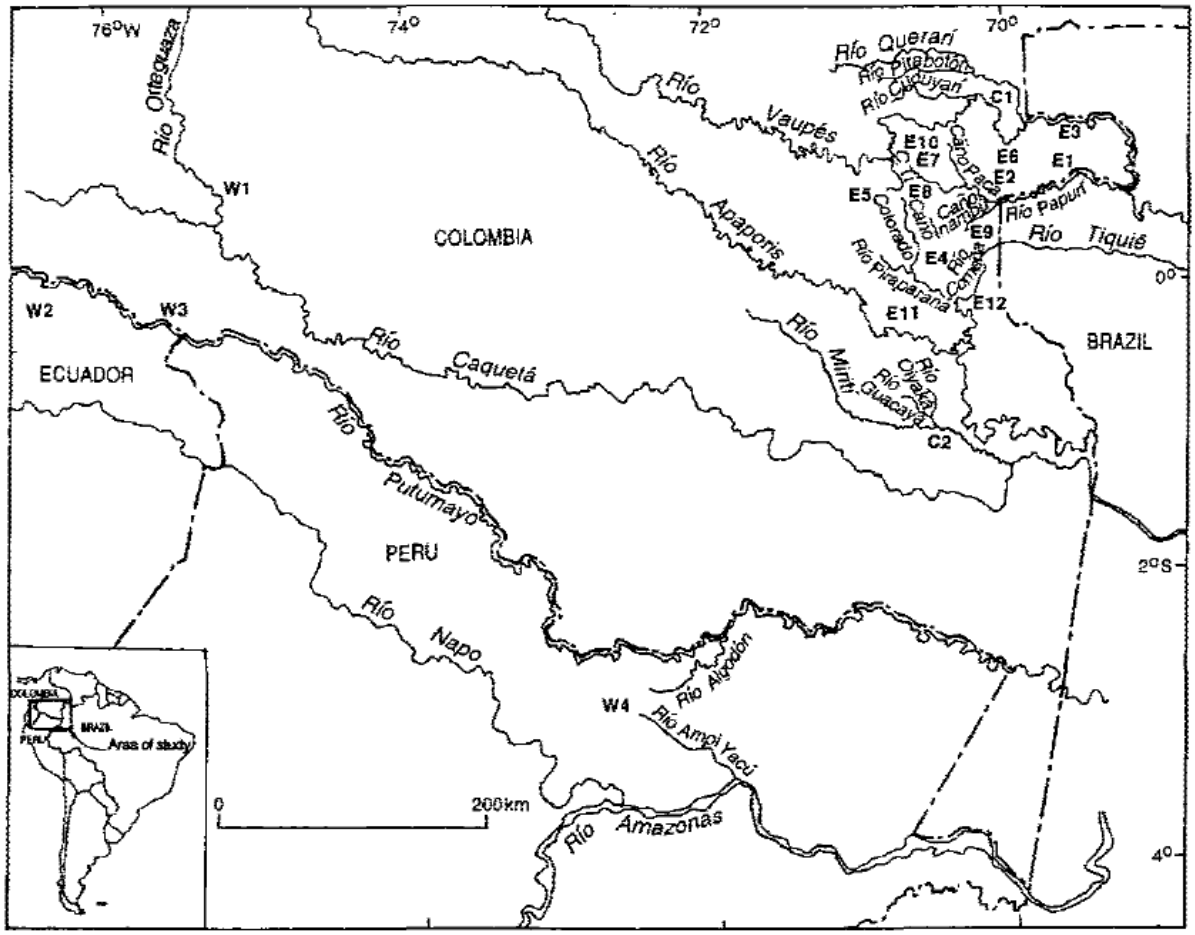
Table 1: The Tukanoan languages

Language	**Abbreviation**	**Sub-branch**	**Branch**	**Sources**
Colombian Siona	C. SIO	Northern WT	WT	(Wheeler 1970; 1987; 2000)
Ecuadorian Siona	E. SIO			(Criollo Quintero 2011; Bruil 2012; 2014; 2016; 2018)
Ecuadorian Sekoya	E. SEK			(Johnson and Levinsohn 1990; Piaguaje, Piaguaje, Johnson, and Johnson 1992)
Peruvian Sekoya	P. SEK			(Vallejos 2013; 2021a; 2021b)
Koreguaje	KOR			(Cook and Crisswell 1993; Cook and Gralow 2001)
Tama	TAM			(Creveaux, Sagot, and Adam 1882)
Máíhɨ́kì	MAI	Southern WT		(Velie and Velie 1981; Farmer 2015)
Tanimuka	TAN	Non-nuclear ET	ET	(Eraso 2015)
Letuama	LET			(Strom 1992)
Kubeo	KUB			(Morse and Maxwell 1999; Morse, Salser, and de Salser 1999; Chacon 2012)
Tukano	TUK	Nuclear ET		(Sorensen 1969; West 1980; Ramirez 1997)
Kotiria	KOT			(Waltz 2007; Stenzel 2013)
Wa'ikhana	WA'I			(Balykova 2019; 2021)
Desano	DES			(Miller 1999; Alemán, López, and Miller 2000; de Lima Silva 2012)
Tatuyo	TAT			(Bostrom 1998; Gomez-Imbert 1982; 2007)
Pisamira	PIS			(González de Pérez 2000; González Muñoz 2016; Rodríguez Preciado 2018)
Karapana	KAR			(Metzger 1981)
Tuyuka	TUY			(Barnes 1990; Barnes and Malone 2000)
Barasana	BSA			(Jones and Jones 1991)
Makuna	MAK			(Smothermon and Smothermon 1993; Smothermon, Smothermon, and Frank 1995)

For ease of exposition, I do not refer to the specific sources I consulted in each individual table below, and instead list these in Appendix A. In addition to the sources listed in Table 1 and

Appendix A, the comparative lexical overview by Huber and Reed (1992) has been invaluable for this thesis.[2] The geographic locations of the Tukanoan groups are presented in Map 1.

Map 1: The geopgraphic locations of the Tukanoan groups and their neighbours, taken from Barnes (1999, 208–09).



Key: W1–KOR; W2–SEK; W3–SIO; W4–MAI; C1–KUB; C2–TAN/LET; E1–PIR; E2–TUK; E3–KOT; E4–BAR; E5–KAR; E6–DES; E7–SIR; E8–TAT; E9–TUY; E10–YUR; E11–BSA/EDU; E12–MAK.

1.2 Why Tukanoan?

As mentioned before, the central question of this thesis is part of a larger question regarding a number of linguistic families of the Northwest Amazon. The Tukanoan family is of special interest in this light, because it is not only one of the larger families of the area (e.g. Aikhenvald 2012), it is also central to the area, and classifiers are omnipresent throughout the family. All Tukanoan languages have some form of nominal classifiers and gender markers, which makes the reconstruction of their noun categorization systems feasible and of particular importance in light of the wider questions introduced above.

[2] I thank Thiago Chacon for sharing his digitalization of this volume with me.

2 Noun categorization systems

Before discussing noun categorization in the Tukanoan family, it is necessary to provide a brief overview of core terminology, typology and theoretical assumptions that form the basis for this thesis.

Section 2.1 discusses some terminological and typological issues, with the main goal of differentiating between noun class on the one hand, and classifiers on the other, where I show that there is no rigid dichotomy, and employ a definition that is based on the degree of synchronic *grammaticalization* of the system of noun categorization at hand. Section 2.2 discusses the development of noun categorization systems with special focus on the process of diachronic grammaticalization and on the notion that classifiers may arise from nouns.

2.1 Some terminological and typological issues

This section discusses some terminological and typological issues with noun categorization systems. Section 2.1.1 provides a definition of classifiers, noun classes and gender, and Section 2.1.2 and 2.1.3 discuss subtypes of respectively noun class, and classifier systems.

2.1.1 Noun classes and classifiers: Distinguishing noun categorization systems

Systems of noun categorization are categorized in various ways throughout the literature. Dixon (1986) subsumes both *gender* and *noun class* systems under the term *noun classes*, and opposes them to *noun classification* (e.g. the numeral classifiers found in East and Southeast Asian languages).

In Dixon's (ibid.) categorization, *noun classes* form an obligatory grammatical system that groups all nouns in a language into a relatively small number of classes, ranging from two (e.g. the feminine-masculine distinction in French) to around twenty (e.g. Bantu noun class systems). Typically, nouns belong to just one class, and other words in the clause often display inflectional agreement with the class of the head noun.

Noun classification systems, on the other hand, consist of large sets, with at least around twenty classifiers, and more than a hundred being common. Typically, nouns will be able to occur with more than one classifier, and some nouns may not be able to be classified, as opposed to *noun classes* which categorize all nouns in a language rather rigidly (id., 106).[3]

[3] According to Dixon (1986), classifiers are always free forms that occur in the same noun phrase as the noun they classify, but always in a separate constituent (e.g. a modifer phrase). In light of new data, especially from South American indigenous languages, these claims have had to be rejected. Classifiers may attach to other morphophonological words (i), and they may occur directly on a nominal head (ii), or in constituents outside of the noun phrase (iii).

(i) EIGHTEENTH CENTURY CHOLÓN (HIBITO-CHOLON)

*ana-čup-**am***	*me-n-eštek*	*Ø-m-a-toŋ*
how.many-CL:PORTABLE-QM	2SG.POSS-RFM-cloth	3SG.S-2SG.O-APL-be.there

'How many dresses do you have?' (Alexander-Bakkerus 2005, 181)

(ii) JACALTEC (MAYAN)

xul	***naj***	*Pel*	*b'oj*	***ya'***	*malin*
came	CL:M.NON.KIN	Peter	with	CL:HUM.RESPECTED	Mary

'Peter came with Mary.' (Craig 1986b, 264)

(iii) ITONAMA (ISOLATE)

*yornoni-**yeta'**-na'-ka*	*ni-mariya*	*pi-ni-'e*	*pi-ni-ku*
leave-ANIM.STANDING.SG-N-F.SG	HON-María	2SG.F-REL-son	3SG.F-REL-house

'María left her son at home' (Crevels 2012, 269)

Increasingly, a strict dichotomy between noun classes and classifiers is being rejected and it may be better to speak of 'prototypical' or 'canonical' types with the abovementioned characteristics, rather than theoretically independent types (e.g. Fedden and Corbett 2017). Therefore, I will not make a strict distinction between the two here, and define prototypical noun class/gender systems as any grammatical morphological system that assigns nouns to classes on a lexical basis, and prototypical nominal classifier systems as any grammatical, pragmatic, morphological system that assigns nouns to classes on the basis of certain selected characteristics of their referents.

It is important to pay attention to the notion of *pragmatic* and *referent*, i.e. something, someone, or some event that is extralinguistic, but to which a linguistic element refers. These notions seem to be central to the distinction between class and classifier in the sense that noun class is typically *lexically determined*, such that a given noun is consistently assigned a particular gender, rather independently of the specific referent it denotes, whereas nominal classifiers are typically *pragmatically determined*, such that a given noun is categorized according to what is relevant for a specific referent or a specific situation (see also Contini-Morava and Kilarski 2013, 291; Kilarski 2013). This distinction is not necessarily obvious, but it is possible to distinguish the two empirically (see for example Corbett et al. (2020) and Franjieh, Corbett, and Grandison (2020). These works show a connection between the size of a categorization system and the rigidity in assigning categories to referents, where smaller systems correlate with more rigid assignment of classifiers to referents, both across speakers and across pragmatic situations, and larger systems correlate with more flexible assignment of classifiers to referents, across speakers and pragmatic situations. This suggests that smaller systems have a more abstract categorization which is not assigned at the moment of speaking, but follows a lexically determined categorization, whereas larger systems have a more *ad hoc* categorization, which is assigned at the moment of speaking.

The difference is scalar (Corbett et al. 2020; Franjieh, Corbett, and Grandison 2020), and probably involves tendencies rather than universals, but it is easy to imagine why these categories would correlate: The semantic oppositions between different categories necessarily become more abstract as the number of categories reduces.

When we take the lexical versus pragmatic opposition as indicative in distinguishing between noun class and classifier, some 'grey areas' remain, where systems with fairly prototypical gender oppositions also categorizes some nouns on a non-lexical basis. For example Spanish (Indo-European) has a masculine-feminine distinction, but some nouns may take either gender, depending of the gender on the referent, e.g. *la/el futbolista* 'the football player', where the feminine article *la* is used with feminine football players, and masculine *el* for masculine ones.[4]

For the purpose of this thesis, it is sufficient to conclude that Spanish has a relatively typical gender system that contains some instances of classifier-like characteristics. In other words, I acknowledge that gender, noun class, and nominal classifiers are not completely separate systems, but I will use the terms to refer to systems that approach either of the abovementioned prototypes –i.e. *noun class* or *gender* for predominantly lexical systems, and *classifier* for predominantly pragmatic systems– providing detailed descriptions where necessary. I use the term *noun categorization device* as an umbrella for the three systems (see

See Aikhenvald (2000a, 10-12) for an overview of some publications that have challenged these suggestions.

[4] One might propose two homophonous nouns *futbolista* with different genders, or point out that this is not a fully productive system, as the same does not hold true for all human nouns such as *la/#el persona* 'the person', the article of which is always feminine. I will refrain from such theorizing here.

also Aikhenvald 2000a). The differences between classifiers and noun classes as outlined above are illustrated in Table 2, although I emphasize that the difference is scalar and a matter of tendency rather than universality.

Table 2: Differences between classifiers and noun classes

	Classifiers	**Gender**
Assignment	Pragmatic	Lexical
Flexibility	Flexible	Rigid
Semantic salience	More	Less
Size of class	Larger	Smaller

Above, we have developed a definition for the terms *noun class* and *classifiers*, but there are many subtypes of both. Types of noun class systems are discussed in 2.1.2, and types of classifier systems in 2.1.3.

2.1.2 Types of noun class systems

As mentioned above, Dixon (1986) subsumes both *noun class* and *gender* under *noun classes*, but the terms are nevertheless associated with slightly different prototypical systems, and different grammatical traditions (e.g. Corbett 1991, 147; Aikhenvald 2000a, 19–20).

Gender generally involves relatively smaller systems of two or three categories, where the referent's gender and/or animacy of referents plays some role (e.g. Aikhenvald 2000a, 19; Chacon and Sateles 2019, 21). Therefore, I use the term *gender* here as referring to any small noun class system that typically involves a semantic distinction of animates' genders, and possibly an inanimate –or neuter– category, but which may also consist of only an animate-inanimate distinction.[5]

Noun class systems, on the other hand, generally involve somewhat larger systems, distinguishing up to around twenty categories, where referents can be more specifically distinguished, e.g. there is a class of nouns predominantly referring to artifacts in Kol (Isolate) (Stebbins, Evans, and Terril 2018, 795–56). Furthermore, whereas gender systems can sometimes said to be devoid of all meaning, noun class systems do have meaning (Allan 1977).

In light of the definition of *gender* as a small and abstract system, and *noun class* as a somewhat larger, and somewhat less abstract system, prototypical gender systems form one end of the spectrum, and prototypical classifier systems the other, such that prototypical noun class systems fall somewhere in between. Although noun class systems are still lexically determined, they typically distinguish more semantic categories than gender, which places them somewhat more toward the pragmatic side of the scale than gender systems.[6]

2.1.3 Types of classifier systems

Within the category of classifiers, there are many different types and subtypes. Classifiers may be distinguished by their morphosyntactic locus (see Aikhenvald 2000a), which also correlates with semantic content. Classifiers may also be distinguished by their grammatical function, i.e.

[5] Noun classes may be assigned on the basis of: i) semantics, i.e. the Ojibwe animate-inanimate distinction (Corbett 1991, 20–21); ii) associated conceptualisations, i.e. Kala Lagaw Ya *kisay* 'moon' is masculine, because it has traditionally been seen as 'grandfather' (Bani 1987, 201); iii) morphology, i.e. German nouns marked with the diminutive *-chen* are normally neuter (Corbett 1991, 50)); iv) phonology, i.e. all Hausa non-sex-differentiable nouns ending in *-aa* are feminine (Aikhenvald 2000a, 25); v) or a combination of the above.

[6] The fact that noun class systems typically have some meaning is also the reason for authors such as Allan (1977) to analyse *classifiers* and *noun classes* as a single category.

whether they participate in agreement or derive nouns. Below, I discus classifiers by morphosyntactic locus (Section 2.1.3.1), and by their grammatical function (Section 2.1.3.2). In this thesis, I use a synthesis of these two ways of distinguishing classifier types, where I define classifier types on the basis of their morphosyntactic locus, but only when the classifier fulfils an agreement function. We will come back to this synthesis in Section 2.1.3.2.

2.1.3.1 Morphosyntactic loci of classifiers

Classifiers may be found on a number of morphosyntactic loci, e.g. with nouns (1a), numerals (1b), possessive pronouns (1c), verbs (1d), or demonstratives (1e). For clarity, it must be noted that I reserve the term *noun classifiers* for those classifiers that occur directly with nouns, as opposed to *nominal classifiers*, or simply *classifiers*, which I use to refer to classifiers regardless of the locus of marking.

(1) a. POPOLOC (OTOMANGUEAN)
su³-ca³ʔ
CL:FLOWER-orchid
'orchid flower' (Veerman-Leichsenring 2004, 421)

b. NEPALI (INDO-EUROPEAN)
pac ***jana*** *mali*
five CL:HUMAN gardener
'five gardeners' (Allassonnière-Tang and Kilarski 2020, 125)

c. AYOREO (ZAMUCOAN)
*g-**achidi*** *tamoco*
3.POSS-CL:PET dog
'his/her/their dog' (Bertinetto 2014, 379)

d. MOTUNA (SOUTH BOUGAINVILLE)
ong *topo* *inak-i-heeto-no-**uru***
DEM.M well look.after-3O.3A-FUT-LINK-CL:HUM
'This is the one (lit. 'human' you will look after well.' (Onishi 1994, 176)

e. OMAHA-PONCA (SIOUAN)
waʔú-ʐĩga ***ðīkhe***
woman-old the.DEIC,CL:SG.SITTING.ANIM
'the (sitting) old woman' (Rankin 2004, 212)

A language may display classifiers in more than one morphosyntactic locus as well. When different morphological sets of classifiers coexist in the same language, one may speak of a system with *different classifiers* (Aikhenvald 2000a, 184–85), e.g. in Itonama (isolate), which has two different morphological sets of classifiers: Seventeen verbal and deictic classifiers, and eight numeral classifiers (2).[7]

[7] It is interesting that deictics pattern with verbs in this respect, rather than with numerals, since one would expect deictics to pattern with the other nominal category that takes classifiers in the language. One possible reason for this phenomenon is that deictics in Itonama are more similar to verbs than to nouns (Crevels 2012, 253).

(2) ITONAMA (ISOLATE)

nik'abï	*o-si-**lo***	***ni**-chïpa*	*uwu*	*wa'ihna*
DEM.ADV.DIST	DV-EX-CL:WINDING	CL:WINDING-two	river	MD
*o-si-**du***		***chas**-k'a'ne*	*iskuwela*	
DV-EX-OVAL.CONTAINER.SG		CL.OVAL.CONTAINER-one	school	

'Over there are two rivers and one school' (Crevels 2012, 254)

The example in (2) contains two classifiers from the deictic and verbal set, and two from the numeral set. Although there is some overlap between the meanings denoted by the classifiers in these functions, their forms are rather different, and deictic and verbal classifiers suffix to the root, whereas the numeral classifiers are prefixes.

When the same, or almost the same morphological set of classifiers is employed in more than one morphosyntactic locus, one may speak of a *multiple classifier system* (Aikhenvald 2000a, 204–28; 2000b, 94; Zavala 2000). In other words, the deictic and verbal set of classifier suffixes are analysed as separate, simply because they attach to different parts of speech. Multiple classifier systems are a typical feature in South American languages that have classifiers, and in particular those of the Northwest Amazon region (Derbyshire and Payne 1990, 246; Aikhenvald 2000b, 94). One such a language is Tariana (Arawakan), which uses noun class markers, noun classifiers, and numeral, possessive, demonstrative, article and verb classifiers (Aikhenvald 2000b). Their respective forms are almost identical across these functions, but their syntactic, semantic and pragmatic uses are different. In the example below, the very same form *-na* 'CL:VERTICAL' is found on an article and a verb (3a), a demonstrative and a noun (3b).

(3) TARIANA (ARAWAK)

a.

*diha-**na**-pe*	*waha*	*wa-ka-ni-**na**-pe*	
ART-CL:VERTICAL-PL	we	1PL-see-TOP.ADV-CL:VERTICAL-PL	

'The high ones (hills) which we see' (Aikhenvald 2000b, 105)

b.

*hane-**na***	*heku-**na***	*ithani-ka*	*di-swa*
that-CL:VERTICAL	tree-CL:VERTICAL	near-DECL	3SG.NF-stay

'It (the baby) is under that very tree (over there)' (id., 103)

As opposed to article, demonstrative, and noun classifiers, the verb classifier is obligatory and does not make a rigid distinction in animacy and as opposed to demonstrative, verb, and noun classifiers, article classifiers have a special class for animate plurals. In this way, classifiers have subtly different morphosyntactic behaviours in each locus (Aikhenvald 2000b, 109).

As exemplified by Itonama, languages with classifiers in more than one morphosyntactic locus may have different semantic and morphological inventories. As Tariana shows, even though the inventories may be very similar, there may still be differences between the grammatical properties of classifiers of different morphosyntactic loci.

Classifiers with different morphosyntactic loci can thus be analysed as separate grammatical categories. What is more, as Aikhenvald (2000b) shows for Tariana, noun class systems and nominal classifiers may co-occur in the same language, and their morphological forms may be similar between noun class and classifier as well (see also Aikhenvald 2000a, 185–87, 223–24, 230–41). As we will see in Section 3, this phenomenon is also found throughout the Tukanoan family.

2.1.3.2 Grammatical functions of classifiers

The functions of classifiers may differ per language, but in the Tukanoan family, one core function of classifiers on modifiers is agreement with the head noun. Another function of Tukanoan classifiers is as noun-formatives, deriving nouns from other nouns or other parts of speech. Since Tukanoan classifiers may fulfil nominalizing functions, I only distinguish multiple classifier types when they fulfil an agreement function with the head noun, and not when deriving nouns from other parts of speech.

As a result of the definitions in Section 2.1.1, classifiers fulfil relatively pragmatical functions, whereas noun classes are more grammatical. Since agreement is a relatively grammatical feature, the fact that many Northwest Amazonian systems involve agreement has been suggested to be the result of the degree to which the noun categorization devices in a language have undergone grammaticalization, such that they are now similar to noun classes, or may even be analysed as noun class systems in the early stages of grammaticalization (Grinevald 2000, 82–83; Grinevald and Seifart 2004). However, it must be noted that the noun categorization devices in such languages can be quite stable and coherent within individual languages (e.g. Seifart and Payne 2007, 384; Petersen de Piñeros 2007). Grammaticalization processes associated with classifiers are further discussed in a diachronic perspective in Section 2.2, which will further clarify the division between classifiers and noun classes and the diachronic interpretation of classifiers in general, which will in turn help identify the different stages of development of the Tukanoan classifier systems.

Whether classifiers may be analysed as agreement markers cannot always be answered straightforwardly. As pointed out by Contini-Morava (2021), most Northwest Amazonian classifier systems do not meet Corbett and Fedden's (2016, 498–99) criteria of *canonical agreement*, as summarized in Table 3 below, where the term *conditions* refers to factors which have an effect on agreement, but which are not directly reflected in the agreement as features such as number, gender, etc. (Corbett 2009, 343). It is also important to note that the term *canonical* refers to logical, rather than empirical, concepts, which entails that completely canonical agreement may be rare in the languages of the world, and indeed, does necessarily occur at all (ibid.).

Table 3: Criteria of canonical agreement (Corbett and Fedden 2016, 499)

controller:	is present, has overt expression of features, and is consistent in the agreements it takes, its part of speech is not relevant
target:	has bound expression of agreement, obligatory marking, doubling the marking of the noun, marking is regular, alliterative, productive; the target has a single controller and its part of speech is not relevant
domain:	agreement is asymmetric (e.g. the gender of the adjective depends on that of the noun), local (i.e. within the NP), and there are multiple domains
features:	lexical, matching values, not offering any choice in values
conditions:	no conditions

According to Contini-Morava (2021), Northwest Amazonian classifiers do not participate in canonical agreement as defined above, since:

i) Northwest Amazonian classifiers can, and often do, lack an overt controller, i.e. there is no obligatory head noun with which classifiers agree;
ii) agreement marking on the target, i.e. the dependent, may be optional;

iii) the features denoted by the agreement are to a large extent a matter of pragmatics, i.e. not exhaustively lexically determined;
iv) there may be pragmatic and syntactic conditions that regulate the occurrence of classifiers;
v) given that our definition of classifiers relies heavily on the notion that they are pragmatically determined, the criteria of *features* and *conditions* are obviously not met, and as Contini-Morava (2021) notes, the occurrence of classifiers on targets is to a large extent subject to pragmatic constraints as well.

Nevertheless, the fact that modifiers may take classifiers that denote some feature of the head noun they modify has often been taken as a leading criterium for authors on languages in the area, e.g. Aikhenvald (2007, 479), Bruil (2014, 137), and Chacon (2012, 238–42). For now, I will use the term of *nominal agreement* in a simplified way, as covariance between a head and a dependent modifier in a relatively systemic way.[8] Therefore, I disregard as agreement all instances of classifiers on nominal heads, regardless of whether they are derived or underived. This is a crucial distinction because in many Northwest Amazonian languages, classifiers may be employed to derive head nouns (see Section 3.1 for a discussion of this phenomenon in the Tukanoan family) (Krasnoukhova 2012, 209–16). In Murui (Witotoan) for example, numerals may bear classifiers when occurring independently (4a), but can take only the generic classifier *-je* when modifying a head noun (4b).

(4) MURUI (WITOTOAN)

a. *da-na*
one-CL:TREE
'one tree' (Petersen de Piñeros 2007, 394, 396)

b. *da-**je*** *rɨño* *da-**je*** *ɨ̈ɨ-ma*
one-CL:GEN woman one-CL:GEN male-M
'One woman and one man.' (ibid.)

Since there is no covariance between the heads *rɨño* 'woman' and *ɨ̈ɨma* 'man', and the modifiers *daje* 'one' in (4b), this example does not show agreement. Numerals may take nominal classifiers, such as *-na* 'CL:TREE' in (4a), but since there is no head with which it agrees, this example does not show agreement either.

Even in classifier systems where modifiers may take classifiers that show covariance with a head in a systemic way, one might argue that the modifier is in fact a derived noun in apposition to another noun, rather than a true modifier to a head noun (Farmer and Vallejos 2021), such as in MAI (5).

(5) MÁÍHH̃KÌ (WESTERN TUKANOAN)
kã-ɲaka kiu-ɲaka
DEM.DIST-CL:SHARP.OBJECT metal-CL:SHARP.OBJECT
'that nail', or 'that sharp thing, a nail', or 'that sharp thing, the metal sharp thing'
(Farmer and Vallejos 2021)

[8] This definition is based on Steele (1978, 610), who puts it: "The term agreement commonly refers to some systematic covariationce between a semantic or formal property of one element and a formal property of another (emphasis in original)." Note that classifiers, which may have meaning, can also entail a systematic covariance of semantic properties in both controller and target.

As mentioned in Farmer and Vallejos (2021), it is not possible to determine exactly which analysis is best, an ambiguity that may also have its effect on the use of the classifiers by speakers. In this study, I uniformly analyse such cases as agreement, but note that more nuanced analyses may be due.

2.1.3.3 Synthesis

As a consequence to the theoretical problems posed by the derivational nature of classifiers in many Northwest Amazonian languages, I discriminate between the different types of classifier on the basis of their morphosyntactic locus, and I distinguish consistently between classifiers that are found on modifiers, and those that are found on heads. In order to avoid confusion with regard to nominalizing functions of classifiers, I only differentiate different types of classifiers morphosyntactically when the word with which a classifier occurs fulfils a distinct syntactic function (cf. e.g. Aikhenvald 2000a). Therefore, in what follows, I do not analyse examples such as *ãmẽdigʉgã* 'small trunk' in (6) as adjectival or verb classifiers, but rather as a noun classifier, since this word fulfils the function of a head noun in this construction.

(6) DESANO (EASTERN TUKANOAN)

yʉhkʉgʉ ãmẽdigʉgã ãrã		
yʉkʉ-gʉ	**~abe-di-gʉ-~ga**	~adi-a
tree-CL:TRUNK	be.small-NLZ:INAN-CL:TRUNK-DIM	be-PFV

'The tree is (very) small.' (de Lima Silva 2012, 188)

2.2 Noun categorization systems in a diachronic perspective

Classifiers have both lexical and grammatical properties, and as such they are often called 'intermediately grammaticalized' (e.g. Craig 1986a; Grinevald 2002, 260–261).

Virtually all linguistic theories distinguish between *grammar* –or *function*– and *lexicon*, but the dichotomy is difficult to define in a completely satisfying way, and authors working on grammaticalization often conceptualise the difference between grammar and lexicon as a continuum (e.g. Langacker 1987, 18–19; Hopper and Traugott 1993, 6–7). Generally mentioned differences are: content; obligatoriness; and the openness of the class (e.g. Contini-Morava (2000) and references therein).

Firstly, lexicon typically refers to content, i.e. objects, actions, and features thereof, whereas grammar typically denotes the way those lexical concepts are related, and the relationship between them and the context in which an utterance is made. Secondly, grammar typically poses obligatory rules whereas the choice of lexicon is typically freer. This is not to say that speakers are free to choose whatever lexical units they like, but the choice is a matter of pragmatics. Grammar, is not entirely independent of pragmatics either,[9] but grammar can be judged correct or incorrect on just the basis of language-internal rules. Thirdly, grammar typically consists of closed classes whereas lexicon consists of open classes (e.g. Dixon 2010, Vol.1, 47–51).

On all three accounts, classifiers have both lexical and grammatical properties. In terms of content, classifiers refer to referents rather than words (e.g. Allan 1977; Senft 2000b, 23; Lucy 2000, 331; Contini-Morava and Kilarski 2013), and a single noun denoting a variety of

[9] It does present a relationship between the context of utterance and the proposition, e.g. by using a past tense to denote that an event took place before the moment of utterance. In this sense, there are pragmatic restraints on what grammatical forms may be used.

referents may be able to be combined with a variety of classifiers based on the specific pragmatically relevant feature, rather than a lexically determined feature of the noun itself (e.g. Kilarski 2013, 280–83, 297–318; Franjieh, Corbett, and Grandison 2020; Corbett et al. 2020). These are lexical, rather than grammatical properties, and classifiers can denote the same meanings as lexical units, such that they can convey rather specific semantic concepts, with some languages containing classifiers that are unique to specific referents, e.g. 'crocodile', 'elephant', in addition to a broader class 'animal' (see Grinevald 2015, 815–16). However, classifiers typically denote relatively general notions of shape, material, and sort, e.g. 'one-dimensional', 'animal', or 'edible', some languages have a classifier that is devoid of any semantic reference (id., 816), and classifiers are sometimes said to mark agreement with a head noun (e.g. Krasnoukhova 2012, 214–16), which are characteristics of more grammaticalized forms.

In terms of obligatoriness, the choice of classifier is a matter of pragmatics: The same referent can be classified by more than one possible classifier, dependent on the relevant, or in Allan's (1977) words *salient*, characteristic of the referent. However, the choice of classifier can also to some extent be governed by language-internal rules (e.g. Erbaugh 1986, 415–25; Zhang 2007, 56). Classifiers may be obligatory or optionally employed, their use may vary depending on their morphosyntactic locus, and some languages have subsets of classifiers that can only occur with nouns or referents of a certain class, e.g. animates.

In terms of the size of the class of classifiers, there is quite some variation. As mentioned before, classifiers come in sets with at least around twenty categories, and more than a hundred being common. While these can be called closed classes, there are also open classifier systems where the syntactic classifier slot can be filled with lexical material.[10]

2.2.1 Less grammaticalized systems

Because of the partly lexical and partly grammatical properties of classifiers, they are often taken as 'intermediately grammaticalized', somewhere in between lexical and grammatical (e.g. Craig 1986a; Grinevald 2002, 260-261). We have seen that nominal classifiers are more grammaticalized than nouns, but less so than noun classes. If we understand the grammaticalization scale as a diachronic process, we would expect to find intermediate stages of it as well, something that is indeed found in the languages of the world.

Lexical items similar to classifiers are *measure terms* and *class terms*. Measure terms are constructions such as English *a **glass of** water*, *a **loaf of** bread*, and class terms are noun-formatives of variable productivity, such as English *-berry* in words like *straw**berry*** (ibid.). These are usually not taken to be classifiers because measure terms because they are deemed more lexical than 'true classifiers' (see also Grinevald 2015, 812), and because measure terms are limited to modification of mass nouns, whereas the classifiers are not (Lyons 1977, Vol. 2, 463). In light of the current discussion, it is most suitable to interpret these on the lexical end but on the same scale as classifiers and noun class markers.

Another classifier-like category are *generic nouns*, such as those found in the Pama-Nyungan family exemplified below (7).

[10] The term *repeater* is sometimes used for classifiers that occur with nouns of the same form (e.g. Aikhenvald 2000a, 103). See Section 2.2.2 for a discussion of this phenomenon and the use of this term.

(7) a. ARRERNTE
kere aherre
game.animal kangaroo
'kangaroo' (Wilkins 1989, 107)

b. YIDINY
minya *ganguul* *jana-ŋ* ***jugi****-il* *gubuma-la*
edible.animal wallaby stand-PRS tree-LOC black.pine-LOC
'The wallaby is standing by the black pine.' (Dixon 1982, 186)

c. KUGU NGANHCARA
minha *pangku*
animal wallaby
'wallaby' (Johnson 1988, 200)

d. YIR-YORONT
minh*-lalpm*
wild.animal-wallaby
'wallaby' (Alpher 2011, 74)

Grinevald (2002) argues that generic nouns and noun classifiers, such as those found in Jacaltec (8), must be analysed as comparable systems with various degrees of grammaticalization.

(8) JACALTEC
xil ***ix*** *ix* *hune'* *hin* ***no'*** *txitam*
saw CL:F.NONKIN woman one my CL:ANIMAL pig
'The woman saw that one pig of mine' (Craig 1986b, 264)

The Pama-Nyungan generic nouns can be analysed as having varying degrees of grammaticalization. In Yir-Yoront, Kugu Nganhcara, and Yidiny, generic nouns are widely employed, but not so much in Arrernte (id., 166-169). In a quantitative study on noun categorization in the languages of Australia, Sands (1995, 270) notes that there is "only occasional cross-referencing or repetition through the clause".

In Yidiny, the use of the generic is a matter of style (Dixon 1982, 186), and in Kugu Nganhcara, it often occurs in addition to a specific noun as a shifting topic, such that it is used at the first mention of a 'noun'.[11] Typically, generic nouns in Pama-Nyungan languages can also function as regular nouns (Wilkins 2000, 209), and in both Yidiny and Kugu Nganhcara, it has an anaphoric function in this use (Dixon 1982, 187-188; Johnson 1988, 198-199). In Arrernte, generic nouns are used to provide some extra information about the context of the referent. The generic noun *kere* 'game animal', for example, is only used in contexts where the animal at hand is in fact hunted or otherwise interacted with in its capacity as *kere* (Wilkins 2000).

[11] Johnson (1988, 198) uses the word 'noun' here, and adds "or when vocabulary is being elicited". It would be interesting to track whether it is in fact newly mentioned *nouns* that activate the use of the generic noun, or newly mentioned *referents*. The use of generic nouns as continuing topics (i.e. tracking of a referent over discourse) and as anaphora might be taken to imply the latter.

Yir-Yoront has an open class of generic nouns (id., 73), although the form *minh-* covers a very broad category of nouns (id., 74-75).[12] In Yidiny and Arrernte, there are around twenty generic nouns respectively (Dixon 1982, 192; Wilkins 1989, 106). No number is given for Kugu Nganhcara by Johnson (1988).[13]

In comparison, Jacaltec noun classifiers function in a way associated with grammaticalization: They do not all occur as full nouns, they fulfil the syntactic roles of anaphora and determiners, and their use is not dependent on discursive context.[14] Table 4 summarizes the four Pama-Nyungan languages and Jacaltec, showing the different abovementioned features.[15]

Table 4: Degree of grammaticalization in four Pama-Nyungan languages and Jacaltec, compared along different parameters

	Also a noun	**Size of class**	**Anaphoric**	**Determiner**	**Style**	**Context**
Jacaltec	No	24	Yes	Yes	~	No
Yidiny	Yes	±20	Yes	No	Yes	No
K. N.	Yes	?	Yes	No	No	No
Y. Y.	Yes	Open class	?	?	?	?
Arrernte	Yes	±19	No	No	No	Yes

Synchronically, the systems compared in Table 4 show varying degrees of grammaticalization, and diachronically, systems such as Arrernte have been argued to be 'incipient' noun classifier systems as well (Grinevald 2002; Passer 2016, 68–69). These different features are therefore an interesting insight in the way that noun classifier systems may develop out of nouns, and the grammaticalization processes involved in such developments.

The relative degree of grammaticalization of the different noun categorization devices is summarized on a scale in Figure 2 below.

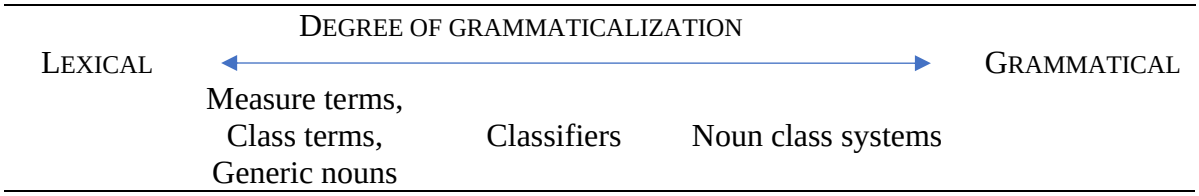


Figure 2: Degree of grammaticalization of different noun categorization devices.

[12] Additionally, Yir-Yoront *minh-* is phonologically reduced to *m-* "in fast speech in certain common species-names" (Alpher 2011, 305). This is understood as an indication of grammaticalization into a class marker by Wilkins (2000, 168, 211), but since it is only occurs in certain common names, it may be more accurate to understand this phenomenon as a process of lexicalization.

[13] Wilkins (2000, 169) cites Johnson (1988) to say that Kugu Nganhcara would only have a small number of classifiers. I cannot find such a claim in Johnson (ibid.).

[14] The Jacaltec classifiers are reduced forms of full nouns (e.g. *naj* 'CL:MAN' < *winaj* 'man'), and these classifiers generally denote broader categories than their full noun counterparts (e.g. *te'* 'CL:PLANT/WOODEN OBJECT'[14] < *te'* 'tree/stick') (Grinevald 2002, 266). The classifier *te'* denotes all plants with the exception of corn, which is a salient cultural item, and has its own clasifier *ixim* (Craig 1986b). The classifier can therefore be said to have undergone semantic bleaching.

[15] In this table, Kugu Nganhcara is indicated as K. G., and Yir-Yoront as Y. Y. in order to fit the page. The tilde '~' in the column for style indicates that there is some mention of style-related uses of noun classifiers in Jacaltec, but these are not their primary use and arguably do not fit the same definition of *stylistics* as do the Yidiny generic nouns. For reasons of space, I will not go into these issues here.

It is evident that some classifier systems are relatively lexical and others are relatively grammatical. As mentioned above, there is variation in the degree of semantic content, obligatoriness and openness of the class that constitutes the classifier system. From a diachronic perspective, noun class and gender systems may develop through further grammaticalization of classifiers (Greenberg 1978, 78; Dixon 1982, 172–73; Corbett 1991, 310–11; Aikhenvald 2000a, 372–75). In turn, classifiers are often said to develop from nouns (*inter alia* Corbett 1991, 311–12; Senft 1993; Aikhenvald 2000a, 352–412; Grinevald 2000, 61; Passer 2016), a notion further discussed in Section 2.2.2 below.

Below, I elaborate on the processes that drive grammaticalization and the lexical material involved in these processes, which will in turn inform my methodology for reconstructing the development of these classifiers in the Tukanoan family.

2.2.2 The development of nominal classifiers

The forces that drive the grammaticalization process and the lexical material it selects are discussed here. This discussion serves to direct the methodology of determining likely etymologies for classifiers in the Tukanoan family, and to provide an indication of the stage when a form grammaticalized, assuming that further grammaticalized forms started this process at an earlier stage.

Grammaticalization is best understood as a series of processes where a lexical element gradually loses its lexical characteristics and gains grammatical ones (i.e. they lose semantic content, gain in obligatoriness of use, and enter a set of limited members, see the introduction of this section). These processes are: *semantic bleaching*; *extension*; *decategorialisation*; and *erosion* (Heine and Kuteva 2002, 2).

Semantic bleaching –or *desemanticisation*– is the loss of meaning of a lexical element, such that its semantic content decreases. Recall for example the Jacaltec (Mayan) noun *te'* 'tree, stick', the meaning of which was broadened to any wooden objects, and all plants (with the exception of corn) (Grinevald 2002, 266). *Extension* –or *context generalization*– is a relatively general concept that broadens the grammatical applicability of an element. This can happen within or across morphosyntactic categories (Heine and Kuteva 2005, 50–58). *Decategorialisation* –or *decategorisation*– is the event that a relatively lexical element loses its lexical characteristics. What characteristics are prototypically lexical is largely language-dependent, and it will be readily accepted that the exact nature of these characteristics has far-reaching effects on the process of decategorialisation. *Erosion* –or *phonetic reduction*– is here understood as an umbrella term for a variety of phonological processes: shortening of forms, e.g. clipping of a syllable; loss of phonological wordhood, i.e. loss of stress or tone, cliticization, and affixation; and in the case of entire phrases, this may entail the loss of wordhood of individual parts, e.g. Brazilian Portuguese (Indo-European) *Vossa Mercê* 'your mercy' > *você* 'PRO.2SG' > *cê*= '2SG' (Vitral 1996, 119).

The different grammaticalization processes co-occur with one another, and processes of semantic bleaching often precede decategorialisation, which in turn may give rise to extension and erosion (Heine and Kuteva 2002; 2005). However, semantic bleaching does not necessarily precede decategorialization when it comes to classifiers: Many languages also contain *repeaters*: Classifiers that have morphologically identical –or almost identical– counterparts as nouns, which indicates a decategorialisation and extension process, but not a semantic bleaching, nor an erosive process. In other words, repeaters may provide some insight in order of the processes involved in the development of classifiers. The term *repeater* is used in two slightly different ways in the literature and therefore requires some clarification.

In a broad definition (e.g. Seifart 2005, 77–81), *repeaters* are classifiers that can also be found as nominal heads (9). In a narrow definition, repeaters are classifiers that 'classify themselves' (Enfield 2014, 120; Aikhenvald 2000a, 103–4), such as in (10) from Lao (Tai-Kadai). In MAI in (9), the nominal root *dórù* 'basket' is used as a free noun in (9a), and as a classifier with a slightly different from *-roru* in (9b). The classifier form has no inherent tone and the initial stop is realised as [r], but these are fairly phonologically predictable processes in the language (Farmer 2015, 20–22).

(6) a. MÁÍHH̃ÌKÌ (TUKANOAN)
***dórù**-ma* *bèè-hɨ* *áhè-yi*
basket-INAN.PL carry-PL.SS.SIM go.down-3.PL.PRS.DECL
'They were going down carrying baskets' (Farmer 2015, 137)

b. *bíà-**roru*** *bèò-re* *sáí-ko*
chili.pepper-CL:BASKET put.on.back-SS.SEQ go-3SG.F.PRS.DECL
'She put her chili pepper basket on her back and went' (ibid.)

In (9b), the classifier *-roru* combines with a noun different from *dórù* itself, which does fit the broad definition of the term *repeater*, since the same form also occurs as a free noun (9a). The classifier *-roru* does not fit the narrow definition, which is reserved for classifiers that can only occur with the noun of the same form, such as in (10).

(10) LAO (TAI-KADAI)
kuu³ *si⁰* *hêt¹* *hang²* *sip²* ***hang²***
1SG.B IRR make nest ten nest
'I'm going to make ten nests (for the chickens)' (Enfield 2014, 120)

The difference between the two definitions of repeaters seems to be whether repeaters are productively derived from the class of nouns, and as a result form an open class, or whether there is simply a number of classifiers that have identical or nearly identical counterparts as nouns but which cannot be freely derived from any noun. For example, Farmer (2015, 36–37) identifies just nineteen repeaters for MAI, whereas repeaters in Lao can be derived productively "whenever it is not obvious what the appropriate numeral classifier is" (Enfield 2014, 121). As mentioned above, openness of the class is an indication of grammaticalization, and as such, repeaters are of interest for the discussion of the development of classifiers from nouns.

Senft (1993) shows that only some of the classifiers –or 'classificatory particles'[16]– found in Kilivila (Austronesian) have undergone semantic bleaching, and that as many as twenty-five classifiers in the language can be analysed as repeaters, where the fact that Senft (ibid.) finds a limited number of twenty-five repeaters in a variety of Kilivila points to the broad definition of repeaters. Therefore, Senft (ibid.) proposes that the Kilivila classifier system originally consisted of only repeaters, and that some of these underwent processes of phonological reduction and/or semantic bleaching. This explanation seems plausible because it accounts for the similarity between constructions with noun classifiers and nominal compounds (Farmer 2015), and because it fits the kind of analysis posed in Section 2.2.1, that classifiers –or at least noun classifiers– arise from nominal compounds.

[16] This term also refers to quantifiers in Kilivila (Senft 1993).

The idea that nominal classifiers grammaticalize from nominal compounds is also discussed for Witotoan languages by Seifart (2007), and for Yagua (Peba-Yaguan) by Payne (2007). However, Payne (id., 465) problematizes a grammaticalization path through phonological repeaters because the syntactic derivational function and semantic bleaching of classifiers would not be explained by it. Nevertheless, the same is true for a grammaticalization path that does not involve repeaters. Since repeaters –in the broad sense– are nothing more than a syntactic subcategory of classifiers, the morphological material of which is productively derived from nouns. The emergence of the syntactic category of classifiers in any capacity may have compounding as its source, and it may therefore not be problematic to accept repeaters as an incipient stage of classifiers, while also maintaining that they are grammaticalized compounds. The processes and intermediate stages of grammaticalization from nouns to classifiers can thus be regarded with some detail, and are summarized in Figure 3, where I provisionally refer to classifiers that are not repeaters of any kind as *full classifiers*.

	DEGREE OF GRAMMATICALIZATION →				
LEXICAL					GRAMMATICAL
Syntactic status	Nominal compounds	Repeaters (Narrow)	Repeaters (Broad)	Generic nouns	Full classifiers
Open class	Yes	Yes	No	Yes	No
Decategorialization Extension	No	Yes	Yes	Yes	Yes
Erosion Semantic bleaching	No	No	No	Sometimes	Yes
Example languages	[Omnipresent]	Lao	Máíhɨ́kĩ	Pama-Nyungan	Mandarin

Figure 3: The development of classifiers from nouns.

Since classifiers have meaning and fulfil a pragmatic function, the context of discourse matters for the choice of classifier, and even when some classifiers have undergone some level of grammaticalization, others may continue to enter the grammatical category when pragmatically required. In other words, repeaters may continue to provide morphological material for classifiers once there is a syntactic category of classifiers available in the language.

By no means do I claim, however, that *all* classifiers develop from nominal compounds through repeaters. For example, Aikhenvald (2000a, 362–65) mentions some languages that have classifiers which may have developed from verbs, and Payne (2007, 465) posits that the Yagua general classifier *-ra* developed directly from the inanimate pronoun *ra*. However, the notion that classifiers in the Tukanoan family likely developed from compounding is of importance because it may also help us to understand what force has driven the languages of the Northwest Amazon to develop the classifier systems that we see today, and why they are similar in some respects.

Once a classifier system is in place, it may further grammaticalize in the sense that it may become a closed set, classifiers may turn to be used obligatorily in certain contexts, and classifiers may be employed in systems of agreement within, or even outside of the noun phrase (e.g. Corbett 1991, 311; Aikhenvald 2000a; Passer 2016). The use of classifiers in agreement systems is a pervasive feature of Northwest Amazonian languages (e.g. Payne 1987; Derbyshire and Payne 1990; Farmer 2015, 125–30) and has been reported in other languages in the wider Amazon area as well (e.g. Krasnoukhova 2012, 218). Due to these characteristics, associated with highly grammaticalized classifier systems, they have been called 'emerging noun class systems' (e.g. Grinevald and Seifart 2004).

In the next section, the Tukanoan family and its noun categorization systems are discussed in more detail.

3 Systems of noun categorization in the Tukanoan family

The most striking features of Tukanoan nominal classifiers are the fact that they blur the lines between classifiers and noun classes on a number of accounts, and that they employ identical, or very similar sets of nominal classifiers in multiple morphosyntactic loci. As mentioned in Section 2.2, classifiers and noun classes cannot be rigidly distinguished, but classifiers are expected to demonstrate relatively lexical features, whereas noun classes are expected to be relatively grammatical. It is specifically in this area that Tukanoan classifiers defy the typology by demonstrating some rather grammatical features. Another typical feature of Tukanoan classifier systems is a difference in patterning between inanimate and animate classifiers, where the former are large sets denoting meanings of *inter alia* shape, material, and use, and the latter are limited to around three or four forms, denoting gender and number.

Below, I discuss the morphosyntactic properties and functions of Tukanoan gender systems (Section 3.1), and classifier systems (Section 3.2). These properties and functions are relevant to the development of the markers since the morphosyntactic and functional differences between these markers justify multiple development paths and provide information as to what these paths may have been.

3.1 Tukanoan gender systems

Gender is marked in all Tukanoan languages and can be functionally differentiated from classifiers in the whole family. Morphologically, the Tukanoan gender markers are also very similar. Throughout the family, animate nouns and person indexes are split into at least three categories: feminine, masculine, and animate plural. These genders are often distinguished on verbs as person indexes or nominalizers, and on demonstratives, pronouns, and animate nouns, where they are often obligatorily marked on those referring to humans and higher animates, and optionally on those referring to lower animates such as most animals. Celestial bodies, such as the sun, moon and stars are often grouped together with animates which is consistent with their traditional role in local ontologies (e.g. Wheeler 1970, 27–28; Gomez-imbert 2007, 17; Jackson 1983, 205–8).[17] In addition to the three aforementioned animate genders, some Tukanoan languages also have a general class, an inanimate class, or both. I treat these markers as *gender* because functionally and morphosyntactically, they belong to the same set. Many authors refer to the combination of animate and inanimate gender markers as 'general classifiers,' which they oppose to 'specific classifiers,' i.e. those referring to shape, material, composition *et cetera* (see Farmer 2015, 126).

[17] In other language families, celestial bodies are also often grouped with animates (Regúnaga 2012, 79, 115–6).

In E. SIO, for example, there are three general classifiers: *-o/-ko* 'CL:F'; *-i̵/-ki̵* 'CL:M'; and *-e/-je* 'CL:GENERAL'. Both general and specific classifiers in the language can occur as noun formatives, but agreement is mostly a function of the general classifiers (Bruil 2016, 8–9). Consider example (11), where the noun *kĩ̵napĩ̵* 'stone' contains the specific classifier *-pĩ̵* 'CL:ROUND', but it is modified by the qualifier *hãiko* 'big' which contains the feminine classifier *-ko*. Note that nominal agreement is not obligatory in the language, as is demonstrated by the other qualifier *pohai* 'white' which does not bear an agreement marker in (11).

(11) ECUADORIAN SIONA (WESTERN TUKANOAN)
(…) *jaico guẽnabë mareña bojai guënabëre*

hai-ko	**kĩ̃na-pĩ̵**	mã-ɾe-jã	pohai
big-CL:F	brilliant.material-CL:ROUND	climb-N2/3SG.PST.NASS-REP	white

kĩ̃na-pĩ̵-re
brilliant.material-CL:ROUND-O
'(…) they climbed a big stone, a white stone.' [20150720sbapa001repi.00:05]

General classifiers mark agreement with the head noun on numerals, other quantifiers, qualifiers, demonstratives, and content question words, whereas specific classifiers are not found on content question words and are not productive, albeit accepted as grammatical, on numerals, qualifiers, and demonstratives (Bruil 2014, 138; 2016, 8–9). Additionally, the generic classifiers are found on verbs as person indexes: *-o/-ko* refers to feminine subjects; *-i̵/-ki̵* to masculine subjects; and *-e/je* to first person and plural subjects (Bruil 2018, 130).[18]

Since the feminine classifier *-o/-ko* often occurs in agreement with inanimate nouns (11) (Bruil 2016, 8), and since the feminine marker often occurs when referring to inanimate subjects in nominalizations (12), it is synchronically more suitable to analyse this marker as a non-masculine gender marker.

(12) ECUADORIAN SIONA (WESTERN TUKANOAN)
a. *ju̱'i̱o huëoja̱'co*

hũ'ĩ-o	wi̵-o-hã'-**ko**
be.sick-3SG.F.PST.ASS	start-CAUS-NLZ-CL:F

'[the story] starts when she got sick.' [20150208srocr001.006]

b. *ja̱o ai yequë jubë jai jubë ja̱obi aiyo ja̱obi aiyo ja̱oni huatotojä̱'ë caoña*

hã-o	ai	jehk-i̵	hubi̵	hai	hubi̵
DEM.DIST-CL:F	more	other-CL:M	bunch	big	bunch

hã-o-bi	ai-o	**hã-o-ni**	wahtoto-hĩ'ĩ̃
DEM.DIST-CLS:ANIM.F-S	big-3SG.F.ASS	DEM.DIST-CL:F-O	take.down-IMP

ka-o-jã
say-2/3SG.F.PST.NASS
'"That one, the other bunch, the big bunch is ripe, take that one down!" she said.'
(Bruil 2014, 349)

[18] I use here the term *subject* to refer to the sole argument of an intransitive clause, or the agent of a transitive clause.

Lastly, the classifiers *-ɾo/-to* 'CL:PLACE', and *-ɾĩ* 'CL:TIME' have "an extensive use as well" (Bruil 2014, 138), although these do not occur as person indexes on finite verbs (ibid.).

A distinction between gender –or 'general classifiers'– and classifiers –or 'specific classifiers'– is ubiquitous in Tukanoan languages, but the exact functions and distributions of gender and classifiers differ across languages. Some features are fairly constant: Many languages have animate gender markers which are morphologically fused with number and a two-way distinction among animates. This two-way distinction entails either a masculine/non-masculine opposition, such as *-o/-ko* in E. SIO, or a feminine/non-feminine opposition.

A feminine/non-feminine distinction can be recognized in the fact that animals are often individualized through the marker that is generally associated with masculine animates (13a, b), and the fact that the default, unmarked form of many other nouns referring to animals is implicitly masculine, of which a feminine form can be derived through suffixation of a feminine noun classifier (13c).

(13) a. KUBEO (EASTERN TUKANOAN)

moa	→	moa-kɨ
fish		fish-CL:M
'(some) fish)'		'a fish'

(Chacon 2012, 237)

b. TATUYO (EASTERN TUKANOAN)

wekɨ́	atí-~bí
tapir	come-3SG.M

'a (male) tapir comes' (Gomez-Imbert 1996, 449–50)

c. IDEM

~ɑ́ɑ́bó-kɨ́	→	~ɑ́ɑ́bó-kó
chicken-CL:M		chicken-CL:F
'chicken, cockerel'		'hen'

(ibid.)

It is interesting to note that these examples contrast with the E. SIO examples in (12). The question of whether the difference can be analysed as a WT versus ET divide, requires a more in-depth look into some other languages in the family.[19]

The forms associated with the animate gender-number markers in Tukanoan languages are fairly constant across the family. In Section 5, I discuss Tukanoan gender markers in more detail, and provide them with a diachronic analysis.

3.2 Tukanoan classifier systems

All Tukanoan languages about which anything is known have classifiers, and in all but BSA and DES, these differ from gender markers in their morphosyntactic distribution. For example, in almost all Tukanoan languages, the gender markers –or the majority of gender markers– are also found as person indexes on the predicate. Gender markers are fused with grammatical number distinctions in all Tukanoan languages, and all –with the exception of TAN and LET,

[19] Verbal agreement in E. SIO seems to conform to this non-feminine gender pattern, in contrast to nominal agreement which contrasts non-masculine to masculine. However, careful consideration is required, since the nominal and verbal morphology are formally very similar, such that it is not always easy to judge whether a certain construction is nominal or verbal.

two closely related languages that have long been in contact with Yukuna (Arawakan)– have noun classifiers.[20]

All Tukanoan languages, with the exception of the SEK varieties[21] and LET, demonstrate some form of agreement with the head noun through classifiers or gender markers. These agreement markers occur on demonstratives and numerals, and most do so on other quantifiers as well. An overview of the functions and morphosyntactic loci of classifiers and gender markers in Tukanoan languages is given in Appendix D.

In many languages in Northwest Amazonia that have noun classifiers, these are employed as *noun-formatives* (Krasnoukhova 2012, 209–11): Devices that derive nouns that are specified for some feature denoted by the classifier. As such, they may derive nouns from other parts of speech, which can then fulfil the syntactic slot of a nominal head, or they may derive nouns from other nouns (ibid.).

The nominalizing function of noun classifiers is exemplified particularly well by languages that do not mark classifiers on modifiers, but do so on heads. Consider (14a, 15) where a noun is modified by a 'bare' numeral *mena* 'two' and *teʔ* 'one', i.e. a numeral that does not agree with the head noun it modifies. In (14b, 15) the numeral fulfils the slot of the nominal head itself, and is marked with a classifier. Muinane has a generic marker *-je* which I do not analyse as an agreement marker, since there is no covariation of the classifier with its controller, i.e. the head with which it "agrees" (see Section 2.1.3.2).

(14) MUINANE (WITOTOAN)

a. ***mena*** *rɨ-ño*
two woman-CL:DR.F
'Two women' (Wojtylak 2017, 151)

b. *mena-**kaɨ*** *kue-mo* *ine!*
two-CL:STICK 1SG-LOC give.IMP
'Give me two (stick-like form, here: cigarettes)!' (ibid.)

(15) PERUVIAN SEKOYA (WESTERN TUKANOAN)

teʔ *ĩmɨ* ***teʔ*** *nomio* *ʤeha-re* *ñui-o* *nokwa-re* *wato-ri*
one man one woman ground-LOC sit-DEP.SG.F banana-O pick-SS
kaʤa–wɨ–ã *ĩʦi–o*
two-CL:CYLINDRICAL-PL.INAN give-3SG.PFV.DE
'There is a man, a woman sitting on the ground picks up banana, and gives him two (cylindrical form, here: bananas)' (Vallejos 2021a)

Examples like these imply that it is the classifier that derives the nominal heads *menakaɨ* 'two cigarettes' in (14b), and *kaʤawɨã* 'two bananas' in (15).

[20] An interesting fact, given the cross-linguistic rarity of noun classifiers (Kilarski 2013, 295–97).

[21] Johnson and Levinsohn (1990, 38–40) analyse E. SEK as having agreement, but the examples they give are very similar to P. SEK constructions discussed by Vallejos (2021a; 2021b), who argues that these are not agreement. Furthermore, Vallejos (p.c. 2021) suggests that the two varieties are similar in this respect.

There is some reason to believe that agreement-like constructions in E. SEK are somewhat more accepted and common than in P. SEK (Johnson and Levinsohn 1990, 38; cf. Vallejos 2021a; 2021b), which might be the result of the intensive contact between E. SIO, which has a further grammaticalized noun categorization system, and which does demonstrate agreement (Bruil 2014; 2016; Farmer and Bruil 2015; Farmer and Vallejos, 2021). However, this hypothesis would require more detailed data and analyses.

Denominal derivations, where nouns are marked with classifiers to form new nouns with different meanings, or other highlighted characteristics, are also extremely common throughout the area, and this is an omnipresent feature of Tukanoan noun classifiers as well. This function of noun classifiers is exemplified in (16), from TUY, where one noun *hoó* 'banana' may appear bare, but from which a number of banana-related terms can be derived, through the suffixation of a classifier (Barnes 1990, 283).

(16) TUYUKA (EASTERN TUKANOAN)

a. *hoó-tõ*
banana-CL:STALK
'banana stalk' (Barnes 1990, 283).

b. *hoó-pũ*
banana-CL:LEAF
'banana leaf' (ibid.)

c. *hoó-poro*
banana-CL:CYLINDRICAL
'banana' (ibid.)

Another function of nominal classifiers that is found widely throughout the family is *individuation*, i.e. a noun with an indefinite number of referents –also called *mass noun*– is individuated, thus conveying a or number of referents, or a single, delimited referent. Given the observations made in Section 2 that classifiers refer directly to referents, it is not unexpected to see these two functions combined: A nominal classifier 'points', as it were, to a specific being or object, or a feature thereof, and as a result, the classifier individuates it as a concrete referent.

Note that *individuation* does not entail singular number, but rather the difference between no specified number and some number, which could, theoretically, be counted. The individuating function of classifiers is exemplified in (17, 18). In KOR, some nouns are understood as a 'generic' (Cook and Criswell 1993, 18), or as an 'unspecified plural' (Cook and Gralow 2001, 22) when they occur without a classifier (17). In E. SIO, inanimate nouns that "belong to a specific noun class" (Bruil 2014, 154) must carry a classifier suffix preceding the plural marker *-ã* when they are pluralized (18b), but do not require it when they are not (17a) (ibid.).[22]

(17) KOREGUAJE (WESTERN TUKANOAN)

a. *d͡zoo*
canoe
'canoes (in general)' (Cook and Gralow 2001, 22)

b. *d͡zoo-wɨ*
canoe-CL:LONG.NARROW.WRAPPED
'a canoe' (ibid.)

[22] E. SEK nouns must carry a noun class marker before they can be pluralized as well (Schwarz (2011), cited in Bruil (2014, 154)).

c. *d͡zoo-wɨ-ã*
canoe-CL:LONG.NARROW.WRAPPED-PL
'canoes (specifically plural)' (ibid.)

(18) ECUADORIAN SIONA (WESTERN TUKANOAN)
a. *nohka*
banana
'banana' (Bruil 2014, 154)

b. *nohka-mo-ã*
banana-CL:CYLINDRICAL.FLEXIBLE-PL
'bananas' (ibid.)

In light of the noun-formative function of classifiers, it is relevant to note that there are strong rules that impede lexical loans from other languages in the Vaupés area, where some of the ET languages (TUK, PIR, KOT, and DES, and to some degree TUY, TAT, SIR, and KUB) are spoken, along with some Arawakan languages (Tariana, and on the outskirts of the area also Baniwa of Içana, Warekena, Baré, and Piapoco) Nadahupan languages Dâw, Hup, and Yuhup, and one Kariban language, Karihona (Aikhenvald 2002, 17–19). In these situations, various strategies can be employed to derive new nouns from language-internal morphology to denote newly introduced objects and concepts, one of which is the extensive use of compounding, a strategy that has been reported for the introduction of new objects, and for biological expert knowledge terms alike in the Chaco, another area in the Americas where lexical loans are avoided (Campbell 2013, 272; List, Messineo, and Brid 2021).

A similar phenomenon has been reported in the Vaupés, e.g. DES *dipa-sabe* (GEN.ADJ-CL:SWAMPY.AREA)[23] 'swamp' > Tariana (Arawakan) *maka-pina* (GEN.ADJ-CL:SWAMPY.AREA) 'swamp' (Aikhenvald 2002, 94). Here, new nouns can be derived through the suffixation of a classifier to a dummy adjective. The constructions in the DES and Tariana examples above are identical, but the forms are not. This is a specific type of calque which forms a productive method for introducing new terminology into a language on the basis of language-internal morphological grounds. The use of classifiers to derive nouns to refer to new objects is found widely, e.g. in C. SIO *uti-pɨbɨ* (paper-CL:PILE.OF.LEAVES) 'book' (Wheeler 1987).

Since classifiers are often grammaticalized nouns (e.g. Passer 2016), since Chacon (2012, 257; 2021) argues that this is also the case for classifiers in Tukanoan languages, since nominal compounds are a frequently used noun-formative strategy, and since this is also one of the main functions of nominal classifiers in Amazonian languages (Krasnoukhova 2012, 209–11), I put forward that Tukanoan noun classifiers have originated from this specific function of nominal compounds, which may also be true for surrounding languages.

Indeed, the Nadahupan language Hup is analysed as having an emergent classifier system, where bound nouns are extensively employed as noun-formatives to denote new concepts to the language (Epps 2008, 267–80), and where there is at least one probable example of a calque construction involving a noun classifier from Tukano: deer-CL:LEAF 'carurú, pokeweed' > HUP *hɔhɔ̃j=g̃æt* (id., 265). A similar scenario may have initiated the widespread phenomenon in other unaffiliated languages. A similar development pattern has also been suggested for Yagua (Peba-Yaguan) classifiers (Payne 2007).[24] It may therefore not be unexpected to see noun

[23] Miller (1999, 44) glosses *dipa* as an article.
[24] Mithun (1986, 380) puts forward a similar theory for the grammaticalization of verbal classifiers.

classifiers across the boundaries of genetic affiliation in the Vaupés area. The fact, however, that classifiers fulfilling this function are also found beyond the Vaupés area –e.g. the WT languages, and a number of Arawakan languages– and the fact that some features of classifiers show strong genetic tendencies in the area suggest that classifiers have not developed independently each language. It is therefore of interest to determine which classifiers grammaticalized at what stage of each family, based on recognizable cognates.

4 Methodology

In order to determine which Tukanoan classifiers can be reconstructed for PT, Proto-Western-Tukanoan (henceforth *PWT*), or Proto-Eastern-Tukanoan (henceforth *PET*), I have gathered all classifiers, their glosses, and some associated nouns that could be found in existing works on the respective Tukanoan languages (see Table 1).

On the basis of these forms, I employ the comparative method to determine which forms are cognate and how these may be reconstructed, taking as a point of departure the sound correspondences as determined in Waltz and Wheeler (1972), Chacon (2014), Chacon and List (2015), and Chacon and Michael (2018). This entails that I look for structural sound correspondences between similar forms in order to determine whether forms in two different hail from the same form in the proto-language. However, the comparative method alone does not always provide conclusive evidence on whether a form be inherited, due to the following three problems:

i) bound forms or affixes may have undergone different sound changes than sounds in free forms or roots did (Chacon 2014; Chacon and Michael 2018), but these bound form sound changes have not yet been described in detail;
ii) some classifiers may have very similar forms and meanings to those of another language, but show unexpected sound correspondences, from the perspective of Waltz and Wheeler (1972), Chacon (2014), and Chacon and List (2015), which may indicate diffusion, analogy, or another phonology-external factor;
iii) not every classifier contains sounds that would display relevant sound laws compared to those classifiers found in related languages, because the sound changes between closely related languages are often minimal (see Chacon 2014).

Therefore, classifiers' meanings are another important indicator for identifying possibly related forms. Comparing meanings cannot yield conclusive evidence for or against cognate status, because detailed justification for all glosses is rare in the grammars available, because the meaning of related forms may show some divergence across languages and branches, because unrelated forms may have similar meanings and uses synchronically, and because direct and indirect diffusion may also be responsible for the semantic similarity.

When it is not possible to determine whether a form is inherited, some characteristics may indicate whether the form grammaticalized recently or at an earlier stage. These characteristics are closely related to the properties of highly grammaticalized forms, since a high degree of grammaticalization may is associated with early development. Characteristics associated with early development are then: phonologically reduced forms; morphologically simplex forms; and syntactically widespread occurrence, i.e. occurrence throughout the noun phrase, rather than simply on the nominal head.

Another factor which may indicate early grammaticalization is whether a form is found throughout the family or not, e.g. if a form is found in Nuclear and Non-nuclear ET languages, this is an indication that the form may have grammaticalized as early as the PET stage, and if a

form is found in both WT and ET languages, this is an indication that the form may have grammaticalized as early as PT stage. With this methodology, it is important to consider which languages are found in which geographical areas, and what is known about language contact between relevant groups, in order to determine whether a form is likely inherited or the result of language contact.

As was discussed in Section 2, classifiers are not lexically, but pragmatically chosen. However, it is important to note that nouns may be associated with prototypical classifiers, such that the choice of classifier is not entirely determined lexically, nor entirely pragmatically, such as in Mandarin Chinese (Sino-Tibetan) (Erbaugh 1986, 415–25; Zhang 2007, 56). Since extensive accounts of emic ontologies are not available in many languages under consideration, since most authors only list a few examples of nouns associated with each classifier, and since these associations seem to be rather prototypical language-internally (Gomez-imbert 2007), I consider prototypical pairings of nouns and noun classifiers as well in determining the common source of classifiers in the Tukanoan languages. In other words, in addition to considering analyses provided by authors on the semantics of Tukanoan classifiers, I take an objective approach to the semantics associated with classifiers' by considering the classifiers prototypically associated with certain nouns. An additional strength of this approach is that noun classifiers are determined as a possible source of the classifier systems in the languages of the area under study, from which the other morphosyntactic functions may have developed (see Section 6).

On the basis of observations in this thesis, it is possible to determine a number of likely cognates for which reconstructions could be proposed either to PT, or to a subbranch of the family. On the basis of the reconstructed forms, I also postulate a number of possible etymologies. Likely cognates, their tentative reconstructions for relevant proto-languages, and their possible etymologies are listed in Appendix B. In turn, it is also possible to postulate a number of sound correspondences for consonants in bound forms, which are listed in Appendix C.

I am keenly aware that this thesis cannot do justice to the detail required to judge conclusively the probability of any of the abovementioned three scenarios. Therefore, I put forward some tentative hypotheses and welcome alternative and additional solutions to the problems and suggestions posed below.

5 Gender

In this section, I give an overview of gender in the Tukanoan family, and provide a diachronic analysis of its development, arguing that these markers can be reconstructed as complex PT forms, where it were essentially the vowels that differentiated the gender features (Section 5.1). Section 5.2 discusses the possibility that the PT gender system developed under Arawakan influence. Section 5.3 argues that at one point the PT gender system contained a two-way gender distinction between animate and inanimate, and that the system distinguished between singular and non-singular number. At a later stage of PT, it also developed a feminine/non-feminine distinction within the animate gender.

As discussed in Section 3.1, the Tukanoan gender and number markers are found in multiple morphosyntactic contexts and although they show some variation between them, as well as between languages, they are fairly constant. Some animate gender and number markers are summarized in Table 5, which lists dependent and independent personal pronouns and nominal markers in eight Tukanoan languages, where I have taken the declarative, present tense, unmarked or direct evidential paradigms of independent verbs in each language.

Table 5: Dependent and independent personal pronouns and nominal markers in eight Tukanoan languages

		C. SIO	E. SEK	KOR	TAN	DES	KUB	KOT	TUK	TAT
M.SG	DEP.PRO	*-hi*	*-kɨ́* *-ɨ́* *-i*	*-mɨ́*	*ki-*[25]	*-bĩ*	*-ãbe* *-bi*	N/A	*-bĩ*	*-~bi*
	IND.PRO	*bágɨ́*	*ĩ̀*	*-kʰi* *-ɨ́*	*i'kí*	*ĩgɨ̃́*	*ɨ̃́*	*tíró*	*kɨ̄ɨ̄́*	*kɨ́ɨ́*
	NOMINAL	*-gɨ́*	*-i* *-kɨ́*	*-ɨ́*	*-ki* *-i* *-ka*	*-gɨ́*	*-kɨ̃́*	*-ɨ́*	*-gɨ́*	*-ɨ́* *-kɨ́*
F.SG	DEP.PRO	*-go*	*-ko* *-o*	*-kʰo* *-o*	*ko-*	*-bõ*	*-ako* *-biko*	N/A	*-bõ*	*-~bo*
	IND.PRO	*bágo*	*ĩo*	*-mo*	*i'kó*	*ĩgõ*	*õ*	*tíkóró*	*koô*	*kóo*
	NOMINAL	*-go*	*-o* *-ko*	*-o*	*-ko* *-o*	*-go*	*-ko*	*-ko* *-o* *-koro*	*-go*	*-o* *-ko*
ANIM. PL	DEP.PRO	*-jɨ́*	*-ʤɨ́* *-ʤe*	*-me*	*~dà-*	*-bã*	*-ibã* *-bã*	N/A	*-bã*	*-~ba*
	IND.PRO	*bákó'á*	*ĩo wa'i*	*-na*	*~i'rá*	*ĩrã* *ẽrã*	*dã*	*tína* *tínũmĩã*	*dãâ*	*~dáa*
	NOMINAL	*-kó'á* *wa'í*	*wa'i*	*-na* *pãi* *romi*	*-~ra*	*-rã*	*-dã* *-wa*	*-a* *-~da* *-ja*	*-rã* *-a*	*-a* *-~ra* *-~da*
Ref.		I	II	III	IV	V	VI	VII	VIII	IIX

It is clear that the three animate genders (masculine or non-feminine, feminine or non-masculine, and animate plural) are fairly constant across all these languages, and that the cross-linguistic variation is only slightly greater than the language-internal variation. For example, it is true that the masculine dependent pronoun *-kɨ́* found in E. SEK is not found in the C. SIO definite declarative present paradigm, where the third person masculine pronoun is *-hi*, but the form *-kɨ́* is found as the dependent perfect, and indefinite past tense masculine person markers (Wheeler 1970, 92–93). Moreover, the masculine nominal marker *-gɨ́* is rather similar to both the dependent pronoun and masculine nominal marker *-kɨ́* found in E. SIO. The same is true throughout the family, and especially the feminine and masculine nominal markers are quite similar. Furthermore, it must be noted that many Tukanoan languages have complex person marking paradigms with fused tense, aspect, mood, evidentiality, person marking and gender, and that most forms listed in Table 5 are found more widely than could be represented here, but with a different grammatical meaning, i.e. C. SIO *-kɨ́*, as exemplified above.

Chacon (in prep.) notes that inanimate demonstratives and nominalizing morphology is rather similar between the Tukanoan languages as well. Consider Table 6 below, adapted from Chacon (ibid.), on which I have expanded slightly for consistency.

[25] The same forms occur suffixed to verbs in TAN, referring to the verb's object (Eraso 2015, 226).

Table 6: Inanimate demonstratives and nominalizing morphology (adapted from Chacon (in prep.))

	NEUTER		GENERIC		LOCATIVE		
	DEM	**NLZ**	**DEM**	**NLZ**	**DEM**	**NLZ**	**TYPE NLZ**
C. Siona		*-ru*		*-je*	*ʔɨ-du*	*∅*	SIMULTANEOUS
E. Sekoya	*i*		*i-je*		*ĩño*		N/A
Máíhɨ̀kì	*ĩ-ti*	*-hai ∅*	*ĩ-ti*	*-hai-je*	*ĩ-ti*	*-hai-je*	FUTURE
Koreguaje		*-d͡ze*	*i-e*	*-d͡ze*		*-na*	SIMULTANEOUS
Tanimuka	*iʔ-ká*		*iʔ-ká*	*-rí-ká*	*i-to / to*		PROCESS
Desano	*i-di*		*i-di*		*i-do*		N/A
Kubeo	*di*	*-ri*	*di=je*	*-e*	*no*	*-rõ*	STATIVE
Tukano	*ti*	*-ti*	*tee*	*-te*	*too*	*-to*	SIMULTANEOUS

On the basis of these comparisons, certain recurring animate and inanimate markers can be recognized. As mentioned before, there is some language-internal variation, and this variation is mainly concentrated on the consonants involved. Chacon (in prep.) shows that the vowels *ĩ, ɨ, o, e, ã, a* are fairly constant throughout his sample, from which he distils some possible reconstructions, which are summarized in Table 7.

Table 7: Gender vowels and their origins (adapted from Chacon (in prep.))

Category	**Reconstruction**	**Etymology**
MASCULINE	**ĩ*	Third person, reanalysed as masculine in the Nuclear ET branch, and affixed to dependent and independent proforms.
	**-ɨ*	Animate singular, reanalysed as masculine, affixed to proforms, nouns, and nominal modifiers.
FEMININE	**-o*	Neuter singular, reanalysed as feminine marker when referring to animates and as locative when referring to inanimates.
LOCATIVE		
GENERIC	**-e*	Uncertain origin, but it occurs consistently as **-je* 'indefinite pronoun'. Synchronically, unspecified gender and extended to 'inanimate plural'.
ANIMATE PLURAL	**-ã / *-a*	Noun formative. Many plural, animate or neuter forms are marked with this vowel.

Chacon (in prep.) analyses **-ɨ* as an animate marker originally unspecified for gender because personal pronouns that do not distinguish between gender, i.e. **jɨ-* '1SG.PRO', and **mɨ-* '2SG.PRO' both take *-ʔɨ*, e.g.: E. SIO *jɨʔɨ, mɨʔɨ* (Bruil 2014, 204); KOR *ǰɨʔɨ, mɨʔɨ* (Cook and Criswell 1993, 31); DES *jɨʔɨ, bɨ̃ʔɨ̃* (Miller 1999, 30); TUK *jɨʔɨ̂, bɨ̃ʔɨ̂* (Ramirez 1997, Vol 1., 322); TAT *jɨ̀ɨ̀, ~bɨ̀ɨ̀* (Gomez-Imbert 1982, 224), all of which refer to the first person singular and second person singular respectively.

It is probable that the PT gender markers were complex forms, consisting of a morpheme with a bilabial or velar consonant in the onset and a vowel as listed in Table 7. In Section 5.1, I discuss the complex nature of these forms. In Section 5.2, I discuss the possible Arawakan influence on the Tukanoan gender markers.

5.1 The complex nature of the PT gender markers

As shown above, the vowels in the gender markers are rather constant, and are associated with the gender categories denoted by the markers. Consonants in the onset of the markers show some more variation within a given gender, and the morphosyntactic locus of the gender marker, rather than its gender value, is indicative of the consonant in the marker's onset.

The onset in dependent pronouns is a bilabial in most languages, a voiced or voiceless velar in others, and in others still, there is no onset. The nominal agreement markers contain a velar or no onset at all, and the laryngeal properties of the velar show some variation: Some languages have markers that contain a voiced or laryngealised velar, others have markers that contain a voiceless velar, and others still have both (see Table 5).

Since the vowels are stable across these functions, but the onsets vary, I argue that the bilabial onset found in dependent pronouns developed from a PT tense-aspect-mood (henceforth *TAM*) marker **-p'ɨ* (see Chacon and Michael 2018), and that the velar found in dependent and independent pronouns, and in nominal gender markers is a development from a PT definite marker **ka*. Thus, I analyse the stable vowels as separate morphemes referring only to gender (see Table 7).

5.1.1 PT **-p'ɨ*

The bilabial in dependent pronouns is found widely throughout the family, i.e. Non-Nuclear ET KUB *-bã̃* 'ANIM.PL', and WT C. SIO has a glottal fricative *-hi* which can be reconstructed as a reflex of PT **p* (Chacon and Michael 2018, 64–70).[26] According to Chacon and Michael (2018), the bilabial can be reconstructed as a PT TAM marker **-p'ɨ* (in the case of an odd number of morae in the suffix group), and **pɨ̄* (in the case of an even number of morae in the suffix group).

I will not repeat their discussion here, but simply note that those gender markers with a bilabial consonant can be reconstructed as a TAM marker giving rise to the onset, which is fused with a gender and number marker giving rise to the vowel. I follow Chacon and Michael (2018) who claim that it is the close central vowel /ɨ/ that was original to the TAM marker, and that it was not until much later that the feminine gender marker **-p'ɨ-õ*/*-pɨ-õ* developed to contrast with it, thus giving rise to a non-feminine gender **-pĩ*/*-p'ĩ*.

The laryngealised bilabial stop in **-p'ĩ* is also found in what Chacon (2014, 293) and Chacon and Michael (2018) reconstruct as PT **-p'ɨ* 'non-third animate person marker'. It is therefore possible that ET masculine singular **-p'ĩ*, feminine singular **-p'õ*, and animate plural **-p'ã̃* all developed as the result of complex morphology, viz.: **-p'ɨ-ĩ* 'ANIM-M'; **-p'ɨ-õ* or **p'ɨ-ĩ-o* 'ANIM-F'; and **-p'ɨ-ã̃* 'ANIM-PL' respectively (see also Chacon and Michael 2018).[27]

5.1.2 PT **ka*

The velar is found in dependent and independent pronouns, as well as in many nominal gender markers, and may be analysed to have developed from a deictic **ka*, which is still found in languages of both main branches of the family: MAI *ká-* (Farmer 2015, 40); KOR *kʰaa-* 'DEM.DIST' (Cook and Criswell 1993, 31). Metzger (1998) suggests that *ka-i / ka-ɨ* and *ka-o* have given rise to the attested forms *ki*/*kɨ*/*kɨ̃ɨ̃* and *ko*/*kõ*/*kõõ*, and this is repeated in Chacon (in

[26] The reconstruction of the bilabial in dependent pronouns in the Tukanoan family is discussed more extensively in Chacon and Michael (2018).

[27] The reconstruction of the feminine person marker as simply **-p'ɨ-õ* is given here in the scenarios that either: i) **-õ* was the result of **ĩ-o* and this had already merged by the time that the feminine person marker developed; or ii) **-õ*, and more specifically the nasalisation of this vowel, had another source, i.e. analogy with masculine **-p'ɨ-ĩ*. Otherwise, a reconstruction as **-p'ɨ-ĩ-o* is more suitable.

prep.). I would add that respectively voiced *gɨ* and *go* may have been the result of a similar complex form.

There are two compelling arguments in favour of Metzger's (1998) analysis that the PT gender markers containing a velar developed as complex forms:

i) there are some language-internal variations of these gender markers without the velar, i.e. P. SEK *mama-kɨ* 'son', *mama-ko* 'daughter', cf. *ʦ̃ĩw-ɨ* 'boy', *ʦ̃ĩw-o* 'girl', which suggests that **-ɨ* and **-o* had already come to be associated with nominal gender opposition before fusing with **-ka*.

ii) feminine classifiers *-ko, -go*, and *-o* may have developed by analogy with the PT marker **-ka-ɨ*, which originally fulfilled an animate singular function, unmarked for masculine/feminine gender according to Chacon (in prep.).

The abovementioned reconstruction of *-kɨ* as PT *-ka-ɨ* 'DEICTIC-ANIM.SG' is particularly interesting as it sheds light on the development of the vowels **ɨ* and **o* into a gender system, where **ɨ* originally was only contrasted to inanimates and non-singular number, and then acquired a new meaning as a non-feminine marker once a second animate gender **o* 'F.SG' developed to contrast it. A similar development of an animate/inanimate distinction to a gendered subdivision of animates through the addition of a feminine marker is reconstructed for the Indo-European family (Meillet 1923, 944).

TAN has a neuter gender *-a/-ka* which may also be marked on determiners in agreement with masculine nouns (Eraso 2015, 325). This is highly interesting since, in all probability, the marker *-kɨ* initially denoted an *animate class*, unmarked for masculine/feminine gender (Chacon, in prep.), such that *-ka* may have initially contrasted with it as a *neuter class*. The development of neuter *-ka* into a non-feminine agreement marker may in turn have been the result of extensive contact with the Yukuna (Arawakan) language, a contact situation which has been hypothesised to have yielded a number of language changes, including the convergence of the closed-central vowel **ɨ* with /i/ (see Aikhenvald 2002, 55; Arias Alvis et al. 2021). The retainment TAN *-ka* as a neuter marker contrasted to *-ki* and *-ko* is an argument in favour of either **-ka* or **-k'(V)* as the source of the gender markers, because we may analyse *ka* as the result of PT **-ka* 'DEICTIC', unmarked for masculine/feminine gender, as opposed to PT **-ka-ɨ* > TAN *-ki*, marked for masculine gender, and PT **-ka-o* > TAN *-ko*, marked for feminine gender.

With regard to the laryngeal properties of the velar, nothing decisive can be said on the basis of TAN, since this language has no voiced velar, as PT laryngealised velar **k'* has converged with PT **k* in this language, yielding synchronic /k/ uniformly, except for word-initial **k'*, which has been lost (Chacon 2014).

The voiced reflexes of PT **-kaɨ* –e.g. E. SIO *-g̱ɨ* (Wheeler 1970, 122), DES *-gɨ* (Miller 1999, 35), TUK *-gɨ* (Ramirez 1997, 201)– are not the regular reflexes of PT **k*, but rather of **k'*, such that Chacon (2014, 294) reconstructs the feminine classifier as **-k'o* rather than **-kao* or **-ko*. Therefore, the reconstruction of **-ka* is somewhat problematic, and it may be better to assume a now obsolete PT suffix **-k'(V)*, consisting of a laryngealised velar and perhaps an unattested vowel of unknown origin.

In TAN, both *-ka* and *-a* are found, but it is not clear how this must be understood since MAI, and supposedly[28] KOR (Chacon, in prep.) also have a neuter marker of the form *-a* (see Farmer 2015, 40–42 for MAI; Chacon, in prep.), albeit perhaps not synchronically productive. In MAI, PT word-initial **k'* has yielded voiced /g/ in some reflexes, and has been lost in others

[28] I have not found any instances of KOR neuter *-a* in Cook and Criswell (1993) nor Cook and Gralow (2001).

(Chacon 2014). Therefore, MAI *-a* could correspond to PT *-k'a*. In KOR, however, PT word-initial **k'* has not been lost altogether, but forms the sole context where an unaspirated /k/ has developed, in contrast with all other reflexes of PT **k* and **k'* which have yielded aspirated /kʰ/ (ibid.), such that one would expect PT **-ka* to yield KOR *#-kʰa*, and PT **-k'(V)* to yield KOR *#-k(V)*, instead of *-a*. Therefore, forms in these languages cannot be taken as conclusive proof for either PT **-ka* or PT **-k'(V)* as the source of the velar.

Metzger (1998) remarks that *ka-/-ka* in Arawakan and Tukanoan languages fulfil the same functions, "viz. a deictic specifier substituting for a known referent, and as a marker of specificity" (id., 44). The fact that the velar is widely found in those gender markers with singular referents but to a lesser extent with plural **ã/a*, and never with generic **e* is explained if the velar developed from the deictic specifier **ka-/-ka* since **e* consistently occurs on indefinite pronouns and unspecified gender, and **ã/a* lacks gender specificity as well. However, it is not necessarily the case that plural **ã/a* has an indefinite function so that it is not immediately clear that **ka-/-ka* were incompatible with it. Nevertheless, inanimate plurals are consistently marked with the suffix *-je*, which seems to have developed from an indefinite marker (see Table 7). It is then plausible that Tukanoan systems –or the PT system– at one point had a system of definite singulars and indefinite plurals. Since there is a corelation between definite singulars and those gender markers containing the velar, I will assume that the velar is indeed a reflex of PT **-ka*.

In conclusion, the discussion shows that the gender markers can be reconstructed as complex forms, consisting of a vowel denoting the gender category of the marker, and a consonantal onset governed by morphosyntactic locus of the marker. I have also indicated that the system originally consisted of an animate/non-animate distinction, that the animate gender would develop into a masculine or non-feminine category upon the development of the feminine marker, and that the markers containing a velar may have contained the deictic marker **ka*. We return to this issue in Section 6.3, where I argue that a similar development (i.e. the addition of additional categories to contrast with PT **-ka*) gave rise to the earliest classifiers of family.

Both the introduction of PT **ka* and the gender categories developed subsequently may have been the result of an Arawakan influence. Below, I discuss this influence in more detail.

5.2 Arawakan influence

As mentioned above, the function of the deictic marker *ka*/*ka-*/*-ka* is consistent across the Arawakan and Tukanoan families (Metzger 1998, 44). Mainly on this basis, Metzger (id., 40) suggests that **ka* may have had an Arawakan origin, which Chacon (in prep.) connects to the nominalizer and subordinator *-ka*, reconstructed by Ramirez (2001) for the Japura-Colombian branch of the Arawakan family. This is an interesting observation, given that the Japura-Colombian branch shares some features of classifiers with the Tukanoan family, e.g. agreement of classifiers on numerals and other quantifiers (Aikhenvald 1994). In terms of gender, there are some interesting similarities between the Arawakan and Tukanoan languages which may indicate an Arawakan origin, or influencing factor, for the development of the gender oppositions found in the Tukanoan systems of today (see also Chacon 2017).[29]

Both Arawakan and Tukanoan gender markers demonstrate a two-way gender distinction between feminine and non-feminine that probably goes back to Proto-Arawakan (Aikhenvald

[29] See also Aikhenvald's (2000a, 70–75) discussion of Paumarí (Arawan) *-ka*. According to Robert M. W. Dixon (in Aikhenvald 2012, 301), this class can be reconstructed for Proto-Arawan, and it would be interesting to pursue whether the *ka-/-ka* markers identified by Metzger (1998) may hail from an Arawan language originally.

1999, 87–88), and these are marked both on verbs and on nouns. As discussed above, the animate class in the Tukanoan family has a similar two-way distinction between feminine and non-feminine (Chacon 2017, 153–54).[30] Moreover, the gender markers may have been fused with number in Proto-Arawakan (Aikhenvald 1999), as they are also in all Tukanoan languages.

The Arawakan and Tukanoan genders show some formal similarities as well: Aikhenvald (ibid.) reconstructs feminine singular prefixes **thu-* or **u-* and suffixes **-thu* or -**-u*, and non-feminine singular prefixes **ɾi-* or **i-* and suffixes **-ɾi* or *-i*, and Ramirez (2001) reconstructs feminine singular prefix **ru-* and suffix **-nu*, and masculine singular prefix **li-* and suffix **-ni*. Chacon (in prep.) notes that the rounded closed-back vowel in Proto-Arawakan feminine singular **u* possibly corresponds to the rounded mid-back vowel **o* in PT, and that the Proto-Arawakan non-feminine singular unrounded closed-front vowel **i* possibly corresponds to PT unrounded closed-front nasal vowel **ĩ* or the unrounded closed-central oral vowel **ɨ*.

The correspondence of Arawakan feminine singular **u* to PT**o* is not unlikely, as Ramirez (2001, 465) only reconstructs a single Proto-Arawakan back or rounded vowel phoneme **u*. Therefore, many Arawakan languages neutralize [u] and [o] (Ramirez 2001, 163–65), such that an Arawakan form **u* may have been acoustically close to PT **o*, although it must be noted that PT may have had an opposition in height between the two rounded back vowels, e.g. PT **p'utu* 'termite' cf. **p'ot'ika* 'aracu fish' (Chacon 2014).

The correspondence between Arawakan **i* and PT **ĩ* or **ɨ* is slightly more problematic. Ramirez (2001, 465) reconstructs a Proto-Japura-Colombian phonemic opposition between front **i* and central **ɨ*, such that Arawakan *i* > PT **ɨ* is not likely. Tukanoan nasalised verbal person markers are not limited to the masculine gender, i.e. DES, TUK, TAT reflexes of Proto-Nucluear-ET masculine **-bĩ*, and feminine **-bõ* and Chacon and Michael (2018, 70, 85) claim that nasalised /õ/ in the feminine form is the result of the gender marked determiner *ĩ-o* 'she'. Here the vowel **o* may be the result of the Arawakan form **-thu*, **-u*, **-nu*. If the masculine PT form **bĩ* has a similar source, e.g. PT **p'ɨ-ĩ-i* 'TAM-3-NF', the final *-i* may have been borrowed from Arawakan **-ɾi*, **-i*, **-ni* 'non-feminine'.

The Proto-Japura-Colombian plural marker **na-* or **-na* might be analysed as the source for Tukanoan nominal animate plural marker *-dã* found in the entire ET branch and KOR (also put forward by Chacon 2017, 253).

On the basis of the similarities observed above, I follow Chacon (2017; in prep.) in arguing that the Tukanoan gender markers show traces of extensive formal as well as functional Arawakan influence. None of the nasal gender-number suffixes found in Table 5 has a velar onset. Therefore, it is plausible that the Tukanoan masculine suffixes *-kɨ* and *-bĩ* have two different sources: **ka-ɨ*, a possibly Arawakan nominalizer with the animate singular marker *-ɨ* found in the independent pronouns attached to it, which is synchronically found in the nominal classifier system of many languages; and **p'ɨ-ĩ-i*, which possibly contained a loan from Arawakan **-ɾi*, **-i*, **-ni* 'non-feminine'. The vowels and their possible Arawakan sources are summarized in Table 8 below.

[30] Some Tukanoan languages in the WT branch seem to have a non-masculine in addition to or instead of a non-feminine gender, but this distinction is also found in some Arawakan languages, e.g. Lokono (Pet 1987, 33–34; Rybka 2017, 263).

Table 8: Possible Arawakan sources of the reconstructed PT forms

PT form	Reconstructed meaning		Arawakan source	Reconstructed meaning
**ka*	Deictic or nominalizer	<	*-ka*	Nominalizer
**i*	Nominal masculine	<	*-ɾi, -i, -ni*	Masculine, non-feminine
**o*	Nominal feminine	<	*-thu, u, nu*	Feminine
**dã*	Animate plural	<	*-na*	Plural

5.3 The PT gender system

On the basis of the observations made above, it is possible to present a tentative reconstruction of the PT gender system, and the developments that occurred to form the synchronic Tukanoan gender systems.

As mentioned before, Chacon (in prep.) argues that the PT gender system at one point contained a two-way animate-inanimate distinction, such that we are presented with the development pattern of the Tukanoan gender-number system in Table 9, where singular number functioned as a 'marked' category, i.e. nouns or pronouns unspecified for number would also have been able to take the non-singular marker *-je*.

Table 9: Development of Tukanoan gender-number marking

	SG	**NSG**	>	**SG**	**NSG**	>	**SG**		**NSG**
							NF	**F**	
ANIM	*-ɨ*	*-je*		*-ka-ɨ*	*-je*/-(*r*)*ã*		*-ka-ɨ*	*-ka-o*	*-*(*r*)*ã*
INAN	*-ka*			*-ka*	*-je*		*-ka*/*-je*/-CL		*-je*

The abovementioned nuance regarding the markedness of singular number is of importance to explain some synchronic functions of *-*(*j*)*e*, e.g. KUB, where it fulfils an inanimate mass noun function (19a) (Chacon 2012), and E. SIO, where it is glossed as "collective inanimate/general" (Bruil 2014, 145), and which are also employed in agreement with functions on demonstratives and other parts of speech that do not productively take classifiers (19b) (id., 138; Bruil 2016, 8) (see also Section 3.1).

(19) a. KUBEO (EASTERN TUKANOAN)

i=***e***	*oko*	*bẽ*	*bo-e-bu*
DEM.INAN.PRX=INAN.MASS	water	well	white-INAN.MASS-COP.3SG.ANIM

'This water is very white (clean)' (Chacon 2012, 240)

b. ECUADORIAN SIONA (WESTERN TUKANOAN)

*i-**je***	*k̨ɨ̃na-ɾo*
DEM.PRX-CL:GEN	hard.material-CL:FLAT.ROUND

'This pan' (Bruil 2016, 8)

In addition to the bilabial stops **p*/*p'* and the velar stops**k*/*k'*, we also find a number of forms with alveolar /t, d, r/, e.g. animate plural *-~ra*/*-~da* (Table 5), neuter *-ru*/*-ri*/*-di*/*-ti*, generic *-di*/*-ti*/*-te*, and locative *-du*/*-do*/*-~ro*/*-to* (Table 6). The vowels involved in these suffixes have to a large extent been discussed above (Sections 5.2.1 and 5.2.2).

Chacon (in prep.) remarks that reflexes of PT **ti* are found in a large number of Tukanoan languages as an anaphoric marker, e.g. MAI *ĩ-ti* (Farmer 2015), TUK *ti* (Ramirez 1997), MAK *i-ti* (Smothermon, Smothemon, and Frank 1995, 35), DES *i-ri* (Miller 1999). I follow Chacon (in

prep.) who suggests that the generic or perhaps indefinite marker *-*e*, and the locative deictic marker *-*o* fused with the PT anaphoric marker **ti*, yielding the attested generic and locative forms mentioned in Table 6 above.

In light of the discussion regarding the voicedness distinctions above, it is interesting to note that the reflexes of **ti* suggest that the alveolar stop occurred in an intervocal position, consider Table 10, based on sound correspondences identified by Chacon (2014).

Table 10: Possible reconstructions of the alveolar in neuter, generic and locative forms

	Context	**C. SIO**	**MAI**	**TAN**	**KUB**	**DES**	**MAK**	**TUK**	**PIS**
**t'*	[_	*ɖ*	*ʔd*	*r*	*d*	*d*	*r*	*d*	*d*
	V_V	*t*	*t*	*t*	*d*	*d*	*d*	*Vʔt*	*t*
**t*	[_	*t*	*t*	*t*	*t*	*t*	*t*	*t*	*t*
	V_V	*t*	*t*	*t*	*d*	*d*	*t*	*t*	*t*
GENERIC/NEUTER		*-ru*	*-ti*		*-ri* *di*	*-di*	*iti*	*-ti* *ti*	*tie*
LOCATIVE		*-du*		*-to* *to*	*-rõ* *no*	*-do*		*-to* *too*	
ANIMATE PLURAL		*-kó'á* *wa'í*	*-na*	*-~ra*	*-dã* *-wa*	*-rã*	*-rã* *-a*	*-rã* *-a*	*-rã*

As Table 10 shows, the voiced reflexes found in C. SIO, KUB, and DES disallow a reconstruction as morpheme-initial plain **t*, and the voiceless reflexes found in the other languages in the table disallow morpheme-initial laryngealised **t'*. Moreover, the facts that MAK has a voiceless stop, and TUK does not display a laryngealisation of the preceding vowel disallow a reconstruction as intervocal **t'*. The fact that C. SIO shows [r, d] is therefore confusing. The E. SIO reflexes present a similar problem for the reconstruction, displaying both voiced *-ɾo* and plain (slightly aspirated) *-to* (Bruil 2014).

I will leave it for future work to discuss whether these findings indicate that bound forms display another set of sound-correspondences, as Chacon (2014) indicates for reflexes of **t'*, or that the Siona forms have another origin than the other Tukanoan languages, a theory that might be reinforced by the diverging vowel in C. SIO. On the basis of the probable bound form cognates identified in this thesis and those identified in Chacon (ibid.), one may distil a number of tentative sound correspondences (see Appendix C). However, the regularity of these sound correspondences cannot be judged fully due to the limited number of cognates identified in this thesis.

Finally, note that languages that display voiceless alveolars in reflexes of **ti-i* and **ti-o* in Table 10, show voiced alveolars in the animate plural markers, i.e. TAN *-to*, cf. *-~ra*, MAK *-iti*, cf. *-rã*, TUK *-ti*, *-to*, cf. *-rã*. This may indicate that the alveolar in these forms has a different origin from the generic, neuter, and locative markers. It is also interesting to note that there is a strong correlation between the nasality of the vowel and the occurrence of the alveolar onset in these forms, which may indicate that the alveolar found in the animate plurals is originally a nasal. This is another argument in favour of the theory that the animate plural form entails a borrowing from Arawakan *-na* (see Section 5.2). Note furthermore that WT forms such as C. SIO *wa'í* are probably a later development, limited to the Siona and Sekoya languages, and which co-exists with the mass noun *wa'i* 'meat, animal, fish' (Bruil 2014).

In conclusion, the gender-number system found in Tukanoan languages probably developed from a system that distinguished only between animacy and between singular and non-singular number. The animate singular gender split into a non-feminine and feminine gender, and at some point, possibly under influence of Arawakan gender distinctions, the forms of which were likely borrowed into the Tukanoan family as verbal person markers. An animacy split also occurred in the non-singular paradigm, where the inanimate non-singular form *-je* may have replaced the inanimate singular form *-ka* as well (see Table 9). The animate plural marker *-~ra* may have developed from Arawakan plural marker *-na*. The generic and neuter suffixes *-ti*, *-di*, *-ri*, and the locative suffixes *-to*, *-do*, *-rõ* probably developed from the anaphoric deictic marker *ti-*, or possibly from the interrogative 'which'.

Once such a system was in place, it could be expanded on, and especially the inanimate class was expanded to a highly elaborate system of 'specific' classifiers. However, if we suppose that the system originally contained only one inanimate singular class *-ka*, which was then expanded by a number of specific classifiers, we would expect the marker *-ka* to appear with all nouns that are not classified by another marker. This is exactly what we find in the Tukanoan family.

6 Classifiers

Some Tukanoan languages have far more nominal classifiers than others, e.g. ranging from around a hundred in MAI (Farmer 2015), DES (Miller 1999), and TAT (Gomez-imbert 2007), to smaller numbers, such as C. SIO and KUB, both of which have seventeen (Johnson and Levinsohn 1990, 28–30; Chacon 2012, 242). Some languages employ repeaters, e.g. DES (Miller 1999, 4), whereas others do not, e.g. E. SIO (Bruil 2014). Some languages have multisyllabic classifiers, whereas others only have monosyllabic classifiers, and in some languages, there are many classifiers that can be linked etymologically to nouns in the same language, whereas some languages only have classifiers with rather opaque etymologies.[31]

Given the processes involved in grammaticalization, as discussed in Section 2.2, I will assume here that the classifiers that grammaticalized at the earliest stages of the family will be monomoraic, have relatively broad meanings, and may not have an easily recognizable source as a noun (see also Grinevald 2004, 1028). Moreover, cognate classifiers that are found in languages of separate subbranches of the family are expected to have been present at an earlier stage of the family. As was discussed in Section 5, the gender markers are found in all Tukanoan languages, and can be reconstructed up to a certain point. Some classifiers in the Tukanoan languages are found throughout the family as well and can be reconstructed, often for a sub-branch (viz. PWT or PET), but sometimes also for PT. As will be shown below, these are often monomoraic morphemes with broad meanings and relatively opaque etymologies. However, for some of these classifiers, it is possible to posit etymologies, and this is exactly what I endeavour to do in this section. Therefore, it is probable that the Tukanoan gender markers are older than the classifiers, and indeed this is also argued in Chacon (2021).

Classifiers, especially in languages that have large amounts of them, can be grouped into certain semantic categories. Barnes (1990, 275), for example, recognizes ten categories: shapes; consistencies; times; botanical items; anatomical concepts; geographical concepts; containers; manufactured items; arrangements; and collections. Miller (1999, 37) recognizes nine

[31] Since these observations are based on other linguists' work, they are somewhat skewed by individual authors' analyses. I attempt to use the unified definitions discussed in Section 2.1, but this issue could not be resolved completely.

categories: generic; abstract; shapes; masses and arrangements; botanical items; geographical concepts; dissociated parts; associative concepts; and designs. Generic classifiers and shapes are cross-linguistically most associated with semantic bleaching (Greenberg 1972, 34–35; Adams and Conklin 1973), and in the Tukanoan languages, these are also the classifiers with the most opaque etymologies. Moreover, as is argued by Adams and Conklin (ibid.), it may be these categories that develop first in an emergent classifier system, and below I argue that this is also the case for the Tukanoan systems.

In the subsequent sections, I discuss some classifiers in four three broad categories: hydronyms and toponyms (Section 6.1); botanical and anatomical terms (Section 6.2); and generic classifiers and shapes (Section 6.3). I will show that hydronyms and toponyms have relatively recognizable etymologies, that it is somewhat more difficult to determine the etymologies of botanical and anatomical classifiers, and that shape and generic classifiers have very opaque etymologies. Problems in the reconstruction of botanical and anatomical classifiers are mainly due to differences in meaning between different languages, where a number of metaphorical and metonymical semantic expansions or shifts may be responsible for the variation. Problems in the reconstruction of shape and generic classifiers, on the other hand, are mostly due to language-internal disparate meanings and a lack of identifiable lexical sources, which is probably a result of their antiquity. On the basis of these observations, I identify three different layers in the development of Tukanoan noun categorization, viz. in chronological order: i) gender markers; ii) shape and generic classifiers; and iii) specific classifiers, such as botanical and anatomical terms and hydronyms and toponyms.

I end this section with a reconstruction of the mechanisms that are responsible for the classifier systems found today, and I argue that the shape and generic classifiers discussed in Section 6.3 originated as an expansion of the gender system, and were later supplemented by specific classifiers such as those discussed in Sections 6.1 and 6.2. Lastly, I refer to Appendix B which summarizes likely classifier cognates discussed in this thesis.

6.1 Hydronymic and toponymic classifiers

Hydronymic and toponymic classifiers are probably a relatively recent development in many Tukanoan languages. I have been able to tentatively reconstruct some classifiers for the PT stage, but others, especially toponymic classifiers, I could only reconstruct up to the PWT, PET, or even Proto-Nuclear-ET stages. Furthermore, the meanings of many of these classifiers can be related to lexical sources in nouns fairly easily, and they display little formal variation between languages. Therefore, it may also be possible that these classifiers were borrowed rather than inherited. However, in what follows, I will assume a scenario in which these classifiers are inherited wherever a reconstruction is possible.[32]

6.1.1 Hydronymic classifiers

I have been able to reconstruct two hydronyms to PT: **-tj'ia* 'CL:RIVER'; and **t'a* 'LAKE'. It is probable that **t'a* did not yet fulfil a classifier function in PT since, in a number of Tukanoan languages, its possible reflex is found in a complex form, seemingly containing the locational marker /ro/, which is also found in a number of other Tukanoan classifiers. There is also some

[32] It would be very interesting to consider these toponymic and hydronymic classifiers in light of the migration paths that the different Tukanoan groups may have undertaken, and possible contact phenomena that may have caused the spread of classifiers throughout the branch. Toponymic and hydronymic classifiers are inherently local in meaning and it may therefore be possible to recognize specific areas to which certain classifiers pertain. However, these questions are best approached interdisciplinarily, and as such fall outside the scope of this thesis and I will leave them for future consideration.

evidence for the claim that **-ja* spread throughout the family after some languages had branched off already, because the likely lexical source of this classifier can be reconstructed to PT **tj'ia* 'river', but this does not systemically yield the reflexes found in the family.

The forms of both these morphemes are interesting since Chacon (ibid.) reconstructs PT **tj'ia* 'river', and **tj'itta* 'lake'. Reflexes of both classifiers are presented in Table 11.

*Table 11: Reflexes of PT *-ja 'CL:RIVER', and *t'a 'LAKE'*

	'River'		**'Lake'**		**Ref.**
	CL	**Example**	**CL**	**Example**	
PT	**-ja*		**t'a*		
C. SIO	*-ja*	*gãtɨ-jã* 'Putumayo river'	*-da*	*hai-da* 'big lake'	I
E. SIO	*-ja*	*bĩʔã-ja* 'Bird river'	*-ɾa*	*gũʔhĩ-ɾa* 'garlopa fish lake'	II
E. SEK	*-ja*	*sao-ja* 'rushing river'	*-ra*	*kũʔhi-ra* 'lake where the Guajara fish'	III
P. SEK	*-ja*	*tʂia-ja* 'river'	*-ra*	*waʔi-ra* 'fishing lake'	IV
KOR	*-d͡ʐa*	*ora-d͡ʐa* 'Río de Chonta'	*-ra*	*d͡ʐia-ra* 'lake'	V
MAI	*-ja*	*mɨ́tò-ja* 'Tobacco River'	*-ra*	*jàò-rà* 'mud puddle'	VI
TAN	*-riá?* *-ríá?*				VII
DES	*-ja*	*dĩbã-ja* 'Poison River'			VIII
TUK			*-ra*	*doé-ra* 'Traíra lake'	IX
KOT			*-taro*	*parí-taro+re* 'lake'	X
KUB	*=ja*	*ihi=ja* 'Vaupés river'			XI
PIS			*=tara*		XII
BSA	*-sa*[33] *-ja*	*kõbẽ-jã* 'Metal River'	*-ra*		XIII
TAT	*-ja*	*bɨkɨ́-ja* 'old river'			XIV
TUY			*-táro*		XV

That the classifier for rivers *-ja* is not simply a synchronic grammaticalization can be seen from the fact that in some languages, there is a difference in form in the noun 'river', and the classifier. Compare Table 11 to: E. SIO *ʂia-* (Bruil 2014, 161); KOR *d͡ʐia-* (Cook and Gralow 2001, 179); DES *dia* (Alemán, López, and Miller 2000, 20); TAT *ɾia* (Gomez-Imbert 1982, 204; Bostrom 1998, 36). The free forms listed here are in concordance with the sound correspondences identified by Chacon (2014, 301–4), but the bound forms show some deviations from these. Following Chacon (ibid.), the phonological context where **tj'* directly precedes the closed front vowel /i/ provides a context where MAI yields /j/, rather than /d/ (see Table 20 and Appendix C), and where the ET languages all yield /r/ or /d/, with the exception of KUB, which yields /h/ in both contexts.

All languages –except for TAN– show *-ja*/*-d͡ʐa*, which suggests that these forms cannot be directly linked to PT **tj'ia* 'river', since the only proto-sound that would yield the approximant found in all these languages is morpheme-initial **j*.[34] Additionally, the front vowel /i/ is dropped in all languages except for TAN, which may also be the result of a phonological erosion process.

[33] I found no examples containing the fricave /s/ and will therefore leave its discussion for future work.

[34] The problem identified here is similar to the problem in the scenario I propose for **-tjũ* > **-jĩ*/*-*jũ* in Section 6.3.1.

Since intervocalic **j* would yield fricatives /c, s, h/ in a number of ET languages, the analysis of **ja* as a PT classifier would require a complex reconstructed noun, in order to explain the morpheme-initial reflexes of **j*. One such an explanation may involve PT anaphoric marker **ti*/*t'i*, with reflexes in a number of languages, i.e.: E. SIO *ti* (Bruil 2014, 252); MAI *íti* (Farmer 2015, 45); KUB *di-* (Chacon and Michael 2018, 76); and TUK *ti-* (ibid.).[35] In other words, the classifier **ja* likely had morpheme-initial **j* at the PT stage, and the noun **tj'ia* 'river' may itself be a morphophonological simplification of Pre-PT ***ti-ja*/*t'i-ja* 'ANA-CL:RIVER'. In this case, the classifier **-ja* would be older than the noun **tj'ia*.

Alternatively, one may suppose a loan from one language where PT **tj'*(*i*)*a*/*-ja* had yielded **-ja*, after the point where these sound laws took place. The fact that TAN did not follow the same phonological reduction process as the other languages, and retains a form close to its free form *riáká* 'river', suggests that the classifier is synchronically derived, as is indeed claimed by Eraso (2015, 348). Therefore, I note a question mark with the TAN reflexes listed in Table 11.

Regarding the reflexes of **-t'a* 'lake' (Table 11), some ET languages demonstrate a bisyllabic form *-taro*/*-tara*, containing /ro/ and in the case of PIS, /ra/. Alternatively, these may be reflexes of the locative **to* (see Section 5), which may suggest that **-t'a* had a relatively lexical status up to a relatively recent stage, partly due to the variation between languages, and partly because the addition of *-t'o* may indicate that **t'a* contained a derivation **-to* itself. That these bisyllabic forms are originally complex is further corroborated by the fact that these are not the only hydronyms and toponyms containing /ro/ (Table 12).

Table 12: Spatial classifiers (containing) -to/-do/-ro

	CL	Gloss	Example	Translation	Ref.
E. SIO	*-to*/*-ro*	PLACE	*ai-ro*	'forest'	I
KOR	*-ro*	PLACE/TIME	*ai-ro*	'forest'	II
MAI	*-to*/*-ro*	PLACE/TIME	*bàì-ro*	'homeland'	III
DES	*-goro*	CLEARED FIELD	*wi'i-goro*	'place where a house used to be'	IV
BSA	*-rodo*	PERIOD OF TIME			V
	-godo	USED/EMPTY AREA	*wíi-godo*	'cleared area for a house'	
TUY	*-kotó*	PREVIOUS HOUSE SITE OR VILLAGE SITE			VI
	-páro	PATIO			

Table 12 shows that classifiers related to spatio-temporal concepts containing what may be reflexes of PT **-to*. The ET languages in this table all show stacking of **-go*, **-ro*, **-pá*, to which a form **-to*/*-do*/*-ro* is attached. Therefore, the bisyllabic forms *-taro* and *=tara* may have originally contained this classifier as well, such that we can identify a PT classifier **-t'a* 'CL:LAKE'. The classifier **t'a* may be a phonologically reduced grammaticalization of PT **tj'itta* 'lake', but since the first syllable of **tj'itta* and **tj'ia* 'river' are rather similar to one another, it is also imaginable that **tj'itta* was already a complex form, consisting of **tj'i*(*a*) 'river?', bearing the lake classifier **-t'a*, which was then lexicalized as simplex morpheme, or

[35] I thank Thiago Chacon for suggesting this analysis to me.

if we reconstruct PT ***ti-ja*/*t'i-ja* 'ANA-CL:RIVER', we may reconstruct ***ti-ta*/*t'i-ta* as well (Thiago Chacon 2021, p.c.).[36]

It is yet unclear what is the etymology of the spatial classifier *-to*/*-do*/*-ro*. Jones and Jones (1991, 4) provide one example with a full noun *roho* 'place' in BSA, which might be a reflex of a PT noun from which these classifiers could have developed, but I have not found any cognates of this word in other languages.[37] It is also not entirely clear whether these forms are in fact diachronically related to the feminine gender as Chacon (in prep.) argues. As will become apparent upon comparison with the shape classifiers in Table 24 in Section 6.2.3, the spatial forms are separate from a near homophonic shape classifier which further obscures the etymological past of this classifier.

6.1.2 Toponymic classifiers

I have been able to reconstruct one toponymic classifier for: PET *-~*páia* 'CL:FLAT'[38]; *-~*ba* 'CL:PATH', which must have grammaticalized at a recent stage of the ET branch; PWT *-*p'o* which seems to have had a meaning 'CL:CAVELIKE' or 'CL:HOLE'; and one PT classifier *-*p'a* 'CL:FLAT.SURFACE'. The fact that many of these classifiers can only be reconstructed for a certain branch of the family may be an indication that toponymic classifiers occurred in a relatively recent stage, when the branches had already split off.

In some ET languages, the classifier for 'path' *-~*ba* has been extended semantically to include rivers as well. This observation has also been made for individual languages, such as TUY (Barnes 1990, 282), and TAT (Gomez-imbert 2007). This classifier has not been reported for any WT language. Classifiers of the form *-~ba*/*-~wa* are listed in Table 13.

*Table 13: Reflexes of Proto-Nuclear-ET *-~ba 'CL:PATH'*

	CL	Gloss	Example	Translation	Ref.
DES	*-bã*	PATH	*ɨgɨ̃ wa-ri-bã*	'his going trail'	I
BSA	*-bã*	PATH, STRING-LIKE			II
TUY	*-bã*	PATH, RIVER			III
PIS	*=bã*	PATH			IV
KOT	*-~ba*	RIVER, STREAM	*pitá-~ba*	'river port'	V
TAT	*-~wa*	PATH			VI

The semantic shift is likely due to the fact that two separate forms co-existed originally, viz. PT *~*p'aʔa* 'path', and *~*p'a* 'creek', i.e. TUK *maʔâ* 'path, line', cf. *maâ* 'river, creek', and but where homophony arose in languages that lost the glottal stop, e.g.: MAI; KUB; DES; BSA; MAK; TAT; KAR; TUY. However, in KOT, the noun *~baa* is translated as 'river, stream' by Stenzel (2013, 126), which is an indication that this noun and the classifier *-~ba* are synchronically associated with one another, even though did retain the glottal stop, i.e. *~bàʔá* 'path', cf. *~bá(a)* 'creek' (id., 459). We may therefore either suppose that the classifier either grammaticalized after the individual langauges lost **ʔ*, or speakers perceived a relation to the noun at the moment when the glottal stop was lost. In either case, grammaticalization of *~*ba* must have occured at

[36] However, this theory does not explain why many Tukanoan languages have voiceless alveolars in their reflexes of the free form **tj'itta*, which is probably why Chacon (2014) originally reconstructed a geminated alveolar **tt* here.

[37] Thiago Chacon (2021, p.c.) suggests that TAN *rõʔo* 'place' (Eraso 2015, 408), and KUB *-no*/*-rõ* 'pronoun, nominalizer' may be cognate to BSA *roho*.

[38] This is not technically a toponymic classifier, but I discuss it in this section because it is somewhat similar in form and meaning to PT *-p'a* 'CL:FLAT.SURFACE', although I argue against cognacy of these two forms.

a relatively recent stage, since within the ET branch, LET, DES, SIR, TUK, and KOT show some retention of the glottal stop (see Chacon 2014, 297, 314–15), such that grammaticalization must have occured after these languages had split off.

Many WT languages –but also ET KUB and TUK– have a classifier for 'flat surface', with a form that may be reconstructed as PT *-*p'a*, or *-*wa*. These classifiers are associated with vertical surfaces, such as walls, but are also found in combination with nouns referring to horizontal surfaces (Table 14).

*Table 14: Possible reflexes PT *-p'a 'CL:FLAT.SURFACE'*

	CL	Gloss	Example	Translation	Ref.
C. SIO	*-ba*	FLAT SURFACE	*wea-ba*	'cornfield (on a vertical flat surface)'	I
E. SIO	*-bã*	WALL	*tɨhtĩ-bã* *nẽa-bã*	'river bank' 'black wall'	II
E. SEK	*-pa*	FLAT SURFACE	*tĩtɨ-pa*	'ravine, rock'	III
P. SEK	*-pa*	FLAT	*wea-pa*	'cornfield'	IV
KOR	*-pa*	FLAT, VERTICAL SURFACE	*kʰad͡za-pa*	'bamboo plank, shelf'	V
KUB	*=wa*	FLAT	*kɨra-i=wa*	'floor, ground'	VI
TUK	*-pá* *-wa*	STRETCHED OUT	*suʔó-wa*	'wicker sieve'	VII

The nasality in E. SIO *-bã* is not found elsewhere. One may suppose a certain influence by aforementioned *-*~ba* 'CL:PATH', where one can imagine a semantic change 'path' > 'river' > 'riverbank'. However, it is unlikely that this classifier would lose its nasalisation in all other cognates with the meaning 'flat surface'. Therefore, I reconstruct a distinct classifier PT *-*p'a* (see Appendix C for sound correspondences in of bound **p'*).

There is another classifier PET *-*~páia* 'CL:FLAT' with reflexes: BSA *-hãi*; and TAT *-*~pái*; MAK *-hãi*; and TAN *-*~ɸáí*, and which is perhaps related to a PT noun *~*páíá* with reflexes such as TAN *~ɸáíá* 'slice' (see also Eraso 2015, 350). Since the PET form is nasalised, one may suspect that the E. SIO form *-bã* 'CL:PATH' is cognate to this classifier, but this hypothesis requires an explanation for why E. SIO would be the only WT language that inherits it from PT, and there is no regular sound change PT *~*p* > E. SIO *-~b*. Therefore, I do not assume that they are related.

Lastly, a number of WT languages share a classifier that can be reconstructed as PWT *-*p'o*, and which refers to cave-like constructions, including notches and areas with overarching structures, such as banana plantations, where interlocking foliage creates a shadow-rich surface. Compare the forms in Table 15.

*Table 15: Reflexes of PWT *-p'o 'CL:CAVE-LIKE/HOLE?'*

	CL	'Plantation'	'Mouth'	'Axe'	Ref.
C. SIO	*-bo*	*ˈo-bo*	*jio-bo*	*ʂu'u-bu*	I
E. SIO	*-bo*	*nohka-bo*	jeoʔ-bo	*ʂɨʔ-bo*	II
E. SEK	*-po*	*noka-po*	*jiʔo-po*	*ʂũʔu-po*	III
P. SEK	*-pa*				IV
KOR	*-po*	*oo-po*	*jɨʔo-po*	*suʔu-po*	V
TAM	*-po*			*su-po*[39]	VI

I have found no similar classifiers in any ET languages. As opposed to some of the classifiers discussed above, the forms listed here present the sound correspondences predicted by Chacon's (2014, 289), such that it is fairly certain that these classifiers where inherited from PWT. It is interesting that many authors list 'axe' as one of the examples of this classifier, even though it might not be immediately associated with cave-like constructions. Wheeler (2000, 185), Johnson and Levinsohn (1990, 29), and Cook and Crisswell (1993, 21) all explain this by offering that this classifier is oriented toward the hole where the haft fits into the axe head. That this classifier seems to have been associated with axes at the PWT stage can now be combined with archaeological insights of the material culture of the area, such that it might be possible to pinpoint with greater precision the temporal frame when these languages were still together.

In conclusion, the Tukanoan hydronymic and toponymic classifiers have probably developed in a relatively recent stage, because the toponymic classifiers cannot easily be reconstructed further back than the PWT and PET stage, because most of the meanings of hydronymic and toponymic classifiers are fairly easily connected to nouns, and because these classifiers may not have been inherited into all the individual languages in which they are found, but might have been borrowed instead, since their forms show little variation between languages.[40] One probable exception is the locative classifier *-to*/*-do*/*-ro* which seems to be very archaic since it is found throughout the family, has an opaque etymology, and shows much language-internal formal variation. This is also in line with what Chacon (in prep.) reconstructs among the gender markers as early as the PT stage (see Section 5).

6.2 Botanical and anatomical classifiers

The botanical and anatomical classifiers I have been able to reconstruct for PT seem to be more archaic than the hydronymic and toponymic classifiers mentioned above. Many of these seem to be found across the family's branches, although it is difficult to determine with certainty whether the forms in the different branches are in fact cognates.

The meaning of some reconstructed classifiers and of lexical source are difficult to identify since these forms seem to be associated with conceptually distinct nouns. In some situations, a metonymic, metaphorical, or otherwise semantic expansion may have given rise to new meanings, i.e. BSA *-i* 'CL:TREE' > *bahi-i* 'net-CL:TREE', where the classifier has probably extended in meaning to include woven objects as these are often made from palm trees (see Section 6.2.1). Other classifiers seem to have more than one source, the forms of which converged phonologically such that they became homophonic, or near-homophonic (see

[39] Note that Creveaux, Sagot and Adam (1882) do not provide gloss-boundaries, nor an explanation of the orthography used. Therefore, my gloss here is an approximation based on Creveaux' <soupo>.

[40] If the classifier *-ja* was borrowed into the family from another language, possible sources may be Quechuan, i.e. Inga *jáku* 'river', or Chibchan, see for example Tunebo *ríʔja* (Huber and Reed 1992, 49).

Section 6.2.2). Afterward, language users may have reanalysed these originally separate classifiers as having some conceptual connection.

Lastly, some classifiers may have had an abstract meaning originally but upon the introduction of new, more specific classifiers for subcategories of the concepts originally classified by these abstract classifiers, these abstract classifiers came to denote a number of disparate concepts (see Section 5.3 and Section 6.3), after which they may have been reanalysed as having some conceptual connection again.

6.2.1 Trees, plants, and hammocks

I have been able to reconstruct three forms that classify trees: PWT **-jɨ̃* (Table 16); PET **-kɨ/-k'ɨ* (Table 17); and PET **-jũ/-jõ* (Table 18). A likely etymology for PET **-kɨ/-k'ɨ* is PT **tjũkkɨ* 'tree' (see also Chacon 2012, 257). This is further discussed in Section 6.2.1.1. The other two classifiers for trees –PWT **-jɨ̃,* and PET **-jũ/-jõ*– may have this noun as their source as well, although this hypothesis is somewhat problematic because KUB has a classifier **-jɨ̃* 'CL:PALM' in addition to a classifier *-kɨ* 'CL:TREE' (id., 247–49), such that it may be necessary to assume a borrowing scenario from PWT **-jɨ̃* into KUB **=jɨ̃*, and PET **-jũ/-jõ*.

The conceptual grouping of trees and hammocks in ET languages furthermore reveals an interesting developmental scenario of classifiers in these languages, where the choice of classifiers becomes slightly lexically –rather than semantically– based, and when new, more specific classifiers enter a language, original conceptual groupings by classifiers become obscured, further entrenching a lexically based classification in the older class.

As shown in Table 16, the reflexes of PWT **-jɨ̃* are both morphologically and semantically very close to one another. Only P. SEK lacks the nasal quality found in all other WT languages (Vallejos 2021a), and MAI demonstrates a front vowel /i/ rather than the central vowel /ɨ/ found in all other WT languages (Farmer 2015).[41]

*Table 16: Reflexes of PWT *-jɨ̃*

	Form	**Example**	**Gloss**	**Translation**	**Ref.**
PWT	**-jɨ̃*				
C. SIO	*-jɨ̃*	*nãdãhã-jɨ̃*	orange-CL:HAS.ROOTS	'orange tree'	I
E. SIO	*-jɨ̃*	*ĩ'si-jɨ̃-ã*	pineapple-CL:TREE-PL	'pineapple trees'	II
E. SEK	*-ñɨ*	*pa'pa-ñɨ*	conambo-CL:TREE	'conambo palm'	III
P. SEK	*-ʤɨ*	*wea-ʤɨ*	corn-CL:PLANT	'corn plant'	IV
MAI	*-ñi*	*sűkí-ñi*	tree-CL:TREE	'tree'	V
KOR	*-ñɨ*	*ɨje-ñɨ*	jungle.grape-CL:HAS.ROOTS	'jungle grape tree'	VI

In some WT languages, the same form is found in nouns for teeth, e.g. C. SIO *gũhi-jɨ̃* (Wheeler 1970, 136), KOR *kõhi-ñɨ* 'tooth' (Cook and Gralow 2001, 39). Authors on these languages have therefore glossed the classifier as 'having (permanent) roots' (see Wheeler 1970, 136; Cook and Criswell 1993, 21). I do not have access to information regarding the classifiers associated with teeth in other WT languages. In ET languages, teeth have other classifiers: e.g. KUB, 'pointed' *kõpi=jo* (Chacon 2012, 248); KUB *upî-ka* 'tooth'; TAT 'rounded *~òpí-a* (Gomez-Imbert 1982, 122; Huber and Reed 1992, 4); MAK 'hollow, tubular' *guhi-ga* (Smothermon and Smothemon 1993, 41; Smothermon, Smothemon, and Frank 1995, 37). Therefore, the use of the 'tree' classifier for teeth may have been a later development, which only took place after

[41] Note that the same vowel correspondence is found in MAI demonstrative and non-third person present verbal paradigms (Farmer 2015, 40, 53).

the WT branch developed, and may therefore have involved a semantic expansion of the concept for trees, for being "planted into the gums", as it were, exactly as was suggested by Wheeler (1970, 136), and Cook and Criswell (1993, 21).

Section 6.2.1.1 discusses **-kɨ/-k'ɨ* and its probable connection to PT **tjũkkɨ*. Section 6.2.1.2 discusses those classifiers referring to palm trees specifically, and their connection to woven objects. Lastly, Section 6.2.1.3 discusses the possibility of **tjũkkɨ* as a common source for PET **-kɨ/-k'ɨ*, and **-jũ/-jõ*, and PWT **-jɪ̃*, and possible borrowing scenarios.

*6.2.1.1 *-kɨ/-k'ɨ* and **tjũkkɨ* 'tree'

The reflexes of PET **-kɨ/-k'ɨ* are slightly more diverse, the voicedness of the velar does not reflect the sound correspondences proposed by Chacon (2014), as we have also seen with regard to the gender markers. Moreover, some languages –viz. TAN, LET, DES, and KOT– have a multisyllabic suffix that can be reconstructed as PT **tjũkkɨ* 'tree' (ibid.), rather than simply a reflex of **-kɨ/-k'ɨ*. This multisyllabic classifier has a form identical or nearly identical to synchronous nouns in a number of languages, e.g. DES *jukɨ* 'tree', cf. *-yukɨ* 'CL:TREE.PL', from which it may have been derived. Chacon (2012, 257) suggests that KUB *=kɨ* 'CL:TREE' that language's noun *hokɨ* 'tree'. If we maintain that both **jɪ̃* and **-kɨ/-k'ɨ* developed from the same noun, a justification for the difference in form must be given.

The fact that the various reflexes of **-kɨ/-k'ɨ* do not reflect the synchronous nouns with PT **tjũkkɨ* as their origin (e.g. the voicedness of the velar in DES *-gɨ*, cf. the voiceless velar in the noun *jukɨ*), the fact that the monosyllabic classifiers have undergone more extensive erosion (viz. multisyllabic and tonal roots, eroded to monomoraic classifiers, unspecified for tone), and the fact that the monosyllabic classifiers have undergone more extensive semantic bleaching (viz. 'trees' > 'any rectilinear object') suggest that the monosyllabic classifiers have undergone grammaticalization before the multisyllabic classifiers did. Note also that there is no productive agreement between the LET classifier *-hũkia* 'CL:TREE' and the head noun, whereas other classifiers do fill an agreement slot in the language. Consider for example the numeral *ĩʔrã-* 'one' in (20), which takes the "general shape classifier" *-bi* 'CL:1D instead of *-hũkia* 'CL:TREE' (Strom 1992, 37, 54).

(20) LETUAMA (EASTERN TUKANOAN)

ĩʔrã-bi	*yapua*
one-CL:1D	tree
'One tree'	

(Strom 1992, 37)

Lack of participation in agreement processes, such as demonstrated by LET *-hũkia* 'CL:TREE' may also point towards the relatively recent grammaticalization of the multisyllabic classifier. It is interesting that LET synchronically has an unrelated form *yapua* 'tree', which could be an Arawakan loan (Rose et al. 2017). Reflexes of **tjũkkɨ* may persist in the synchronically simplex word *õterikia* 'fruit tree' < PRE-LET **õte?-rikia* (Strom 1992, 61), although only laryngealised PT **tj'i* > LET *ri*, whereas plain **tj* > LET *h* (Chacon 2014, 302). Lastly, it is interesting to note that where the two both the mono- and multisyllabic forms persist synchronously in the same language (viz. in DES and KOT), it is the monosyllabic form which denotes the singular number, and the multisyllabic form which denotes the plural. This may be an effect of the individualizing function of classifiers. As mentioned, individualization and singularization are not identical, but in DES it has taken on a singularizing function (de Lima Silva 2012, 117).

The multisyllabic forms have morpheme-initial sounds that are associated with word-initial position, i.e. PT **tj* > DES and KOT *j* word-initially, but *s* intervocalically, whereas the monosyllabic forms can be better related to intervocalic reflexes of PT **k*, i.e. DES *-gɨ*, rather than *#-kɨ*, and KOT *-kʉ*, rather than *#-kʰʉ* (Chacon 2014). There are some reflexes that have voiced realisation of the velar that are not found in Chacon (ibid.), e.g. TUY, PIS, TUK, but Chacon (ibid.) does not discuss voicedness in these languages. On the basis of sound correspondences of bound forms and earlier works (i.e. Chacon 2012, Chacon and Michael 2018), I propose a number of regular sound correspondences in bound morphemes (see Appendix C).

On the basis of these correspondences, it is likely that the monomoraic classifiers underwent grammaticalization at a fairly early stage –and given the uniformity of the forms and meanings associated with them, this may well be as early as PET– whereas the multisyllabic classifiers may have grammaticalized at a later stage, after these languages split off. Therefore, I have indicated the multisyllabic forms with a question mark in Table 17 below.

*Table 17: Reflexes of PET *-kɨ/-k'ɨ*

	Form	**Example**	**Translation**	**Ref.**
PET	**-kɨ/-k'ɨ*			
TAN	*-~hukí*? *-~húkí*?	*hó'ba-~hukí ~kárá-ka* big-CL:PLANT star.apple-N	'big star apple tree'	I
LET	*-hũkia*?			II
DES	*-gɨ* (SG) *-jukɨ* (PL)?	*gãri-gɨ* sugar-CL:TREE.SG *pũ-jukɨ* hammock-CL:TREE.PL	'sugar cane' 'hammocks'	III
KUB	*=kɨ*	*hokɨ=kɨ* tree=CL:TREE	'tree'	IV
BSA	*-ɨ* *-gɨ* *-kɨ*	*ruh-ɨ* torso-CL:TREE	'tree trunk'	V
MAK	*-kɨ* *-gɨ*	*kũba-kɨ* laurel-CL:TREE	'laurel tree'	VI
TUY	*-gɨ*	*ati-gɨ-dɨpɨ* DEM.PRX-CL:TREE-CL:BRANCH	'the branch of this tree'	VII
PIS	*=gɨ*	*betʃu=gɨ* ceremonial.staf=CL:RECTILINEAR	'ceremonial staff (spec.)'	VIII
KAR	*-kʉ*	*di-kʉ jukʉ* which-CL:TREE tree	'which tree […]?'	IX
KOT	*-kʉ* (SG) *-jʉkʉ* (PL)?	*sé-kʉ́* cucura-CL:TREE.SG	'cucura tree'	X
TUK	*-gɨ* *-khɨ*	*ũju-gɨ* avocado-CL:RECTILINEAR	'avocado tree'	XI
TAT	*-ɨ* *-kɨ*	*~kádéè-ɨ* *star.apple*-CL:RECTILINEAR	'star apple tree'	XII

6.2.1.2 Palm trees

In many languages in the ET branch, there is an additional classifier for palm trees, only lacked in TAN, which employs the 'tree' classifier *-~hukí/-~húkí* for palms as well (Eraso 2015, 348).

The WT languages also employ their respective 'tree' classifiers for palms. Words for different palm types are listed in Table 18, most of which were taken from Huber and Reed (1992, 175), except the E. SIO example, which comes from Bruil (2014, 139), the E. SEK example, which comes from Johnson and Levinsohn (1990, 29), the KOR example, which comes from Cook and Criswell (1993, 21), and the MAI example, which comes from Velie and Velie (1981, 19).

Table 18: Classifiers associated with palm trees

	Classifier	**Example**	**Translation**
E. SIO	*-jɨ*	*wa'ho-jɨ̃*	'type of palm (sp.)'
E. SEK	*-ñɨ*	*kõsa-ñɨ*	'hungurahua palm (*Oenocarpus bataua*)'
KOR	*-jɨ̃*	*ɨne-jɨ̃*	'chontaduro palm (*Bactris gasipaes*)'
MAI	*-ñi*	*bidi-ñi-hã*[42]	'wiririma palm (*Astrocaryum jauari*)'
TAT	*-jõ*	*ɨ̃de-jõ*	'chontaduro palm (*Bactris gasipaes*)'
DES	*-jũ*	*ɨ̃rɨ̃-jũ*	
TUK	*-jõ*	*ɨ̃rẽ-jõ*	
KOT	*-õ*	*ɨ̃rẽ-õ*	
KUB	*=jɨ̃*	*ɨ̃rẽ=jɨ̃*	
PIS	*=wõ*	*ɨ̃dẽ=wõ*	
BSA	*-jõ*	*ɨ̃dẽ-jõ*	
TAN	*-hũki*	*ɨ̃rẽrita-jũki-a*	
MAK	*-jõ*	*hata-jõ*	
KAR	*-jõ*	*ɨ̃dẽ-jõ*	
TUY	*-wõ*	*ɨ̃dẽ-wõ*	

The ET classifiers *-jõ*, *-jũ*, *-õ*, and *-wõ* may be analysed as reflexes of PET **-jũ/-jõ*, although for as far as I am aware, there is no structural sound-correspondence between ET /j/ and PIS, TUY /w/, such that these forms may have another source, although the labialisation of the approximant might be the result of influence of the rounded vowel. Strikingly, KUB employs the form *=jɨ̃* identical to that found in the WT languages, even though KUB's 'tree' classifier *=kɨ* is different from that found in the WT branch. Chacon (2012, 247) glosses *=jɨ̃* as a shape classifier for 'hollow objects', and lists a number of non-palm nouns that also employ this classifier: e.g. *koa=jɨ̃* 'a pan'; *dori=jɨ̃* 'fish trap'; *hapu-i=jɨ* 'flute'. No other ET language has a similar overlap, such that the KUB classifier *=jɨ̃* 'palm tree' may have been semantically extended to refer to hollow, mostly elongated objects, perhaps since palm trees can, in Chacon's (2012, 247) words, "become and be used as hollow".

Since the classifiers **-jɨ̃* 'tree' in PWT, and **-jũ/-jõ* 'palm' in PET languages are somewhat similar in form, and in the case of KUB identical, it is interesting to consider whether they may have a common source. This is especially interesting since Chacon (2012, 349) remarks that it is unexpected to see the 'tree' classifier *=kɨ* marked on nouns such as *pãu=kɨ* 'hammock', and *ˈpapi=kɨ*[43] 'fishnet', since these products are made from the material of palm trees, such that one would expect the palm classifier *=jɨ̃* on these nouns. This is a pattern found across the ET branch, as is demonstrated in Table 19.

[42] Velie and Velie (1981) list many words with the morpheme *-hã*, which probably is an inanimate copula (id., 121; Farmer 2015, 92–93).

[43] Chacon (2012, 349) erroneously cites this word as "*ˈpapi=ka* 'fishnet=CL:3D'". The correct form *ˈpapi=kɨ* is found in Huber and Reed (1992, 86), and in Morse, Salser, and de Salser (1999, 245).

Table 19: Classifiers associated with woven objects

	Classifier	**Examples**	
		'Hammock'	'Net'
E. SIO	*-dɨ*	*hãĩ-dɨ*	
C. SIO	*-dɨ*	*hãõ-dɨ*	*wãte-dɨ*
E. SEK	*-rɨ*	*hãɨ-rɨ*	
P. SEK	*-rɨ*		
KOR	*-rɨ*	*hão-rɨ*	
MAI	*-rɨ, -tɨ*	*hãĩ-rɨ*	*mate-rɨ*
TAT	*-ɨ*	*pũ-ɨ*	
DES	*-gɨ*	*pũ-gɨ*	*wẽhẽdi-gɨ*
TUK	*-gɨ, -khɨ*	*pũ-gĩ*	*wẽhẽ-khɨ*
KOT	N/A	*pɨ̃dõ*	*wãjɨ̃dõ*
KUB	*=kɨ*	*pãu=kɨ*	
PIS	*=gɨ*	*pũ=gɨ*	
BSA	*-ɨ, -gɨ, -kɨ*	*hũɨ-gɨ*	*bahi-ɨ*
TAN	N/A	*pãũ-a*	
MAK	*-gɨ*		*jori-gɨ*
KAR	*-ʉ*	*pũ-ʉ*	
TUY	*-gɨ*	*pũũ-gĩ*	*bapi-gɨ*

There is quite some overlap between the classifiers associated with woven objects and those associated with trees in the ET languages, and those languages that have divergent classifiers for woven objects, are those that use gender for these nouns –i.e. feminine gender in KOT (Waltz 2007, 382, 409), and neuter in TAN– or lack the more recently grammaticalized classifier *-jukɨ*, as is the case in DES. The same is not true for WT languages, which bear classifiers that can be reconstructed as **-t'ɨ*.

PWT **-tɨ/-t'ɨ* 'CL:WOVEN.OBJECTS' is likely a reflex of the PWT verb **tɨõ-* 'to weave': C. SIO *tɨõhi* 'to weave' (Wheeler 1987, Vol. 2, 88); MAI *tɨ̃ãji* 'to weave a hammock or bag' (Velie and Velie 1981, 56–57); KOR *tʰʉomʉ* 'to weave, to mould' (Cook and Gralow 2001, 116), or if not, the verb and the classifier most probably have a common lexical source.[44] Again, this sheds some light on the voicing of voiceless obstruents in the onsets of bound forms, which, as has been pointed out, do not always follow the patterns identified by Chacon (2014, 294).

The fact that we find **-tɨ/-t'ɨ* in WT may indicate that the classifier **-kɨ/-k'ɨ* grammaticalized at the PET level –clarifying why the classifier is associated with two, synchronically seemingly disparate semantic fields– but not before that, since one would expect the same form to show up in the WT branch as well. Alternatively, it is imaginable that **-kɨ/-k'ɨ* was originally present in PWT as well, but was lost.

It is probable that PT **tjũkkɨ* 'tree' grammaticalized into PT or PET **-kɨ/-k'ɨ* fulfilling a noun-formative function, and presumably originally associated exclusively with trees, and later expanding semantically to woven objects as well. Since woven objects are made from palm trees specifically, this scenario indirectly presupposes lack of a specific 'palm' classifier **-jũ/-jõ* at the stage of PET where the semantic expansion of **-kɨ/-k'ɨ* to include woven objects took place, as one would expect **-jũ/-jõ* to occupy this function. Alternatively, there may have been a classifier **-jũ/-jõ* which denoted something other than palm trees.

[44] As was pointed out to me by Thiago Chacon (p.c.), the PWT verb may entail a denominal verb, where the rounded back vowel *-o* fulfils a verbalizing function.

This analysis is of particular importance because it shows how a Tukanoan classifier, which is assigned on a pragmatic rather than lexical basis, does exhibit some lexical preference: A previously semantic grouping of concepts –i.e. palm trees, non-palm trees, and woven objects– must have shifted to a somewhat more lexical grouping of "noun classes", since the addition of another classifier specific for palms did not affect nouns referring to woven objects, which would be expected if the conceptual grouping was still based on semantics. In Section 6.3 below, I work out this notion in more detail.

Above, we have identified three interrelated forms that classify trees: PWT *-*jɨ̃*; PET *-*kɨ*/-*k'ɨ*; and PET *-*jũ*/-*jõ*. It is likely that PET *-*kɨ*/-*k'ɨ* arose from a reflex of PT **tjũkkɨ*, but etymologies for the other two forms are not easily discernible. It is possible that PWT *-*jɨ̃* and PET *-*jũ*/-*jõ* have a common source or that the ET branch borrowed the classifier from the WT branch, but it is also possible that these two forms are not related at all.

6.2.1.3 A common source?

Since the semantics of the 'palm' classifier are rather similar across languages, and since it is probable that classifiers start out with relatively specific meanings which are expanded over time, PET *-*jũ*/-*jõ* may have started out as referring to a specific type of palm tree, or as referring to palm trees in general. One possible lexical source for this classifier could be PT **jũkka* 'cumare palm (*Astrocaryum aculeatum*)' > C. SIO *ñuka* 'chambira (*Agave Americana Astrocosum*)' (Wheeler 1987, Vol. 2, 61), E. SEK *ñukua* 'cabuya, chambira' (Piaguaje et al. 1992, 94), KOR *ñuka* 'cumare' (Cook and Gralow 2001, 84), KUB *ñuka* 'cumare' (Morse, Salser, and de Salser 1999, 232), TUK *jõkâ-puri* 'tucum palm (*Astrocaryum vulgare*)' (Ramirez 1997, Vol. 2, 246), KOT *ñɨkɨ* 'tucum, or cumare palm'[45] (Waltz 2007, 173–74).

Possible arguments against this analysis are the fact that KOT demonstrates the central vowel /~ɨ/ lexically, but the rounded mid-back vowel /~o/ as a classifier, and that KUB demonstrates the rounded closed-back vowel /~u/ lexically, but the central vowel /~ɨ/ in its classifier. Another possible counter-argument against **jũkka* as the source for *-*jũ*/-*jõ* may be the fact that the palm species associated with the form **jũkka* are used in weaving practices (e.g. Wheeler 1987, Vol. 2, 61; Ramirez 1997, Vol. 2, 246; Eraso 2015, 338), such that one might expect this classifier to occupy the 'woven objects' classifier slot upon grammaticalization, in addition to the 'palm tree' classifier slot, although this would by no means have necessarily been the case.[46]

It is impossible to say with any certainty what is the source of PWT *-*jɨ̃* since the phonological evidence that this monomoraic morpheme contains is too scarce such that many possible lexical sources remain available, one of which may be PT **jɨ̃ka* 'leg' (Chacon 2012, 104), but reflexes of the classifier *-*jɨ̃* are not semantically associated with legs in any Tukanoan language. Alternatively, *-*jɨ̃* could be the result of grammaticalization of the first syllable of the root PT **tjũkkɨ* 'tree', which also yielded PET *-*k'ɨ*/-*kɨ*. In this scenario, the second syllable was deleted in order to render a monosyllabic classifier, and the onset cluster **tj* was simplified to **j*, a hypothesis which is strengthened by the reflexes of laryngealised PT **tj'* in WT languages, as summarized in Table 20, which is an adaption from Chacon (2014, 302–03).

[45] Waltz (2007, 173) translates this word into Spanish as "cumare", but cites the Linnean term associated with the tucum palm: *Astrocaryum vulgare*.

[46] As Table 19 shows, the WT branch does not use a reflex of **jũkka* 'cumare palm', which may provide evidence that **jũkka* only grammaticalized into its classifier in PET, if that is indeed the source of PET *-*jũ*/-*jõ*.

*Table 20: Reflexes of PT plain *tj and laryngealised *tj' in WT languages, adapted from Chacon (2014, 302–3)*

PT	Context	C. SIO	E. SIO[47]	C. SEC	MAI	KOR	KUE
**tj*	[_	*s*	*s̱*	*s*	*s* *tʃ*/_*i*	*s*	*t* *s*/_*e*
	V_V]$_{STEM}$	*s*	*s*	*s*	*s*	*s*	*t*
*tj'	[_	*s̱*	*s*	*s̱*	*d*	*d͡z*	*j*
	[_i	*s̱*	*s̱*	*s̱*	*j*	*d͡z*	*r* [l]

These reflexes show that PT **tj* was simplified as a simple onset in all WT languages in intervocalic position, and that **tj'* was simplified as a simple onset, where the SIO and SEK languages preserve the original laryngealisation in synchronic /s̱/, and the other languages show either /d/, /j/, or in the case of KUE [_i, /r/. Moreover, it is cross-linguistically common for biconsonantal clusters to delete the first consonant, rather than the second, when simplified in intervocalic contexts (e.g. Wilson 2001, 148), which would render **-j'ũ*, or possibly **-jũ*, if a language does not have laryngealised approximant in its consonant inventory. See Appendix C for a comparison and tentative reconstruction of consonants in bound forms in the Tukanoan family.

The vowel quality found in the WT reflexes of this hypothetical reconstruction provides a further problem since there is no reason to suppose a consistent vowel shift PT **ũ* > PWT **ɨ̃*. Waltz and Wheeler (1972, 136) do give a number of vowel shifts PT **u* > *ɨ*, but none that specifically affect the WT branch, nor in the specific context of **tjũkkɨ*, nor do they consider nasality. One may hypothesise that PWT **-jɨ̃* therefore would not be a clipped, but a contracted reduction of PT **tjũkkɨ* 'tree', such that it would have been the noun's final vowel that is retained, rather than the first syllable. However, I am not aware of any regular contractions of this kind in these languages. Therefore, I cannot conclude decisively that PT **tjũkkɨ* is indeed the lexical source of the PWT classifier **-jɨ̃*. A better understanding of the diachrony of vowels in the family will shed more light on this issue.

ET *-jõ*/*-jũ*/*-wõ*/*-õ* may entail a loan from WT *-jɨ̃* 'CL:TREE'. This theory would explain the similarity of the WT forms to KUB *=jɨ̃* 'CL:palm', but would require an explanation for the backing and rounding of the central vowel /~ɨ/ to / ~u/, and /~o/ in many ET languages. It is true that nasalised vowels often undergo a quality shift backwards (Schourup 1973, 201–3), and there is also some evidence that nasalised vowels shift downward (Ruhlen 1978, 229–30),[48] where the latter is specifically found in the absence of a conditioning nasal that would cause it phonologically (Beddor, Krakow, and Goldstein 1986),[49] such that the phonological reduction to a monomoraic, bound form may have prompted this change. However, it seems unlikely that the same vowel shift occurred in all the independent languages of the ET branch. Therefore, it

[47] Being published in the same year as Bruil (2014), Chacon (2014) does not yet include E. SIO data. However, he does mention some cognates on the basis of which he comes to his reconstructions, which allow us to determine what reflexes are found in E. SIO, e.g.: [_ PT **tj'oʔa* 'distant' > E. SIO *soʔo* (Criollo Quintero 2011); and [_i word-medial PT **tj'ia* 'river' > E. SIO *s'iaja* (ibid.).

[48] In terms of Element Theory, this may be explained by supposing a single element |L| which is responsible for both nasalisation and low vowels .

[49] This tendency is not found in languages where nasal vowels occur outside of phonological contstrains (Beddor, Krakow, and Goldstein 1986), something which does occur in Tukanoan languages, but which is not yet understood completely with regard to the underlying suprasegmental properties of affixes.

may have occurred at some proto-stage in the Nuclear ET branch, after KUB had split off, as to explain why KUB retains the central vowel. The hypothesis that Nuclear PET *-*jũ*/-*jõ* and KUB -*jɨ̃* entail a loan from PWT *-*jɨ̃* would shed an interesting light on past contact between the two branches of the family, and on KUB's as of yet elusive role in it. At this time, there is no decisive linguistic evidence for either scenario, although the loaning scenario may be slightly more convincing.[50]

In conclusion, there may be some reason to analyse PT **tjũkkɨ* as the source of all three classifiers, albeit indirectly through borrowing across two branches.

6.2.2 Seeds, eyes, and grains

A number of Tukanoan languages have a classifier for seeds -*ka*/-*ga*, and some ET languages have classifiers that seem to be reflexes of a root *-*kape*/-*japе*. Consider Table 21, where classifiers associated with seeds are listed in a number of languages.

Table 21: Classifiers associated with seeds

	Classifier	**Example**	**Translation**	**Ref.**
C. SIO	*-ga, -ka*	*ãõda-ka-dɨ*	'seed (diminutive)'	I
E. SEK	*ka*	*ora ka*	'chontaduro seed'	II
P. SEK	*-ka*			III
MAI	*-a, -ga*	*érí-gà*	'shapaja seed'	IV
KOR	*-a* (SG) *-ka-ã* (PL)	*ape-ka-ã*	'dice, marbles'	V
KUB	*=jabe*	*otéi=jábe*	'seed'	VI
TUY	*-pe*	*oteri-pe*	'seed'	VII
PIS	*=pe*	*bĩpĩ=pe*	'açaí seed'	VIII

It is interesting to note that there are some languages that do not have a classifier -*pe*, but which nonetheless have a word 'seed' that contains this syllable: TUK *oteripe*[51]; TAN *japе*; MAK *ahe*[52] (Huber and Reed 1992, 159). In some languages there is a classifier which is bisyllabic, containing a form /ka/ similar to that found in the WT branch, in addition to a form that may be a reflex of *-*pe*: BSA -*kahe* 'plants grown from fruits', and -*kahero* 'stalk, base of fruit' (Jones and Jones 1991, 54); and KOT -*kapa* 'shoot, seedling' (Stenzel 2013, 150).

I analyse PT *-*ka-ape* as an originally complex classifier containing -*ka* and -*ape*. My reasons for this analysis are as follows:

i) languages may have two independent classifiers -*ka*, and -*pe*;
ii) in TUY, KAR and TAT, the noun *kaapea* 'eye' contains a long vowel (Huber and Reed 1992, 6), suggesting Proto-Nuclear-ET **ka-ape*;

[50] It would be interesting to approach this question interdisciplinarily, including insights on the area's plant distribution and material cultures, i.e. some groups may be specialized in certain weaving techniques, which could further inform our understanding of the development and spread of these classifiers.
[51] The form *ote-* –also found in e.g. in DES *oteri*, TAT *oterike*, KAR *oterike* (Huber and Reed 1992, 159)– is probably aa reflex of PT **otte* 'to plant' (Thiago Chacon 2021, p.c.). It is therefore possible to analyse -*pe* as a separate morpheme in this noun, at least historically.
[52] The MAK and BSA reflexes of PT **p* are both /h/.

iii) KOR and E. SEK show an irregular insertion of a velar in the plural, which may be explained by assuming an originally complex form **-ka-ape*, where **-ka* is the inanimate morpheme discussed in Section 5.1.2. [53]

In the following subsections, I discuss these four arguments in more detail. (Sections 6.2.2.1–3). Lastly, Chacon (2012, 257) suggests PET **t'ɨka* 'fruit' as a possible source of KUB *=dɨ* 'CL:ROUNDED'. Since this classifier occurs with the same nouns as other languages' *-ka* and *-pe* (e.g. *jako=dɨ* 'eye' (id., 246)), one might suspect that the /ka/ in **t'ɨka* could also be the source of *-ka* in other languages. However upon closer inspection, this does not seem likely (Section 6.2.2.4).

6.2.2.1 Two separate classifiers within the same language

The classifier *-a*/*-ka*/*-ga*/*-gã* occurs as an independent classifier 'sphere' alongside *-pe* in a number of languages. Consider: TUY *-ga*/*-gã*, cf. *-pe* (Barnes 1990, 275); WA' I *-ga*/*-ka*/*-a*/*-dia*, cf. *-pe*; KOT *-ka*, cf. *-kapa* (Stenzel 2013, 123, 129); and BSA *-a*, cf. *-kahe*/*-kahero* (Jones and Jones 1991, 51, 54). Below, this opposition is demonstrated in TUY (21), and WA'I (22).

(21) TUYUKA

a. *bií-ro* *bií-wɨ* *ati-**gá*** *sɨ̃ɨ́-**gá***
like-ADV like-INAN.PST.EV DEM.PRX.SG-CL:3D.GEN snail-CL:3D.GEN
hɨ̃ɨ́-wa
say-IMP
'Say: "This snail was like that"' (Barnes 1990, 290)

b. *ati-**pe**-sotoʔa-pɨ*
DEM-CL:3D.NONSPHERE-on.top-LOC
'on top of that non-spherical thing' (Barnes and Malone 2000, 440)

(22) WA'IKHANA

a. *kɯ-**ga***
cassava-CL:ROUND.NPL
'cassava tuber' (Balykova 2019, 10)

b. *ihki-**pe***
inajá-CL:SEED
'*inajá* seed' (id., 7)

Notice that Barnes and Malone (1990, 275; 2000, 451) gloss *-pe* as a 'non-spherical' classifier (21), which is quite different in meaning from the classifier *-ka*/*-ga*/*-a*, which is often associated with round objects in Tukanoan languages, e.g. WA'I (22) (Balykova 2019). The fact that *-ka*/*-ga*/*-a* and *-pe* co-exist within a single language is an indication that forms like BSA *-kahe* have a complex form *-ka-pe* as their source.

6.2.2.2 A long vowel in independent forms

There are some languages where **ka-ape* has lexicalized as a noun (Huber and Reed 1992, 6), and grammaticalized into a classifier again (see Stenzel ibid., 121, 128).[54] I reconstruct a Proto-

[53] See Balykova (2019, 18–22) for a discussion of the various irregular plurals of *-a*/*-ka*/*-ga*/*-gã* in the Tukanoan family.
[54] See also Waltz (2007, 49).

Nuclear-ET noun **kaape* ‘eye’, with reflexes: TUY, *kaapea* (Barnes and Malone 2000, 450); KAR *kaapea*; TAT *kaapea*; BAR *kahea*; MAK *kahea*; KOT *kapariaka*; and TUK *kapea* (Huber and Reed 1992, 6). The long vowel suggests originally complex **ka-ape*, where **ka-* may have been the PT ‘inanimate singular’ marker discussed in Section 5.1.2.

The classifier **-ape* may be a development of a root PET **jape-a* ‘seed’, with reflexes: TAN *ɟàɸéá* ‘seed’ (Eraso 2015, 116); KOT *japa* ‘drop, seed’ (Waltz 2007, 320); TUK *japá* ‘pointy end’ (Ramírez 1970, Vol 2., 236).

6.2.2.3 Irregular plurals in KOR and E. SEK

Furthermore, the KOR classifier demonstrates an exceptional morphophonemic behaviour: when denoting singular number, the ‘small and round’ classifier takes the form *-a* (e.g. *ɨo-a* ‘lightbulb’), but when pluralized, it demonstrates an allomorph containing a velar stop and a nasalised vowel *-kʰã* (e.g. *ɨo-kʰa-ã* ‘lightbulbs’) (Cook and Criswell 1993, 26). A similar behaviour is reported for E. SEK (e.g. *ñako-a* ‘eye’, cf. *ñako kã* ‘eyes’) (Piaguaje et al. 1992, 90).[55] The nasality in these forms suggests that *kã* is actually underlyingly /ka-ã/, containing the plural suffix *-ã*. This allomorphy is evidence of an older property of the velar, which could be analysed as the same form *-ka* discussed in Section 5.1.2, where its function was at some point an ‘inanimate singular’, but which was broadened to also denote inanimate plurals in TAN, MAI, and KOR (see Chacon, in prep.).

This hypothesis would require an explanation for the loss of the velar in the singular allomorph. Recall from Section 2.2.2 that E. SIO and E. SEK both require individualization of inanimate nouns before these can be pluralized. The KOR and E. SEK allomorphy may then be explained by assuming that *-ka*/*-kʰa* developed from the inanimate marker **-ka*/*-ga*/*-a* discussed in Section 5. Since this form is analysed to have fulfilled an individuating function originally, the originally underlying form of the plural may have been **-kʰa-a-ã*, consisting of an individuating morpheme *-ka*, the classifier *-a* ‘CL:SEED’, and the plural marker *-ã* respectively.

Alternatively, the allomorphy may be explained by assuming that the classifier for seeds in these languages was originally the same as the inanimate marker, rendering the KOR and E. SEK plural forms simply **-kʰa-ã* and **-ka-ã* respectively, where the moraic quantity of the suffix group is even, which in the case of the bilabial triggered the fortis allophone PT **p*, with reflexes *p, b, h, ∅*. If the even number of morae in the plural is responsible for the form including the velar and the odd number of morae in the singular form *-a* is responsible, this could also provide some important insight into the allomorphy of other classifiers and the seemingly irregular sound correspondences of the bound forms discussed above (see Appendices).

Furthermore, the same irregularity regarding insertion of the velar stop is found in the KOR form associated with seeds, and with the E. SEK form associated with eyes. This is an argument in favour of the hypothesis that both classifiers share the same source. Consider Table 22, based on Huber and Reed (1992, 6), and expanded with Cook and Crisswell (1993, 19), Velie and Velie (1981, 44), and González de Pérez (2000, 391).

[55] In TUY, another allomorphy is found, where the ‘generally spherical’ classifier *-ga*/*-gã* has the plural form *-pa* (Barnes 1990, 275). Plural forms containing *-pa* also occur in three other irregular classifiers: Singular *-rɨ* ‘pot’, *-wɨ* ‘hollow interior’, and *-wa* ‘shallow container’ have the plural forms *-párɨ*, *-páwɨ*, and *-páwa* respectively (Barnes 1990, 280). See Section 6.3 for an analysis of PT **-pa*/*-p’a* as an old plural or ‘collective’ marker.

Table 22: Classifiers associated with 'eye', adapted from Huber and Reed (1992, 6)

	Classifier	**Example**
C. SIO	*-ga*	*jãko-ga*
E. SEK	*-a* (SG) *-ka-ã* (PL)	*jãko-a*
MAI	*-a*	*jãko-a*
KOR	*-a* (SG) *-ka-ã* (PL)	*ñako-a*
TAN	*-a*	*jãko-a*
BSA	*-a*	*kahe-a*
KUB	*=dɨ*	*jako-dɨ*
TUY	*-a?*	*kapea*
PIS	*=ga*	*kãpé-gà*
TUK	*-a*	*kaape-a*
TAT	*-a*	*kaape-a*

In conclusion, it is likely that there was a complex PT form **-ka-ape*, which contained the inanimate marker **-ka*, as well as a separate classifier **-ape*. In some languages these fused, and in other these classifiers still exist independently of each other. The source of the classifier **-ape* is not clear, but it can maybe be reconstructed to PET, where it may have grammaticalized from **jape* 'seed', although this hypothesis is a tentative one. In Section 6.3.1, I further discuss the development of *-ka*/*-k'a* into a shape classifier.

*6.2.2.4 Some remarks on PET *t'ɨka 'fruit'*

The form *=dɨ* 'CL:ROUND' found in KUB is interesting, since it is different from forms found in other languages. Chacon (2012, 257) identifies PET **t'ɨka* 'fruit'[56] as a possible source, suggesting a semantic expansion from fruits to eyes. This theory gains plausibility from the fact that PIS also has a single classifier for fruits and eyes: *ũpũ=ga* 'banana'; *ɨ̃dẽ=ga* 'peach palm fruit (*Bactris gasipaes*)'; *pika=ga* '*uaitutu* (forest fruit)' (González de Pérez 2000; Rodríguez Preciado 2018). One might suspect that to *-a*/*-ka*/*-ga*/*-gã* would then also be related to this source, but many languages where *-ga*/*-ka* is associated with a noun 'eye' have a separate form *-rika*/*-dɨka* with a noun 'grain of corn', which is formally rather close to PET **dɨka*, such that an explanation would be required for the co-existence of *-ga*/*-ka* and *-rika*/*-dɨka* in these languages. Compare Table 22 with Table 23, which was based on Huber and Reed (1992, 162), expanded with Bruil (2014, 146), Vallejos (2021a), and Eraso (2015, 338).

[56] This proto-form has reflexes in ET languages such as: DES *dɯka* (Alemán, López, and Miller 2000, 23); KOT *dicha* (Stenzel 2013, 25); TUK *dɨká* (Ramirez 1997, Vol 2., 43). I did not find any cognates in WT languages.

Table 23: Classifiers associated with grains of 'corn'

	Classifier	**Example**
E. SIO	*-ka*	*wea-ka*
P. SEK	*-ka*	*wea–ka*
TAN	*-ka*	*óá-ka*
DES	*-dɨka*	*oho-dɨka*
BSA	*-rika*	*oho-rika*
TUY	*-rika*	*ho-rika*
MAK	*-rika*	*oho-rika*
KAR	*-rika*	*o-rika*
TAT	*-rika*	*ó-ríka*

As shown by Tables 21 and 22, at least BSA, TUY, and TAT have two different classifiers for both semantic categories. Moreover, the forms found in the WT languages are also slightly different in both tables such that it is likely that these are in fact two near-homophonic classifiers in these languages, rather than reflexes of a single source.

In conclusion, it is implausible that the classifiers found in Table 21 all have **t'ɨka* as their source, although KUB *=dɨ* 'CL:FRUIT' probably does have this form as its source. The semantic expansion found in KUB *=dɨ* which came to refer to eyes as well may have occurred only in KUB, and perhaps also PIS, but it is also possible that at an earlier stage PIS had a separate classifier **=diga/=dɨga* 'CL:FRUITS', and another *=ga*, which denoted a category including eyes, where phonological erosion, perhaps instigated by some perceived conceptual similarity between the two categories, caused these classifiers to become homophonous at some point.

6.2.3 Flowers, ears, and concave things

Many Tukanoan languages have a classifier *-to/-do/-ro*, the meanings of which are often difficult to identify comprehensively. As mentioned in Section 5, and 6.1, it denotes meanings related to space, but in many languages, such as TAT, it also denotes objects with concave, or oblong shape, or a "not very defined shape", as well as flowers, ears, and certain other body parts and unrounded fruits which are not classified by *-ka/-ga/-a* (Gomez-imbert 2007, 419), and is employed as a general inanimate singular marker (ibid.). Moreover, it is also found in the verbal paradigms of some languages, where it has been analysed as a further grammaticalization of the classifier (ibid.; Bruil 2014, 77), and has developed as an animate singular marker in KOT (Stenzel 2013, 129–30).

However, upon closer inspection, the *-to/-do/-ro* classifiers denoting 'concave shapes' are probably not the same as those denoting spatial concepts. Consider Table 24, which summarizes nouns referring to pots and cups that are marked with classifiers similar to *-to/-do/-ro*.

Table 24: Nouns referring to pots and cups

	Concave objects			**Ref.**
	CL	**Example**	**Translation**	
C. SIO	*-do*	*gɨna-do*	'metal pot'	I
E. SIO	*-do*	*ɡ̰ɨna-do*	'metal pot'	II
E. SEK	*-ro*	*kɨna-ro*	'metal pot'	III
P. SEK	*-ro*	*ɡ̰ɨ̃na-ro*	'metal pot'	IV
KOR	*-ro*	*kɨna-ro*	'metal pot'	V
MAI	*-to, -do, -ro*	*kíú-dò*	'metal pot'	VI
TAN	*-hòtó*	*~í'ra-hòtó*	'one pot'	VII
DES	*-du* *-soro*	*koa-duka-du* *bãta-soro*	'gourd bowl' 'clay pot'	VIII
KOT	*-to, -ro*	*biá-tó*	'cooking pot'	IX
KUB	*=do*	*kopo=do*	'cup'	X
BSA	*-tʉ, -rʉ, -sotʉ*			XI
TAT	*-to, -~do, -ro*	*wáà-ro*	'calabash container'	XII
TUY	*-rɨ*	*bɨkɨ-rɨ*	'old pot'	XII

Comparison between Table 12 (Section 6.1.1) and Table 24 reveals a difference in form between the spatial classifier and the classifier associated with pots and cups, e.g.: E. SIO *-to*/*-do* 'CL:PLACE', cf. simply *-do* 'CL:FLAT.ROUND' (Bruil 2014, 143–44); and BSA *-tʉ*/*-rʉ*/*-sotʉ* 'CL:POT-LIKE', cf. the quantity, spatio-temporal, and manner nominalizer *-to*/*-do*/*-ro* (Jones and Jones 1991, 50–51, 141–42).[57] The disparate meanings of *-to*/*-~do*/*-ro* in languages like TAT may therefore be the result of homophonization of two classifiers with different lexical sources, i.e. a spatio-temporal marker and a classifier for pots. In addition to the distinct forms of 'spatial' and 'concave' marker, WA'I also demonstrates a distinction in morphosyntactic behaviour between to homophonous markers *-do* denoting 'CL:MISCEL' on the one hand, and 'CL:CONCAVE' on the other, where the former participates in nominal agreement (23a), whereas the latter does not (23b) (Balykova 2019, 7).

(23) WA'IKHANA (EASTERN TUKANOAN)

a. *yʉ'ʉ akaro su'tiro duu*

yʉ'ʉ	aka-**do**	su'ti-**do**	du-u
1SG	one-CL:MISCEL	clothes-CL:MISCEL	buy-VIS.PFV.1

'I have bought one piece of clothing.' (Balykova 2019, 6)

b. *yʉ'ʉ bahtido wa'mabahti duu*

yʉ'ʉ	bati-**do**	~wa'ba-**bati**	du-u
1SG	basket-CL:CONCAVE	new-CL:BASKET	buy-VIS.PFV.1

'I have bought a new basket.' (id., 7)

In WA'I, it can be observed that *-do* entails two homophonous markers because it does occur in agreement in its 'miscellaneous' function, and therefore occurs in agreement on the numeral *aka-* 'one' in (23a), but not in its 'concave' function, such that the repeater *-bati* 'CL:BASKET'

[57] One might suppose that Máihɨ̀kì provides another example of this, since Farmer (2015, 235) lists a plural *-noa* for 'concave things', but not for 'places'. However, the same plural form was found for 'places' (id., 78), such that these forms have an identical plural form.

from the head noun in (23b) is employed to mark agreement on the adjective *~wa'ba* 'new'. In other words, there are both formal and functional reasons to assume multiple sources for *-do*.

The source of the classifiers in Table 24 may be the PT noun **sot-* 'pot/ceramics/clay' identified by Chacon (2014), with reflexes: C. SIO *soto*; E. SEK *soto*; KOR *t^{h}ot^{h}o*; MAI *toto*; RET *hoto*; KUB *hodo*; DES *sodo*; BSA *cotʉ*; TUK *sɨtɨ*; TAT *hotɨ*; KOT *situ* (ibid.) One argument in favour of this scenario is the fact that BSA demonstrates the same deviant vowel /ʉ/ with respect to the other languages listed above, both in the noun *cotʉ* and the classifier *-tʉ*, *-rʉ*, *-sotʉ*.

It is interesting that some languages show two classifiers: One monosyllabic and one bisyllabic classifier, where the bisyllabic classifiers show reflexes of PT **soto* 'pot', in accordance with Chacon (ibid.), and it is probable that these are a later grammaticalization than the monosyllabic classifier. If the monosyllabic form grammaticalized from the same lexical source, it is possible that the disyllabic form developed due to the emerged homophone of the monosyllabic classifier for pots and the spatio-locative marker, which then lead to a re-analysis of the monosyllabic classifier as referring to a number of different semantic categories, such as ill-defined shapes.

This does not explain the entirety semantic fields covered by the classifier, i.e. 'unrounded fruits' do not straightforwardly fall into a 'spatio-locative', nor a 'concave' category. It is possible that there are other nouns that also became homophonous to *-to*/*-~do*/*-ro*, but which are now unidentifiable. However, since *-to*/*-~do*/*-ro* shares a number of properties with the gender markers (see Section 5), such as overlap with the verbal paradigm, and since it contrasts with the generic classifier *-ka*/*-ga*/*-a* on a number of points, it is plausible that this classifier was originally part of a PT gender system, where **-to*/*-t'o* and **-ka*/*-ga*/*-a* at some point both denoted some subcategory of the inanimate gender.

If this is indeed the case, it is probable that these classifiers denoted very broad semantic categories originally, and only received the current disparate status as the result of an addition of more specific classifiers into the system. In this light, it is even possible that both the spatial classifiers and the 'concave' classifiers have the same source after all, but became reanalysed as two separate classifiers, which might in turn explain their slightly different forms. The issue of gender markers such as **-ka*/*-ga*/*-a* and possible **-to*/*-t'o* with broadly defined meanings developing into classifiers with separate, more specified meanings is further discussed in Section 6.3.

6.3 Shape and generic classifiers

As discussed in Section 5, the Tukanoan gender system may have originally consisted of an animate-inanimate distinction, which also distinguished singular and non-singular number. At some point, the animate paradigm was expanded, adding a feminine gender and extending the animacy distinction to the non-singular number. As a result of the addition of the feminine gender, the non-feminine gender was reanalysed as masculine in a number of constructions, but was retained as an unmarked gender in a number of constructions.

A similar process may have occurred in the inanimate gender, expanding the inanimate gender with a number of more specific classes. As discussed above, gender markers and classifiers display a number of morphosyntactic distinctions in many Tukanoan languages. This fact lends evidence to the suggestion that the classifiers are a more recent development than the gender markers. This suggestion is also corroborated by the fact that many Tukanoan languages contain many classifiers that are not found in any other languages in the family, indicating that these particular classifiers may have developed after the point in time when that language split off from a shared ancestor. A few classifiers are shared by many different languages in the

family, and are found both in the WT and ET branches, indicating that they may have developed at the PT level.

Widely shared classifiers are generally monomoraic and relatively abstract, and if a language has a small number of classifiers, these are very likely be monomoraic and abstract as well. Recall from Section 2.2 that these are characteristics of grammaticalization processes: The process of erosion renders bimoric morphemes monomoraic, and the process of semantic bleaching renders more concrete morphemes more abstract. Moreover, in some languages, the more abstract classifiers share some morphosyntactic properties with gender markers in the language, which other, more specific classifiers do not share, e.g. E. SIO classifiers *-ɾo/-to* 'CL:PLACE', and *-ɾĩ* 'CL:TIME' have a relatively more extensive use in the grammar than other classifiers do (Bruil 2014, 138). Based on these observations, I assume that these classifiers grammaticalized at a relatively early stage of the family, where *-ka*/*-ga*/*-a* developed first (Section 6.3.1).

Based on the idea that classifiers originated as an expansion of a more archaic gender system, one would expect the classifiers to fulfil functions similar to that of PT **-ka* '(inanimate) singular', and to have meanings complementary to it, similarly to the way that the feminine gender marker came to fulfil a meaning complementary to the archaic animate marker *-kɨ*, which in turn came to denote masculine gender. This is indeed borne out in the data, where classifiers are mainly characterised by denoting singular inanimate concepts. Moreover, classifiers that are likely highly archaic (i.e. *-ka*/*-ga*/*-a*, *-to*/*-do*/*-ro*, and *-wɨ*) are likely to have stood in mutual opposition to one another in the earliest stages of the classifier system. In other words, when *-ka*/*-ga*/*-a* was in place, *-to*/*-do*/*-ro*, and *-wɨ* developed as its counterparts in the classifier system. This is further discussed in Section 6.3.2.

There are two notable exceptions of probably archaic classifiers which were originally probably not opposed to **-ka*, i.e. animate collective *-hɨ*/*-bɨ*/*-wɨ*, and *-je*. In Section 6.3.3, I argue that these did not develop as oppositions to inanimate singular classifier **-ka*, but already in the gender system, as opposition to singular animate and inanimate alike.

6.3.1 The antiquity of *-ka*/*-ga*/*-a*

Beside the fact that *-ka*/*-ga* is widely found, that its etymology is rather opaque, and that its form is monosyllabic, as is expected from the most archaic classifiers, there are two additional reasons for assuming that *-ka*/*-ga* was among the very first classifiers in the family:

i) KOT and WA'I have an allomorph of *-ka*/*-ga* of the form *-dia*/*-ria* 'CL:ROUND', which probably consisted of a nominaliser *-di* and + *-a* 'CL:ROUND' (Balykova 2019);
ii) TAN and LET *-ka*/*-ga* retains a relatively gender-like distribution (Eraso 2015; Strom 1992).

These observations suggest a unique position of *-ka*/*-ga*/*-a*, and lend further evidence to my claim that the Tukanoan gender system developed out of the gender system, i.e.: if *-ka*/*-ga*/*-a* is the oldest classifier in the family, it is likely that it is a development of the older, inanimate PT gender marker *-ka*/*-k'a*. Therefore, the abovementioned two claims are further discussed in Sections 6.3.1.1 and 6.3.1.2 below.[58]

[58] The discussion on WA'I, KOT, TAN, and LET that follows here is in large part based on observations in Balykova (2019).

6.3.1.1 -ka/-ga/-a in complex forms -di-a/-ri-a 'CL:ROUND'

In WA'I, the allomorph of *-ka*/*-ga*/*-a* 'CL:ROUND' that has the form *-dia* occurs the possession marker *yaa*, on the two nouns *bʉkʉ* 'old' and *~wa'ba* 'new', and on numerals (24a) (Balykova 2019, 10, 15). In KOT, the classifier *-ria* denoting "round and usually somewhat elongated" objects is found on some nouns, but it is the only inanimate classifier that occurs on numerals in modifier position, and even occurs when the noun with which the numeral agrees takes *-ka* 'CL:ROUND' (24b) (Stenzel 2013, 177–8). Moreover, Balykova (2019, 12) suggests that these nouns are also in fact nominalizations of respectively the verbs *wʉ* 'fly' and *phuti* 'blow'.

(24) a. WA'IKHANA
akaria vela
aka-**dia** vela
one-CL:ROUND spark.plug
'one spark plug' (Balykova 2019, 4)

b. KOTIRIA
phʉaria bʉhsoka
phʉá-**ria** bʉsó-ka
two-CL:ROUND canoe-CL:ROUND
'two canoes' (Stenzel 2013, 178)

The fact that WA'I *-dia* occurs specifically on the nouns *bʉkʉ* 'old' and *~wa'ba* 'new' is one of the reasons for Balykova (2019) to suppose that this marker developed from classifier *-a* in combination with the nominaliser *-di*, since these specific nouns were originally verbs. Since other classifiers did not fuse with the nominaliser –e.g. *~tudu-di-be'to* 'car (spin-NLZ.3NPL-CL:RING)'– and since nominal agreement probably initially involved nominalisation, the fusion of *-di* and *-a* is an indication that *-a* was among the first classifiers to develop.

The fact that KOT *-ria* 'CL:ROUND' is the only inanimate agreement marker on numerals in the language can be understood as an intermediate stage of a process that led to full numeral agreement systems in other languages. Since only *-ria* can occur on modifying numerals in KOT, it is probable that this marker was the first classifier to develop an agreement function in the language at all, i.e. assuming more than one classifier at that stage of the language would require an explanation for why only *-ria* developed into an agreement marker on numerals.

6.3.1.2 The retainment of a gender-like distribution of -ka/-a in TAN and LET

In TAN and LET, there is a probable cognate of the *-ka*/*-ga*/*-a* form: *-a*/*-ka* (Gomez-Imbert 2007, 421). However, in these languages, *-a*/*-ka* functions as an inanimate gender marker, and does not occur in the same slot on the noun as the nominal classifier does (25a). Moreover, classifiers cannot occur on nouns, whereas the neuter marker *-a*/*-ka* can and often does, just like the animate gender markers do.

(25) a. TANIMUKA
*hó'ba-**ó-ká*** *biá-Ø*
big-CL:3D-N pepper-N
'big, round pepper' (Eraso 2015, 244)

b. LETUAMA
bãẽkaraka-bi *rĩpi-**a***
three-CL:1D branch-N
'three branches' (Strom 1992, 55)

In (25b), it is shown that the noun *rĩpi* 'branch' must take a marker from the small gender-like class, while the numeral *bãẽkaraka* 'three' takes a classifier. These observations show that *-a*/*-ka* in TAN and LET do not pattern with classifiers. Since *-a*/*-ka* is a likely cognate with *-ka*/*-ga*/*-a*, and since Balykova (2019, 15-18) shows that the neuter meaning of the classifier did not develop out of a classifier 'CL:ROUND', it is likely that it was the other way around, i.e. the neuter marker developed into a classifier where it came to classify 'round' things. This is in line with my hypothesis that the Tukanoan classifiers developed out of the PT gender system.

It is likely that the TAN and LET classifiers display a different morphosyntactic distribution from the other Tukanoan languages because the TAN and LET classifier systems are a later development than those of the other Tukanoan languages, and it is possible that TAN and LET classifier systems these developed in contact with Yukuna (Arawakan), rather than under a Tukanoan influence. Balykova (2019, 17-18) shows that the classifier systems of TAN and LET are probably relatively recent developments, and this claim is further corroborated by my discussion of the TAN classifiers, i.e. *-ria* 'CL:RIVER', and *-~hukí* 'CL:TREE' (Sections 6.1.1, and 6.2.1).

6.3.2 Shape classifiers in opposition to each other

As discussed in Section 5, and in Chacon (in prep.), once the animate/inanimate paradigm was in place in PT, a more fine-grained distinction within the animate paradigm developed, specifying gendered distinctions **-ka-o* 'feminine, singular', and **-ka-ɨ* 'non-feminine, singular'. A similar expansion of the inanimate paradigm would therefore entail contrasting to the inanimate marker **-ka* 'inanimate, singular'.

In the animate paradigm, the contrast that emerged was feminine/non-feminine, a contrast widely found cross-linguistically (Corbett 2013). Within the inanimate paradigm, a similar cross-linguistic tendency exists, distinguishing three basic shapes –long, round, and flat– with a possible secondary parameters of rigidity, relative size, fullness, regularity of shape, part vs. whole, and with regard to the shape 'long', also horizontal vs. vertical, and "edgedness" (e.g. Adams and Conklin 1973, 5; Denny 1986, 303; Grinevald 2015, 811–12). Indeed, these categories are also found in the Tukanoan languages, and analysed as developments at a stage before the other classifiers in these languages (Gomez-imbert 2007; da Silva and Chacon 2017).

Da Silva and Chacon (2017) identify a number of relatively abstract 'general' classifiers in the three ET languages KUB, TUK, and KOT, which are summarized in Table 25 below. These classifiers are mainly classifiers of shape and correspond fairly accurately to the abovementioned cross-linguistic tendencies. However, it must be noted that in the sources to which I have access, the classifiers KUB *=kɨ*, and KOT *=kʉ* are glossed as 'tree', and 'tree, shaftlike' respectively (Chacon 2012, 249; Stenzel 2013, 122), although Da Silva and Chacon (2017) gloss them as 'rectilinear' (see Table 25). The divergence is true for many of the other classifiers listed by Da Silva and Chacon (2017), such that the apparent semantic convergence is perhaps not purely empirical, but rather the result of a bias informed by the literature itself. Therefore, in the table below, I have provided the glosses and translations as listed in the sources on these languages, in addition to the more convergent glosses provided by Da Silva and Chacon (ibid.). The glosses and examples listed here are from: KUB Chacon (2012, 242–51);

TUK West (1980, 119–21), Huber and Reed (1992, 96), and Ramírez (1997, Vol. 1, 202-204); and KOT Stenzel (2013, 122–25).

Table 25: Abstract shape classifiers in KUB, TUK and KOT

Gloss per Da Silva and Chacon (2017)	**Form**	**Gloss per source**	**Example**
3D	KUB: *=ka*	Three-dimensional, complementary to *=bo* 'oval' and *=dɨ* 'rounded'	*kɨ̃i=ka* 'manioc tuber'
	KUB: *=bo*	Oval, at least part of the body perceived as resembling a ball, yet not fully rounded	*kɨ̃rã=bo* 'a big rock or stone'
	KUB: *=dɨ*	At least most salient part of the body perceived ad rounded, or circular.	*jako=dɨ* 'eyes'
	TUK: *-ká/-ga*	Rounded, cylindrical, spherical	*kapê-ga* 'eye'
	KOT: *-ka*	Round	*~tá-ká* 'stone'
RECTILINEAR	KUB: *=kɨ*	Trees, woven objects	*hokɨ=kɨ* 'a tree'
	TUK: *-kɨ́/-gɨ*	Rectilinear, long and straight	*ɨ̃rí-pa-gɨ́* 'saws'
	KOT: *-kʉ*	Tree, shaftlike	*~jʉchʉ́-kʉ* 'leg'
CYLINDRICAL	KUB: *=jɨ̃*	Palm trees, hollow objects	*koa=jɨ* 'a pan'
	TUK: *-pɨ́/-wɨ*	Tubular, hollow, vehicle, instrument	*bupu-wɨ́* 'shotgun'
	KOT: *-dʉ*	Cylindrical and straight	*~josá-dʉ* 'spear, arrow'
	KOT: *-paro*	Cylindrical and curved	*hó-páró* 'banana'
CONCAVE/CONVEX	KUB: *=do*	Convex bodies, usually with a bulgy outward extremity	*hõbɨ̃=do* 'navel'
	KOT: *-to/-ro*	Concave	*phʉʔʉ́-ró* 'basket'
CONTAINER	KUB: *=bɨ*	Container, anything that (potentially) contains a substance	*tãu=bɨ* 'bottle'
	TUK: *-tɨ/-rɨ*	Pots, pans	*kome-tɨ* 'metal cooking pot'
FILIFORM	KUB: *=bɛ̃*	Thin line	*hõbɨ̃=bɛ̃* 'umbilical cord'
	KUB: *=bũ*	Thick line (relative to *=bɛ̃* 'thin line')	*pɨ̃kõ=bũ* 'tail'
	KOT: *-da*	Filiform, ropelike, threatlike	*phoá-dá* 'strand of hair'

BLADE	KUB: *=we*	Blade, flat with smooth substance, but not really two-dimensional	*ãbũ=we* 'hand'
	KOT: *-~phi*	Bladelike	*joa+ri-~phi* 'machete'
FLAT	KUB: *=wa*	Large flat areas, flat surface	*kɨ̃rã=wa* 'slab'
	TUK: *-wa/-pha*	Stretched out	*suʔó-wa* 'wicker sieve'
	KOT: *-phata*	Flat	*~khubú-phata* 'bench (made of boards)'

I do not claim that the classifiers with diverging translations between the original source and Da Silva and Chacon (2017) are necessarily incompatible. The original source translations are often slightly more specific than those listed by Da Silva and Chacon (ibid.), but their meanings do generally overlap, and it is possible that there are examples that justify a more abstract gloss of the classifiers. However, these adjustments require additional scrutiny and a discussion of possible homophony or polyonymy, given that I have already identified certain synchronic classifiers that are possibly the result of converged homophonous lexical sources, e.g. *-to/-do/-ro*.

Here, I will limit myself to classifiers which are likely cognates and which have abstract meanings of shape, relatively uniformly. For example, I exclude 'CL:BLADE', since KUB *=we* and KOT *-~phi* are probably not related,[59] and I also exclude 'CL:RECTILINEAR', as this classifier has been discussed under Section 6.2.1, and its abstraction from 'CL:TREE' to rectilinear is probably a relatively recent semantic extension, a path of abstraction also identified as cross-linguistically common in Adams and Conklin (1973). The classifiers listed in Table 25 which fall within these parameters are: *-ka/-ga* 'rounded'; *-to/-do/-ro* 'concave'; and possibly *-wa/-pha/-phata* 'flat'. It is therefore likely that these classifiers were among the very first to develop.

Family-wide, forms *-ka/-ga/-a* are widely found, though with a wide array of different glosses. Section 5 extensively discusses the reflexes of **-ka/-k'a/-a*, where I argue that this classifier arose as an inanimate gender category in PT. As mentioned in Chacon (2012, 248), this morpheme classifies objects that are not covered by the categories *=bo* 'oval' and *=dɨ* 'rounded', and Gomez-Imbert (2007, 419) lists a number of seemingly disparate semantic domains classified by *-a/-ka*, containing fruits, tubers and seeds not classified by *-to/-~do/-ro*, body parts and organs, and a number of natural and cultural objects, such as stones, rounded sieves, flutes, axes, fish traps, blowpipes, canoes, and cars, amongst others. Gomez-Imbert (ibid.) suggests that a number of metaphorical and metonymical processes are responsible for the seemingly disparate semantics listed here. While synchronically, an overarching abstract category may be posited, and while there may be some culturally informed explanation for perceiving concepts such as acne, flutes, axes, and canoes as saliently 'rounded', the diachronic explanation may be more complex. As established in Section 2.2, the process of grammaticalization involves semantic bleaching and loss of pragmatic influence on the use of a morpheme. Since **-ka/-k'a/-a* has probably grammaticalized at an early stage, it is probable that it was originally contrasted to only very few other nominal categories.

[59] Furthermore, I only found one possible cognate for either classifier: TAT *-~pái/-~pii/-~phi* 'blade' (Gomez-imbert 2007, 242), which probably corresponds to KOT *-~phi*.

I propose here that initially, all nouns were categorised by a very small number of categories, as is indeed asserted as a property of gender by Dixon (1986, 105), Aikhenvald (2000a, 21), and Grinevald (2000, 56). As a result, early gender markers or classifiers with which PT *-*ka*/-*k'a*/-*a* contrasted must have been very broad categories, such that a semantic category (if gender was indeed assigned on a semantic basis at all) must have been fairly abstract to begin with. Given that many authors list -*ka*/-*ga*/-*a* as a classifier for rounded objects –e.g. Table 25, but also Gomez-Imbert (2007, 419) regarding TAT, Jones and Jones (1991, 51) regarding BSA, Barnes (1990, 275) regarding TUY, and González de Pérez (2000, 384) regarding PIS– 'roundedness' may have been the meaning, or part of the meaning, of this PT classifier. However, there are also examples where -*ka*/-*ga*/-*a* is associated with cylindrical, hollow objects: BSA *tõro-a* 'flute' (Jones and Jones 1991, 50); MAK *jea-ga* 'firearm' (Smothermon, Smothemon, and Frank 1995, 37); and indeed this 'cylindrical, hollow' category seems to fit concepts such as TAT 'flute' and 'blowpipe' as listed by Gomez-Imbert (2007, 419) more easily than a category 'rounded'. Moreover, Smothermon, Smothermon, and Frank (1995, 37) list *habiti-ga* 'small canoe/airplane' and *haho-ka* 'big canoe/airplane' under their category 'hollow or tubular', concepts that may fall within a similar category as TAT canoes and cars.

Since this classifier demonstrates the same morphophonological pattern –i.e. voicedness and lack of the velar in the onset– across these semantic categories, I suppose that PET *-*ka*/-*k'a*/-*a* have the same marker as their source. Since reflexes of this marker have different meanings in different languages synchronically, it is not possible to determine the core semantic category denoted by this classifier in PET. However, as discussed above, following Chacon (in prep.), I suppose that *-*ka* was a PT inanimate gender marker. If this PT *-*ka* and PET *-*ka*/-*ga*/-*a* are related, the meaning of the PET marker may have been quite broad still. Since the category of synchronic -*ka*/-*ga*/-*a* is –at least to some extent– defined by contrast to other classifiers. One such an opposition is that between -*ka*/-*ga*/-*a* and as -*to*/-~*do*/-*ro*, and since -*ka*/-*ga*/-*a* also denotes hollow, tubular objects in a number of languages, perhaps it is also contrasted to a category 'cylindrical, hollow' which can be reconstructed as PT *-*wɨ* on the basis of family-wide comparison (Table 26).

*Table 26: Reflexes of PT *-wɨ and PET *-ka/-ga/-a referring to cylindrical/hollow concepts*

	Classifier	**Examples**			**Ref.**
		'Blowpipe'	'Flute'	'Canoe'	
C. SIO	*-wɨ*	*hɨo-wɨ*	*huri-wɨ*	*jo-gu*	I
E. SIO	*-wɨ*	*hɨo-wɨ*		*jo-wɨ*	II
E. SEK	*-wɨ*	*hɨo-wɨ*	*pĩʔko-wɨ*	*jo-wɨ*	III
P. SEK	*-wɨ*			*ʤo-wɨ*	IV
KOR	*-wɨ*	*hɨo-jɪ̃*	*phĩʔko-wɨ*	*joo-wɨ*	V
MAI	*-bɨ*	*hu*		*jou*	VI
TAN	*-bi*		*a'ɸé-bí ɸúɸúo-ká*		VII
DES	*-su* *-du/-ru*		*tadi-su*	*gasi-ru*	VIII
TUK	*-wɨ*	*pekâ-wɨ*[60]	*buaa-wɨ́*	*juki-sɨ́*	IX
KOT	*-ka*	*púkà*	*phuti-ria*	*bʉho-ka*	X
KUB	*=jɪ̃*	*pĩo=jɪ̃*	*hapu-i=jɪ̃*	*hiado=kũ*	XI
PIS	*=ga/=a*			*kũmũ=a*	XII

[60] I found no example of 'blowpipe' in Ramírez (1997). The form cited here is translated as 'firearm' (id., Vol. 1, 214).

BSA	*-ka/-ga/-a*	*buhu-a*		*kũbũ-a*	XIII
TAT	*-ga/-a*	*bupu-a*		*kũbũ-a*	XIV
MAK	*-ka/-ga*	*buha-ga*		*haho-ka*	XV
KAR	*-a*	*bupu-a*		*kũbũ-a*	XVI
TUY	*-wɨ*	*bupu-wɨ*		*juku-soro*	

Table 26 shows the reflexes of PT **-wɨ* and PET **-ka/-ga/-a* on nouns that refer to cylindrical or hollow objects. The fact that that some languages employ a reflex of **-wɨ* and other a reflex of PET **-ka/-ga/-a* is an indication that these two classifiers may have had an overlap in meaning at some stage in the history of the family.

I have now identified three classifiers that contrast to one another in certain semantic domains, where the choice of classifier seems to be lexically determined in at least some languages: PT **-wɨ*; PT **-to/-t'o*; and PT **-ka/-ga/-a*. It is difficult to say anything conclusive about their original meanings, but **-wɨ* seems to be related to hollow, oblong, or tubular concepts, **-to/-t'o* to concave concepts, and **-ka/-ga/-a* to rounded things, body parts, and hollow concepts.

Moreover, as mentioned above, PT **-to/-t'o* may have a connection to the homophonous spatio-temporal marker, which is interesting since a number of WT languages employ *-wɨ* for certain spatial concepts, e.g. 'beach': E. SIO *mehã-wɨ* (Bruil 2014, 135); E. SEK *meha-wɨ* (Johnson and Levinsohn 1990, 28; Piaguaje et al. 1992, 80); KOR *meha-wɨ* (Cook and Gralow 2001, 78).[61] However, this *-wɨ* may not be historically the same as the reconstructed PT classifier of the same form in Table 26, since the same noun root is found with another, but similar form in C. SIO *meha-bɨ* 'beach'. Moreover, there is a minimal pair in this language, between the *-wɨ* and *-mɨ*: *gɨna-wɨ* 'metal container/bottle/drum', cf. *gɨna-mɨ* 'sky (metal-CL:HIGH)'. The same noun contains the classifier *-wɨ* in E. SIO *ɉɨna-wɨ* 'sky', such that we are presented with a puzzle: Either **-wɨ* and **-bɨ/-mɨ* were two or three different classifiers at the PWT stage, which would require an explanation as to why they were all subsumed under **wɨ* in all languages but C. SIO, or **-bɨ/-mɨ* is a more recent development in C. SIO, which would require an explanation as to what prompted this development.[62] At this time, I cannot answer these questions, and I leave them for future research.

Based on the idea that these three classifiers were among the very first to develop, as discussed above, I posit the following scenario: The PT gender system contained one 'inanimate' category **-ka*, which at some point was contrasted to **-to/-t'o*, **-wɨ*, and perhaps to some other marker or markers as well. In line with the aforementioned cross-linguistic tendencies, these classifiers denoted quite abstract meanings of shape, rigidity, fullness, and possibly size. At some point, additional categories were introduced into the system, with increasingly more specific meanings, giving rise to a system that shares many features with a classifier system, but also retains some of the characteristics associated with gender markers, such as the possibility of nominal agreement. Since new categories were more specific in meaning, and possibly arose productively from nominal compounding, the previous system – where a very small amount of nominal markers could denote all nouns in the language– was supplemented by a new system of a larger amount of nominal markers, which could denote a specific subgroup of concepts. As a result, the older markers came to categorize a number of apparently disparate concepts, complemented with a number of more specific classifiers, such

[61] See also KUB *hia=wɨ* 'island' (Morse, Salser, and De Salser 1999, 167).
[62] A possible source for *-mɨ* 'CL:HIGH' is the C. SIO verb *ɨmɨ-* 'to be high' (Wheeler 1987, Vol. 2, 87).

that the older nominal markers either lost a recognizable meaning, and were thus reanalysed as 'generic' classifiers, or were reanalysed as homophonous markers, with separate meanings. Consider for example Table 27, which posits a diversification scenario which gave rise to the current classifiers associated with nouns referring to teeth and tongues in PWT and PET, where in both scenarios, it is PET which retains the more archaic classifier, and PWT which undergoes a development such that a newly grammaticalized classifier substitutes the more archaic one.[63]

*Table 27: Diversification scenario of classifiers associated with reflexes of PT *k'õpi 'tooth', and PT *tj'eme 'tongue'*

<table>
<tr><th>Early PT</th><th></th><th colspan="2">Late PT</th><th>PWT addition:</th></tr>
<tr><td rowspan="3">*-ka 'INAN'</td><td rowspan="3">↗</td><td colspan="2">*-ka/-k'a/-a 'CL:ROUNDED'</td><td rowspan="3">*-jɨ̃ 'CL:TREE'
*-jo 'CL:LONG.THIN'</td></tr>
<tr><td colspan="2">*-to/t'o 'CL:CONCAVE/SPATIO-TEMPORAL'</td></tr>
<tr><td colspan="2">*-wɨ 'CL:HOLLOW.OBLONG/TUBULAR'</td></tr>
<tr><td colspan="3">*k'õpi 'tooth' ⟶ *k'õpi-k'a</td><td rowspan="2">↗
↘</td><td>PWT
*k'õhi-jɨ̃

PET
*k'õpi-k'a</td></tr>
<tr><td colspan="3">*tj'eme 'tongue' ⟶ *tj'eme-t'o</td><td>PWT
*tj'eme-jo

PET
*~tj'ebe-~t'o</td></tr>
</table>

The scenario proposed here is perhaps counterintuitive since it involves a diachronically rare phenomenon of semantic narrowing. A classifier which originally may have been so broad as to refer to anything inanimate became specific to rounded things –with the exception of objects more saliently concave, or hollow and tubular– and eventually came to denote a number of seemingly disparate semantic categories –i.e. body parts, fruits, and means of transport– once other semantic fields, originally covered by the classifier received a more specific classifier. It may also come as a surprise that the more specific classifiers developed more recently than the Tukanoan gender markers, as it is often asserted that gender markers develop from classifiers, and not the other way around (e.g. Dixon 1982, 171–73; Corbett 1991, 139–41; Aikhenvald 2000a, 372–73), and Amazonian classifier systems which share a number of features with typical gender or noun class systems have been characterised as emerging noun class systems (Grinevald and Seifart 2004). However, my proposal does fit within the conventional idea that gender markers are further grammaticalized than nominal classifiers (as was also discussed in Section 2.2), the related notion that the Tukanoan specific classifiers are later developments than the gender markers (see also Chacon 2021; in prep.),[64] and the idea that the Tukanoan shape classifiers are among the classifiers that developed earliest (da Silva and Chacon 2017). Lastly, as follows from Section 2.1, the relative lack of pragmatic motivation for these classifiers is the hallmark of prototypical gender, which hence offers an explanation of the other prototypical gender-like features of these classifiers.

[63] The data on which the generalizations in Table 27 are based mainly come from Huber and Reed (1992), and the reconstructed forms are taken from Chacon (2014).

[64] Likewise, the Arawakan languages demonstrate relatively consistent gender systems, whereas classifiers and classifier systems can only be reconstructed to a limited degree (Aikhenvald 1994).

As was furthermore discussed in Section 2.2, the Tukanoan classifiers probably had lexical nouns as their sources and fulfilled a function of noun classifier originally. It is had been established cross-linguistically that anaphoric elements may develop into gender markers if there is already a gender system in place (Aikhenvald 2000a, 375). If these anaphoric markers are already grammatical in nature, e.g. a locative expression or a demonstrative (see Corbett 1991, 313–4), they may function on par with the existing gender markers, but since the emerging Tukanoan classifiers had meaning, they retained a slightly different grammatical function in these language. This scenario explains why the Tukanoan gender and classifier systems share some overlap, but are not identical.

Since many other Northwest Amazonian language families show similar patterns, it would be interesting to consider whether a similar scenario could also be posited for these languages, although discussion of these other families falls outside the scope of this thesis.

6.3.3 Animate collective

Many Tukanoan languages have one 'animate collective' classifier (Gomez-imbert 2007, 424), e.g. E. SIO *-hɨ/-bɨ* (Bruil 2014, 141); KUB *=wɨ* (Chacon 2012, 251), and there are also some inanimate classifiers referring to collective or otherwise non-singular referents, e.g. E. SIO *-(j)e* 'CL:INAN.COLL/GEN'. The abovementioned hypothesis that

However, these are exceptions and these forms may demonstrate slightly deviant morphosyntactic behaviours, e.g. BSA *-se* 'CL:GEN' is one of the few classifiers in that language that do not require the plural marker *-ri* in plural deverbal nouns (Jones and Jones 1991, 33). General *-je* is also found in the gender paradigm which explains the disparate status of this morpheme.

It is possible to analyse the animate collective classifier –e.g. E. SIO *-hɨ/-bɨ*– as an originally complex morpheme consisting of the PT animate gender **-ɨ* in combination with some bilabial consonant, although it is not clear which proto-form this would entail, i.e. PT **-p-i-ɨ* > E. SIO *-hi*, cf. KUB *#=pi*, PT **-p'-ɨ* > E. SIO *-bɨ*, KUB *#=bɨ*, PT **-w-ɨ* > E. SIO *#-wɨ*, KUB *=wɨ*. These examples show that the source for the bilabial in these animate collective markers is not easily reconstructable. However, it is relevant to note that the TUY classifier *-ga* 'CL:3D/SPHERICAL' has a suppletive plural, where a bilabial consonant takes the place of the velar: SG *-ga*, cf. PL *-pa* (Barnes 1990, 279). This suppletion suggests that there may have been a plural or collective number marker involving a bilabial consonant at some point, which may also be responsible for the bilabial consonant in the animate collective classifiers. Further evidence for a plural marker containing a bilabial is found in WA'I, where the plural counterpart of nonplural *-ga* 'CL:ROUND' is *-poka*, e.g. *~ʉta-ga* 'stone', cf. *~ʉta-poka* 'stones' (Balykova 2019, 8). This bilabial may furthermore have been subject to the same lenis/fortis distinction observed by Chacon and Michael (2018). This would explain the deviant sound correspondences in the forms listed above. The E. SIO and KUB forms are compared in Table 28 below, where I make a tentative assumption that the plural marker containing the bilabial may have had fortis form **pa*, and lenis **p'a*, and where the vowel is based on the WA'I form of the irregular plural.

Table 28: A possible reconstruction of the bilabial in animate collective markers in E. SIO and KUB

	PT	**E. SIO**	**KUB**
FORTIS	*p(a)-i-ɨ*	*-hi*	–
LENIS	*p'(a)-ɨ*	*-bɨ*	*-wɨ*

According to Chacon and Michael (2018, 72), PT suffix groups with an even amount of syllables triggered the fortis form of the bilabial **p*, whereas suffix groups with an odd number of morae triggered the lenis form **p'*. Evidence for this distinction is found in the E. SIO forms, since the form containing *h*, a reflex of the fortis bilabial, contains the front vowel *i*, whereas the form containing *b*, a reflex of the lenis bilabial, does not, which is explained if we suppose that the form *hi* originally had an odd number of morae in its suffix group, viz. *-i-ɨ*, where the front vowel is perhaps a remnant of the proximate demonstrative *i*, and the form *-bɨ* had an odd number of morae in its suffix group, viz. simply *-ɨ*.

Animate plural **-pɨ/-p'ɨ* did therefore probably did originally develop opposed to inanimate singular **-ka/-k'a*, but already at the stage of the gender system, and may therefore perhaps rather be analysed as originally opposing the singular animate marker, or the singular animate and inanimate marker alike. Recall that a similar analysis was given for PT **-je* in Section 5.3, but that the exact stage of its development could not be determined.

7 Conclusion

In this thesis, I have argued that the noun categorization markers in the Tukanoan languages are, at least morphosyntactically speaking, essentially an expansion of the gender system that was already in place at the PT level. Some classifiers can be reconstructed to PT, whereas others are most probably later developments.

The lexical material that formed for the Tukanoan inanimate classifiers is probably predominantly nominal, although some apparent exceptions are also found. I have found at least two WT classifiers of possible verbal origin: PWT **tɨõ-* 'to weave'; C. SIO *-mɨ* 'CL:HIGH', although it is also possible that the verbs and classifiers are two separate developments of the same noun, which may or may not persist in any of the described Tukanoan varieties. The forms of the classifiers *-ka/-ga/-a*, *-to/-do/-ro*, and *-wɨ* most probably developed out of the gender system.

A summary of the cognates I have identified, reconstructed proto-forms and possible etymologies of the classifiers discussed in this thesis is found in Appendix B. The relevant sound correspondences these cognates' consonants are condensed in Appendix C, where I propose a number of reflexes for bound forms on the basis of the reconstructed classifiers. Note, however, that a greater sample of classifier cognates is necessary in order to increase the reliability of these sound correspondences, and in order to determine what phonological contexts play a role in the attested reflexes.

A number of reconstructed forms –i.e. PT **-k'ɨ/-kɨ*, **-k'o/-ko*, **-k'a/-ka*, and **-t'o/-to*– display an apparent allomorphy which is reminiscent of the lenis/fortis distinction identified for bilabial **-p/-p'* by Chacon (2014) and Chacon and Michael (2018), and which may provide an additional argument that these classifiers are most archaic to the family. It is therefore possible PT at some point had a morphophonological rule that caused all stops in the onsets of bound to display a lenis/fortis alternation. At this point, it is difficult to say what environment triggered the allomorphy. The allomorphy may have been triggered by:

i) the number of syllables in the suffix group, similarly to the PT **-pɨ/-p'ɨ* allomorphy (ibid.);
ii) another morpheme that triggered the allomorphy, similarly to a reconstructed development in TUK where PET **-sa* 'to go', and **-ti/-di* 'NLZ' provided an

environment for a fortis/lenis allomorphy of PET *-*p*/-*p'*, *-*t*/-*t'*, and *-*k*/-*k'* to TUK *w*/*p*ʰ, *r*/*t*ʰ, and *g*/*k*ʰ respectively (Chacon 2016, 274–76);

iii) another, yet to be identified environment that triggered the allomorphy.

ET languages show more forms that demonstrate this allomorphy which may be an argument that these forms are more archaic than the forms reconstructed for the WT branch, or that the ET branch retained both the fortis and lenis allomorphs longer (see Appendices).

I have also discussed terminological and typological issues of classifiers in a wider sense, and proposed a defining property for distinguishing classifiers and gender markers, where classifiers are essentially fulfil a more pragmatic function, whereas gender markers fulfil a more grammatical function. I have also identified a terminological issue regarding *repeaters*, where I distinguish between repeaters in the narrow sense, i.e. classifiers that occur with nouns from which they are productively derived, and repeaters in the broad sense, i.e. classifiers that have the same form as nouns in the lexicon, but which do not necessarily have to occur together.

Lastly, I have discussed a number of problems in reconstructing certain classifiers where great progress could be made in interdisciplinary projects, e.g. the hydronymical and topographical classifiers, the classifiers for trees in both main branches of the family, and the notion that axes are classified as **-p'o* 'CL:CAVE-LIKE' or perhaps 'CL:HOLE'. Questions regarding possible loans from one language or branch within the family to another may be yet be answered by comparing material culture, geographic and biological information with the linguistic data, which could in turn provide improved insights into past contacts of the Tukanoan family.

References

Adams, Karen L, and Nancy Faires Conklin. 1973. “Toward a Theory of Natural Classification.” In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, edited by Claudia Corum, T. Cedric Smith-Stark, and Ann Weiser, Vol. 9:1–10. Chicago Linguistic Society.

Aikhenvald, Alexandra Y. 1994. “Classe Nominal e Gênero nas Línguas Aruák [Nominal Class and Gender in Arawakan Languages].” *Boletim Do Museo Paraense Emílio Goeldi* 10: 137–59.

———. 1999. “The Arawak Language Family.” In *The Amazonian Languages*, edited by R. M. W. Dixon and Alexandra Y. Aikhenvald, 65–106. Cambridge: Cambridge University Press.

———. 2000a. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.

———. 2000b. “Unusual Classifiers in Tariana.” In *Systems of Nominal Classification*, edited by Gunter Senft, 93–113. Cambridge: Cambridge University Press.

———. 2002. *Language Contact in Amazonia*. Oxford: Oxford University Press.

———. 2012. *The Languages of the Amazon.* Oxford: Oxford University Oress,

Alemán, Tulio M., Reinaldo H. López, and Marion Miller. 2000. *Wirá Ya, Peamasa Ya Wererituri: Desano-Español, Diccionario Bilingüe de 896 Palavras*. Santafé de Bogotá: Editorial Alberto Lleras Camargo.

Alexander-Bakkerus, Astrid. 2005. *Eighteenth Century Cholón*. Utrecht: LOT. http://www.jstor.org/stable/10.1086/598772.

Allan, Keith. 1977. “Classifiers.” *Language* 53 (2): 285–311.

Allassonnière-Tang, Marc, and Marcin Kilarski. 2020. “Functions of Gender and Numeral Classifiers in Nepali.” *Poznan Studies in Contemporary Linguistics* 56 (1): 113–68. https://doi.org/10.1515/psicl-2020-0004.

Alpher, Barry. 2011. *Yir-Yoront Lexicon*. Berlin/New York: De Gruyter Mouton. https://doi.org/https://doi-org.ezproxy.leidenuniv.nl/10.1515/9783110872651.

Arias Alvis, Leonardo, Nicholas Q. Emlen, Sietze Norder, Nora Julmi, Magdalena Lemus Serrano, Thiago Costa Chacon, Jurriaan Wiegertjes, Austin Howard, and Rik Van Gijn. 2021. *Interpreting Mismatches between Linguistic and Genetic Patterns among Speakers of Tanimuka (Eastern Tukanoan) and Yukuna (Arawakan).* Presented at SALSA XIII Biennial Conference, Charlottesville.

Balykova, Kristina. 2019. “What do oranges and hammers have in common? The classifier ‘round’ in Wa’ikhana and other East Tukano languages.” *LIAMES* 19: 1–24. http://dx.doi.org/10.20396/liames.v19i0.8655262

———. 2021. “Interaction between Grammatical Number and Nominal Classification System in Wa’ikhana.” Presented at Amazonicas VIII, Góias/Brasília. https://www.youtube.com/watch?v=7F6stVlImZc.

Bani, Ephraim. 1987. “Garka a Ipika: Masculine and feminine grammatical gender in Kala Lagaw Ya.” *Australian Journal of Linguistics* 7: 189–201. http://dx.doi.org/10.1080/07268608708599381

Barnes, Janet. 1990. “Classifiers in Tuyuca.” In *Amazonian Linguistics: Studies in Lowland South American Languages*, edited by Doris L. Payne, 273–92. Austin: University of Texas Press.

———. 1999. “Tucano.” In *The Amazonian Languages*, edited by R. M. W. Dixon and Alexandra Y. Aikhenvald, 207–26. Cambridge: Cambridge University Press.

———, and Terrell Malone. 2000. “El Tuyuca.” In *Lenguas Indígenas de Colombia: Una Visión Descriptiva*, edited by María Stella González de Pérez and María Luisa Rodríguez de Montes, 437–52.

Beddor, Patrice Speeter, Rena Arens Krakow, and Louis M. Goldstein. 1986. “Perceptual Constraints and Phonological Change: A Study of Nasal Vowel Height.” *Phonology Yearbook* 3: 197–217. https://doi.org/10.1017/s0952675700000646.

Bertinetto, Pier Marco. 2014. “Ayoreo.” In *Lenguas de Bolivia, Tomo III: Oriente*, edited by Mily Crevels and Pieter Muysken, 369–413. La Paz: Plural editores.

Bostrom, Paula Kaye. 1998. *Nominalizations and Relative Clauses in Tatuyo: A Prototype Approach*. The University of Texas, Arlington PhD dissertation. https://doi.org/10.1016/b978-0-08-102025-8.00003-x.

Bruil, Martine. 2012. *Documentation of Ecuadorian Siona.* Endangered Languages Archive. http://hdl.handle.net/2196/00-0000-0000-000D-EA53-3. Accessed in January 2021.

———. 2014. *Clause-Typing and Evidentiality in Ecuadorian Siona*. Leiden University, Leiden PhD dissertation.

———. 2016. *Heterogeneity and Variation in the Use of Ecuadorian Siona Classifiers.* Presented at Combining gender and classifiers in natural language, Surrey.

———. 2018. “The Development of the Portmanteau Verbal Morphology in Ecuadorian Siona.” *Journal of Historical Linguistics* 8 (1): 128–67. https://doi.org/10.1075/jhl.17003.bru.

Campbell, Lyle. 2013. “Language Contact and Linguistic Change in the Chaco.” *Revista Brasileira de Linguística Antropológica* 5 (2): 259–91. https://doi.org/10.26512/rbla.v5i2.16268.

Chacon, Thiago Costa. 2012. *The Phonology and Morphology of Kubeo: The Documentation, Theory and Description of an Amazonian Language*. University of Hawai’i, Mānoa PhD dissertation.

———. 2014. “A Revised Proposal of Proto-Tukanoan Consonants and Tukanoan Family Classification.” *IJAL* 80 (3): 275–322.

———. 2016. “The Reconstruction of Laryngealization in Proto-Tukanoan.” In *The Phonetics and Phonology of Laryngeal Features in Native American Languages*, edited

by Heriberto Avelino, Matt Coler, and Leo W. Wetzels, 258–84. Leiden/Boston: Koninklijke Brill NV.

———. 2017. “Arawakan and Tukanoan Contacts in Northwest Amazonia Prehistory.” *PAPIA: Revista Brasileira de Estudos Crioulos e Similares* [Brazilian Journal of Creole and Similar Studies] 27: 237–65. https://d1wqtxts1xzle7.cloudfront.net/57015262/Chacon_2017_Arawakan_and_Tukanoan_contacts_in_Northwest_Amazonia_prehistory.pdf?1531834901=&response-content-disposition=inline%3B+filename%3DArawakan_and_Tukanoan_contacts_in_Northw.pdf&Expires=1632135951&Si.

———. 2021. “Origem e Convergência nos Sistemas de Classificação Nominal em Línguas Tukano [Birth and Convergence of Gender and Nominal Classifier Systems in Tukanoan Languages].” Presented at Amazonicas VIII, Góias/Brasília. https://www.youtube.com/watch?v=jPSQ9d-L_cg.

———. In prep. “Rumos, Picadas e Caminhos na Mata Densa: Explorações Diacrônicas Sobre os Sistemas de Gênero na Família Tukano [Directions, Trails and Paths in the Dense Forest: Diachronic Explorations on the Gender Systems of the Tukano Family].”

———, and Johann-Mattis List. 2015. “Improved Computational Models of Sound Change Shed Light on the History of the Tukanoan Languages.” *Journal of Language Relationship* 13 (3): 177–203. https://doi.org/10.31826/9781463236984-003.

———, and Lev Michael. 2018. “The Evolution of Subject-Verb Agreement in Eastern Tukanoan.” *Journal of Historical Linguistics* 8 (1): 59–94. https://doi.org/10.1075/jhl.16024.cha.

———, and Rodrigo do Prado Sateles. 2019. “Explorações Sobre Padrões de Indexação e Classificação Nominal de Argumentos em Línguas Indígenas da América do Sul [Explorations in Patterns of Indexation and Nominal Classification of Arguments in Indigenous Languages of South America].” *LIAMES* 19: 1–35. https://doi.org/10.20396/liames.v19i0.8655674.

Contini-Morava, Ellen. 2000. “Introduction.” In *Between Grammar and Lexicon*, edited by Ellen Contini-Morava and Yishai Tobin, ix–xxxii. Amsterdam/Philadelphia: John Benjamins Publishing Company.

———. 2021. “Reflections on Nominal Classification.” Presented at Amazonicas VIII, Góias/Brasília.

———, and Marcin Kilarski. 2013. “Functions of Nominal Classification.” *Language Sciences (Oxford)* 40: 263–99. https://doi.org/10.1016/j.langsci.2013.03.002.

Cook, Dorothy M., and Linda L. Criswell. 1993. *El Idioma Koreguaje (Tucano Occidental)*. Santafé de Bogotá: Editorial Alberto Lleras Camargo.

———, and Frances L. Gralow. 2001. *Diccionario Bilingüe Koreguaje-Español, Español-Koreguaje*. Bogotá: Editorial Alberto Lleras Camargo.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139166119.011.

———. 2009. “Agreement.” In Die Slavischen Sprachen / The Slavic Languages, edited by Tilman Berger, Karl Gutschmidt, Sebastian Kempgen, and Peter Kosta, 342–54. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110214475.1.6.342.

———. 2013. “Sex-Based and Non-Sex-Based Gender Systems.” In *The World Atlas of Language Structures Online*, edited by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/31.

———, and Sebastian Fedden. 2016. “Canonical Gender.” Journal of Linguistics 52 (3): 495–531. https://doi.org/10.1017/S0022226715000195.

———, Sebastian Fedden, Michael Franjieh, Alexandra Grandison, and Erich Round. 2020. “Gender: New Horizons.” https://www.youtube.com/watch?v=JmJDV5q7zqQ.

Craig, Colette. 1986a. “Jacaltec Noun Classifiers: A Study in Grammaticalization.” Lingua 70 (4): 241–84. https://doi.org/https://doi.org/10.1016/0024-3841(86)90046-X.

———. 1986b. “Jacaltec Noun Classifiers: A Study in Language and Culture.” In *Noun Classes and Categorization*, edited by Colette Craig, 264–93.

———. 1986c. *Noun Classes and Categorization*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Creveaux, Jules, P. Sagot, and Lucien Adam. 1882. *Roucouyenne, Arrouague, Piapoco et d’Autres Langues de la Région des Guyanes*. Paris: Librairie-Éditeur J. Maisonneuve.

Crevels, Mily. 2012. “Itonama.” In *Ambito Andino*, edited by Mily Crevels and Pieter Muysken, 233–94. La Paz: Plural editores.

Criollo Quintero, Ligia Mariá. 2011. *Aibë Coca: Historias de Nuestros Antepasados*. Ibarra: Creadores Gráficos.

Denny, J. Peter. 1986. “The Semantic Role of Noun Classifiers.” In *Noun Classes and Categorization*, edited by Colette Craig, 297–308.

Derbyshire, Desmond C., and Doris L. Payne. 1990. “Noun Classification Systems of Amazonian Languages.” In *Amazonian Linguistics: Studies in Lowland South American Languages*, edited by Doris L. Payne, 243–71.

Dixon, R. M. W. 1982. *Where Have All the Adjectives Gone? And Other Essays in Semantics and Syntax*. Berlin: Mouton. https://doi.org/10.1515/9783110822939.185.

———. 1986. “Noun Classes and Noun Classification in Typological Perspective.” In *Noun Classes and Categorization*, edited by Colette Craig, 105–12.

———. 2010. *Basic Linguistic Theory*. Oxford: Oxford University Press.

Enfield, Nick J. 2014. “Nominal Classification in Lao: A Sketch.” *STUF - Language Typology and Universals* 57 (2–3): 117–43. https://doi.org/10.1524/stuf.2004.57.23.117.

Epps, Patience. 2008. *A Grammar of Hup*. Berlin/New York: Mouton de Gruyter.

Eraso, Natalia. 2015. *Gramática Tanimuka, una Lengua de la Amazonía Colombiana*. Université Lumière Lyon 2, Lyon PhD dissertation.

Erbaugh, Mary S. 1986. "Taking Stock: The Development of Chinese Noun Classifiers Historically and in Young Children." In *Noun Classes and Categorization*, edited by Colette Craig, 399–436.

Fabre, Alain. 2007. "Morfosintaxis de los Clasificadores Posesivos en las Lenguas del Gran Chaco (Argentina, Bolivia y Paraguay)." *UniverSOS: Revista de Lenguas Indígenas y Universos Culturales* 4: 67–85.

Farmer, Stephanie Jo. 2015. *Establishing Reference in Máihɨ̀kì*. University of California, Berkeley PhD dissertation.

———, and Martine Bruil. 2015. "Gender and Noun Classifiers in Two Western Tukanoan Languages." Presented at Combining Gender and Classifiers in Natural Language, Surrey.

———, and Rosa Vallejos. 2021. "A Comparative Study of the Morphosyntax of Western Tukanoan Noun Classifiers." Presented at Amazonivas VIII, Góias/Brasília. https://www.youtube.com/watch?v=giCemK7oZRA.

Fedden, Sebastian, and Greville G. Corbett. 2017. "Gender and Classifiers in Concurrent Systems: Refining the Typology of Nominal Classification." *Glossa: A Journal of General Linguistics* 2 (1): 1–47. https://doi.org/10.5334/gjgl.177.

Franjieh, Michael, Greville G. Corbett, and Alexandra Grandison. 2020. "How Language Shapes Thought: Categorization in the South Pacific." Presented at ESRC, Surrey. https://www.youtube.com/watch?v=sHSnW4UmlRc.

Gil, David. 2013. "Numeral Classifiers." In *The World Atlas of Language Structures Online*, edited by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/chapter/55.

Gomez-Imbert, Elsa. 1982. *De la Forme et du Sens dans la Classification Nominale en Tatuyo: Langue Tukano Orientale d'Amazonie Colombienne*. Université de la Sorbonne, Paris PhD dissertation.

———. 1996. "When Animals Become 'Rounded' and 'Feminine': Conceptual Categories and Linguistic Classification in a Multilingual Setting." In *Rethinking Linguistic Relativity*, edited by John J Gumperz and Stephen C Levinsohn, 438–69. Cambridge: Cambridge University Press.

———. 2007. "Tukanoan Nominal Classification: The Tatuyo System." In *Language Endangerment and Endangered Languages: Linguistic and Anthropological Studies with Special Emphasis on the Languages and Cultures of the Andean–Amazonian Border Area*, edited by Leo Wetzels, 301–428. Leiden: CNWS Publications.

González de Pérez, María Stella. 2000. "Bases Para El Estudio de La Lengua Pisamira." In *Lenguas Indígenas de Colombia: Una Visión Descriptiva*, edited by María Stella González de Pérez and María Luisa Rodríguez de Montes, 373–94.

———, and María Luisa Rodríguez de Montes. 2000. *Lenguas Indígenas de Colombia: Una Visión Descriptiva*. Santafé de Bogotá: Instituto Caro y Cuervo.

González Muñoz, Sara. 2016. *Descripción Morfológica de los Adjetivos Calificativos en la Lengua Pisamira.* Universidad del Valle, Cali MA thesis.

Greenberg, Joseph. 1972. "Numeral Classifiers and Substantival Number: Problems in the Genesis of a Linguistic Type." *Working Papers on Language Universals* 9: 1–39.

———. 1978. "How Does a Language Acquire Gender Markers?" In *Universals of Human Language, Vol. 3: Word Structure*, edited by Joseph Greenberg, 47–82. Stanford, CA: Stanford University Press.

Grinevald, Colette. 2000. "A Morphosyntactic Typology of Classifiers." In *Systems of Nominal Classification*, edited by Gunter Senft, 50–92.

———. 2002. "Making Sense of Nominal Classification Systems: Noun Classifiers and the Grammaticalization Variable." In *New Reflections on Grammaticalization*, edited by Ilse Wicher and Gabriele Diewald, 259–75. Amsterdam/Philadelphia: John Benjamins Publishing Company.

———. 2004. "Classifiers." In *Morphologie: Ein Internationales Handbuch Zur Flexion Und Wortbildung (2. Halbband) / Morphology: An International Handbook on Inflection and Word-Formation (2nd Half)*, edited by Geert E. Booij, Christian Lehmann, Joachim Mugdan, Stavros Skopeteas, and Wolfgang Kesselheim, 1016–31. Berlin: Mouton de Gruyter. https://www-degruyter-com.ezproxy.leidenuniv.nl/document/doi/10.1515/9783110172782.2/html.

———. 2015. "Classifiers, Linguistics Of." In *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition), edited by James D Wright, Second Edi, 811–18. Oxford: Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-08-097086-8.53003-7.

Grinevald, Colette, and Frank Seifart. 2004. "Noun Classes in African and Amazonian Languages: Towards a Comparison." *Linguistic Typology* 8 (2): 243–85. https://doi.org/10.1515/lity.2004.007.

Hammarström, H., R. Forkel, M. Haspelmath, and S. Bank. 2020. "Glottolog." 2020. https://glottolog.org/resource/languoid/id/tuca1253.bigmap.html#6/-0.770/288.124.

Heine, Bernd, and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge University Press. https://doi.org/10.1017/CBO9780511613463.

———. 2005. *Language Contact and Grammatical Change. Cambridge Approaches to Language Contact*. Cambridge University Press. https://doi.org/10.1017/CBO9780511614132.

Hopper, Paul J, and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Huber, Randal Q, and Robert B Reed. 1992. *Vocabulario Comparativo: Palabras Selectas de Lenguas Indígenas de Colombia*. Santafé de Bogotá: Asociatión Instituto Lingüístico de Verano.

Jackson, Jean E. 1983. *The Fish People: Linguistic Exogamy and Tukanoan Identity in Northwest Amazonia*. Cambridge: Cambridge University Press.

Johnson, Orville J., and Stephen H. Levinsohn. 1990. *Gramática Secoya*. Quito: Instituto Lingüístico de Verano.

Johnson, Steve. 1988. "The Status of Classifiers in Kugu Nganhcara Nominals." *Aboriginal Linguistics* 1: 198–204. https://openresearch-repository.anu.edu.au/bitstream/1885/10567/1/Johnson_StatusClassifiers1988.pdf.

Jones, Wendell, and Paula Jones. 1991. *Barasano Syntax*. Dallas, Texas: Summer Institute of Linguistics/The University of Texas at Arlington.

Kilarski, Marcin. 2013. *Nominal Classification: A History of Its Study from the Classical Period to the Present*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Krasnoukhova, Olga. 2012. *The Noun Phrase in the Languages of South America*. Radboud Universiteit, Nijmegen PhD dissertation.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*. Stanford, CA: Stanford university press.

Lima Silva, Wilson de. 2012. *A Descriptive Grammar of Desano*. University of Utah, Salt Lake City, PhD dissertation.

List, Johann-Mattis, Cristina Messineo, and Nicolás Brid. 2021. "Beyond Structure: Towards an Analysis of Semanc Patterns in the Gran Chaco." Presented at Amazonicas VIII, Góias/Brasília. https://www.youtube.com/watch?v=cehPlrPZuBk.

Lucy, John A. 2000. "Systems of Nominal Classification: A Concluding Discussion." In *Systems of Nominal Classification*, edited by Gunter Senft, 326–41.

Lyons, John. 1977. *Semantics.* Cambridge University Press. https://doi.org/10.1017/CBO9780511620614.004.

Mason, J. Alden. 1950. "The Languages of South American Indians." In *Handbook of South American Indians*, edited by Julian H. Steward, 156–317. Washington: United States Government Printing Office.

Meillet, Antoine. 1923. "Le Genre Féminin Dans Les Langues Indo-Européennes." *Journal de Psychologie Normale et Pathologique* 20: 943–44.

Metzger, Ronald G. 1981. *Gramatica Popular del Carapana*. Bogotá: Instituto Lingüístico de Verano.

———. 1998. "The Morpheme KA- of Carapana (Tucanoan)." *SIL Electronic Working Papers* 14: 1–47. http://www.sil.org/silewp/1998/003/silewp1998-003.pdf.

Miller, Marion. 1999. *Desano Grammar*. Studies in the Languages of Colombia, 6. Dallas, Texas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Mithun, Marianne. 1986. "The Convergence of Noun Classification Systems." In *Noun Classes and Categorization*, edited by Colette Craig, 179–97.

Morse, Nancy L., and Michael B. Maxwell. 1999. *Gramática del Cubeo*. Santafé de Bogotá: Editorial Buena Semilla.

Morse, Nancy L., Jay K. Jr. Salser, and Neva de Salser. 1999. *Diccionario Ilustrado Bilingüe: Cubeo-Español, Español-Cubeo*. Santafé de Bogotá: Editorial Alberto Lleras Camargo.

Onishi, Masayuki. 1994. *A Grammar of Motuna (Bougainville, Papua New Guinea).* Australian National University, Canberra PhD dissertation.

Passer, Matthias Benjamin. 2016. *The Typology and Diachrony of Nominal Classification*. Universiteit van Amsterdam, Utrecht PhD dissertation.

Payne, Doris L. 1984. "Evidence for a Yaguan-Zaparoan Connection." *Work Papers of the Summer Institute of Linguistics, University of North Dakota* 28: 131–56.

———. 1985. *Aspects of the Grammar of Yagua: A Typological Perspective.* University of California, Los Angeles PhD dissertation. https://pure.mpg.de/rest/items/item_401038/component/file_401037/content

———. 1987. "Noun Classification in the Western Amazon." *Language Sciences* 9 (1): 21–44. https://doi.org/10.1016/S0388-0001(87)80007-4.

———. 1990. *Amazonian Linguistics: Studies in Lowland South American Languages*. Austin: University of Texas Press.

———. 2007. "Source of the Yagua Nominal Classification System." *IJAL* 73 (4): 447–74. https://doi.org/10.1086/523773.

Pet, Willem Jan Agricola. 1987. *Lokono Dian, The Arawak Language of Suriname: A Sketch of its Grammatical Structure and Lexicon.* Cornell University, Ithaca PhD dissertation.

Petersen de Piñeros, Gabriele. 1994. 2007. "Nominal Classifiers in Uitoto." *IJAL* 73 (4): 389–409.

Piaguaje, Ramón, Elías Piaguaje, Orville E. Johnson, and María de Johnson. 1992. *Vocabulario Secoya*. Quito: Instituto Lingüístico de Verano.

Ramirez, Henri. 1997. *A Fala Tukano dos Ye'pâ-Masa* [The Language of the Ye'pâ-Masa]. Manaus: CEDEM.

———. 2001. *Línguas Arawak da Amazônia Setentrional: Comparação e Descrição* [Arawakan Languages of the Setentrional Amazon: Comparison and Description]. Manaus: Editora da Universidade do Amazonas.

Rankin, Robert L. 2004. "The History and Development of Siouan Positional with Special Attention to Polygrammaticalization in Dhegiha." *Sprachtypologie Und Universalienforschung (STUF)* 57 (2): 202–27. https://doi.org/10.1524/stuf.2004.57.23.202.

Regúnaga, María A. 2012. *Tipología del género en lenguas indígenas de América del Sur*. Bahía Blanca, Argentina: Editorial de la Universidad Nacional del Sur.

Rodríguez Preciado, Iveth Patricia. 2018. *Aspectos de la Morfología Nominal y Verbal de la Lengua Pisamira: Una Lengua de la Familia Tucano Oriental del Noroeste Amazónico*. Universidade Nacional do Brasil, Brasília PhD dissertation.

Rose, Françoise, Magdalena Lemus Serrano, Natalia Eraso, Thiago Chacon, and Françoise Rose. 2017. "A New Look into Arawak-Tukanoan Contact : The Yukuna-Tanimuka Bidirectional Hypothesis." Presented at SSILA Annual Meeting, Austin.

Rybka, Konrad. 2017. "Contact-induced phenomena in Lokono (Arawakan)." In *Boundaries and Bridges*, edited by Kofi Yakpo and Pieter Muysken, 257–82. Boston/Berlin: De Gruyter Mouton.

Sands, Kristina. 1995. "Nominal Classification in Australia." *Anthropological Linguistics* 37 (3): 247–346. https://www.jstor.org/stable/30028415.

Schourup, Lawrence C. 1973. "A Cross-Language Study of Vowel Nasalization." *Ohio State Working Papers in Linguistics* 15: 190–221.

Schwarz, Anna. 2011. "Specificity Efforts in Discourse and the Typology of Nominal Classification." Presented at ALT 9, Hong Kong.

Seifart, Frank. 2005. *The Structure and Use of Shape-Based Noun Classes in Miraña (North West Amazon)*. Radboud Universiteit, Nijmegen PhD dissertation.

———. 2007. "The Prehistory of Nominal Classification in Witotoan Languages" *IJAL* 73 (4): 411–46.

———, and Doris L. Payne. 2007. "Nominal Classification in the North West Amazon: Issues in Areal Diffusion and Typological." *IJAL* 73 (4): 381–87.

Senft, Gunter. 1993. "A Grammaticalization Hypothesis on the Origin of Kilivila Classificatory Particles." *STUF* 46 (2): 100–112. https://doi.org/10.1524/stuf.1993.46.14.100.

———. 2000a. *Systems of Nominal Classification*. Cambridge: Cambridge University Press.

———. 2000b. "What Do We Really Know about Nominal Classifier Systems?" In *Systems of Nominal Classification*, edited by Gunter Senft, 11–49.

Silva, Lucas Ferreira da, and Thiago Costa Chacon. 2017. "Classificadores Nominais em Kubeo, Wanano e Tukano: Uma Análise Comparativa [Nominal Classifiers in Kubeo, Wanano and Tukano: A Comparative Analysis]." In *23º Congresso de Iniciação Científica da UnB e 14º do DF* [23rd Congress of Cientific Initiation of UnB and 14th of DF]. Universidade Nacional do Brasil, Brasília.

Smothermon, Jeffrey R., and Josephine H. Smothemon. 1993. *Masa Ye, Gawa Ye Rãca: Ãmara Tuti [Macuna-Español Diccionario]*. Santafé de Bogotá: Editorial Alberto Lleras Camargo.

———, and Paul S. Frank. 1995. *Bosquejo Del Macuna: Aspectos de La Cultura Material de Los Macunas; Fonología; Gramática*. Santafé de Bogotá: Editorial Alberto Lleras Camargo.

Sorensen, Arthur P. Jr. 1969. *The Morphology of Tukano*. Columbia University, New York PhD dissertation.

Stebbins, Tonya, Bethwyn Evans, and Angela Terril. 2018. "The Papuan Languages of Island Melanesia." In *The Languages and Linguistics of the New Guinea Area: A Comprehensive Guide*, edited by Bill Palmer, 775–894. Berlin/Boston: De Gruyter Mouton. DOI: 10.1515/9783110295252-007.

Steele, Susan. 1978. "Word Order Variation. A Typological Study." In *Universals of Human Language*, edited by J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik, 585–623. Stanford: Stanford University Press.

Stenzel, Kristine. 2013. *A Reference Grammar of Kotiria (Wanano). A Reference Grammar of Kotiria (Wanano)*. Lincoln/London: University of Nebraska Press. https://doi.org/10.2307/j.ctt1ddr99n.

Strom, Clay. 1992. *Retuamã Syntax*. Studies in the Languages of Colombia 3. Dallas, Texas: Summer Institute of Linguistics.

Vallejos, Rosa. 2013. "El Secoya del Putumayo: Aportes Fonológicos para la Reconstrucción del Proto–Tucano Occidental. Línguas Indígenas Americanas." *LIAMES*, 13: 67–100.

———. 2021a. "Nominal Classification without Grammatical Agreement: Evidence from Secoya." *IJAL* 87 (3): 423–455.

———. 2021b. "Secoya Classifiers: Morphosyntac Variation and the Emergence of Grammatical Agreement." Presented at Amazonicas VIII, Góias/Brasília.

Veerman-Leichsenring, Annette. 2004. "Popolocan Noun Classifiers: A Reconstruction." *IJAL* 70 (4): 416–51. https://doi.org/10.1086/429208.

Velie, Daniel, and Virginia Velie. 1981. *Vocabulario Orejon*. Yarinacocha, Pucallpa, Perú: Ministerio de Educación/Instituto Lingüística de Verano.

Vitral, Lorenzo. 1996. "A Forma Cê e a Noção de Gramaticalização [The Form 'cê' and the Notion of Grammaticalisation]." *Revista Estudos Da Linguagem* 4 (1): 115–24.

Waltz, Nathan E. 2007. *Diccionario Bilingüe: Wanano o Guanano-Español, Español-Wanano o Guanano*. Bogotá, D.C.: Editorial Buena Semilla.

———, and Alva Wheeler. 1972. "Proto Tucanoan." In *Comparative Studies in Amerindian Languages*, edited by Esther Matteson, 119–49. De Gruyter Mouton.

West, Birdie. 1980. *Gramatica Popular Del Tucano*. Bogotá, D.C.: Instituto Lingüístico de Verano.

Wheeler, Alva. 1970. *Grammar of the Siona Language*. University of California, Berkeley PhD dissertation.

———. 1987. *Gantëya Bain: El pueblo Siona del Río Putumayo Colombia*. Lomalinda, Meta: Editorial Townsend.

———. 2000. “La lengua Siona.” In *Lenguas Indígenas de Colombia: Una Visión Descriptiva*, edited by María Stella González de Pérez and María Luisa Rodríguez de Montes, 181–98.

Wilkins, David P. 1989. *Mparntwe Arrernte (Aranda): Studies in the Structure and Semantics of Grammar*. The Australian National University, Canberra PhD dissertation. https://doi.org/10.25911/5d78d91feadf1.

———. 2000. “Ants, Ancestors and Medicine: A Semantic and Pragmatic Account of Classifier Constructions in Arrernte (Central Australia).” In *Systems of Nominal Classification*, edited by Gunter Senft, 147–216.

Wilson, Colin. 2001. “Consonant Cluster Neutralisation and Targeted Constraints.” *Phonology* 18: 147–97.

Wojtylak, Katarzyna. 2017. *A Grammar of Murui (Bue): A Witotoan Language of Northwest Amazonia*. James Cook University, Queensland PhD dissertation.

Zavala, Roberto. 2000. “Multiple Classifier Systems in Akatek (Mayan).” In *Systems of Nominal Classification*, edited by Gunter Senft, 114–46.

Zhang, Hong. 2007. “Numeral Classifiers in Mandarin Chinese.” *Journal of East Asian Linguistics* 16 (1): 43–59. https://doi.org/10.1007/s10831-006-9006-9.

Appendix A: References with tables

Throughout this thesis, I refer to sources in tables by Roman numerals. Below, for each of the numerals corresponding to those found in the respective tables, I list relevant references for the forms found in these tables. Tables not listed below are based on Huber and Reed (1992) unless otherwise indicated in the introduction of the table in the text above.

Table A: References with tables

Table	Sources consulted:
5	I) Wheeler (1970); II) Johnson and Levinsohn (1990); III) Cook and Criswell (1993); IV) Eraso (2015); V) Miller (1999); VI) Chacon (2012); VII) Stenzel (2013); VIII) Ramirez (1997, Vol. 1); IIX) Bostrom (1998), and Gomez-Imbert (2007).
11	I) Wheeler (2000); II) Bruil (2014); III) Johnson and Levinsohn (1990); IV) Vallejos (2021a); V) Cook and Crisswell (1993); VI) Farmer (2015); VII) Eraso (2015); VIII) Miller (1999); IX) Ramírez (1997, Vol. 1); X) Stenzel (2013); XI) Chacon (2012); XII) González de Pérez (2000, 392); XIII) Jones and Jones (1991); XIV) Gomez-Imbert (2007); XV) Barnes (1990).
12	I) Bruil (2014, 144); II) Cook and Gralow (2001, 29, 48, 50); Farmer (2015, 235); IV) Miller (1999, 42); Jones and Jones (1991, 55); and VI) Barnes (1990, 78).
13	I) Miller (1999, 39); II) Jones and Jones (1991, 52); III) Barnes (1990, 282); IV) González de Pérez (2000, 384); V) Stenzel (2013, 126); VI) Gomez-Imbert (2007).
14	I) Wheeler (2000, 185); II) Bruil (2014, 141); III) Johnson and Levinsohn (1990, 29); IV) Vallejos (2021a); V) Cook and Crisswell (1993, 21); VI) Chacon (2012, 250); and West (1980, 121).
15	I) Wheeler (2000, 185); II) Bruil (2014, 142); III) Johnson and Levinsohn (1990, 29); IV) Vallejos (2021a); V) Cook and Crisswell (1993, 21); and VI) Creveaux, Sagot, and Adam (1882, ca. 53).
16	I) Wheeler (2000, 186); II) Bruil (2014); III) Johnson and Levinsohn (1990); IV) Vallejos (2021a); V) Farmer (2015, 173); VI) Cook and Criswell (1993).
17	I) Eraso (2015, 203); II) Strom (1992, 54); III) Miller (1999, 52); IV) Chacon (2012, 249); V) Jones and Jones (1991, 42); VI) Smorthermon and Smothermon (1995, 37); VII) Barnes (1990, 287); VIII) González de Pérez (2000, 384); IX) Metzger (1981, 175); X) Stenzel (2013); XI) Ramirez (1997, 111); XII) Gomez-Imbert (1982, 316).
21	I) Huber and Reed (1992, 159); II) Johnson and Levinsohn (1990, 79); III) Vallejos (2021a); IV) Farmer (2015, 233); V) Cook and Gralow (2001, 30); VI) Huber and Reed (see (I)), Maxwell and Morse (1999, 180), and Morse, Salser, and de Salser (1999, 99); VII) Barnes (1990, 275); VIII) Rodriguez Preciado (2018, 73).
24	I) Wheeler (2000); II) Bruil (2014, 139); III) Johnson and Levinsohn (1990); IV) Vallejos (2021a); V) Cook and Crisswell (1993); VI) Farmer (2015); VII) Eraso (2015); VIII) Miller (1999, 37), and De Lima Silva (2012, 153); IX) Stenzel (2013); X) Chacon (2012); XI) Jones and Jones (1991); XII) Gomez-Imbert (1982, 107); and XIII) Barnes (1990, 274).

26	I) Wheeler (1970, Vol. 2, 138); II) Bruil (2014, 148, 160); III) Johnson and Levinsohn (1990), and Piaguaje et al. (1992, 106); IV) Vallejos (2021a); V) Cook and Criswell (1993), and Cook and Gralow (2001); VI) Huber and Reed (1992), and Farmer (2015); VII) Eraso (2015, 242); VIII) Miller (1999, 23), and De Lima Silva (2012, 184); IX) Ramírez (1997, Vol. 2, 267, 279); X) Stenzel (2013, 178, 464); XI) Chacon (2012, 123, 247); XII) Gonzalez Muñoz (2016, 78); XIII) Huber and Reed (see VI), and Jones and Jones (1991); XIV) Huber and Reed (see VI), and Gomez-Imbert (2007); XV) Smothermon and Smothermon (1995); XVI) Huber and Reed (see VI).

Appendix B: Classifier cognates

The cognates and their reconstructions discussed in this thesis are listed below. Since a number of classifiers are limited to a single branch of the family, and for ease of exposition, I show the cognates in two tables: Table B1 for the ET branch, and Table B2 for the WT branch. In Table B1 below, I have tentatively assumed that PET *-*jõ* 'CL:PALM' indirectly has **tjũkkɨ* as its source, and that TUY and PIS -*wõ* 'CL:PALM' are cognates of -*jõ* but these are by no means certainties (see Section 6.2.1.2), hence the question mark in *-*jõ?* and *-*wõ?* below. Since the TAN and LET classifiers are fairly similar in form and function, I have grouped them together here.

Lastly, as is demonstrated below, a number of markers display a form of allomorphy which can be reconstructed for to the PT stage. This allomorphy may be the result of a lenis/fortis distinction such as the one reconstructed for bound **p'*/*p* in Chacon and Michael (2018) (see Appendix C). Moreover, the fact that the ET cognates show more forms with this allomorphy may indicate that these forms are either more archaic, or that the two forms were retained longer in this branch as opposed to the WT branch.

Table B1: ET classifier cognates, reconstructions, and etymologies

PT Gloss	PT	PET	PET Gloss	TAT	DES	TUK	KOT	KUB	PIS	BSA	TAN LET	MAK
'M'	*-k'ɨ *-kɨ	*-k'ɨ *-kɨ	'M'	-kɨ -ɨ	-gɨ	-gɨ	-ɨ	-kɨ̃	-kɨ -ɨ -gɨ	-ɨ	-ki -i	-ɨ
'F'	*-k'o *-ko	*-k'o *-ko	'F'	-o -ko	-go	-go	-o -ko -koro	-ko	-o -ko -go	-o	-o -ko	-o
'INAN'	*-k'a *-ka	*ga *a *ka	'CL:3D'	-a -ka		-ga -kha	-ka	-ka	-ga	-ga -a -ka	-ka	-ga -ka
'LOC'	*-t'o *-to	*-t'o	'CL:CONCAVE'	-to -ro -~do	-ru		-to -ro	-do	-rõ	-ro -to		-tẽ-ro
'path'	*-~p'a *-~pa	*-~ba *-~pa	'CL:PATH'	~wa	-bã		-~bá		-bã	-bã		
'tree'	*tjũkkɨ	*tjũkkɨ	'CL:TREE' 'CL:CILINDRICAL' 'CL:WOVEN'	-ɨ -kɨ	-gɨ	-gɨ -khɨ	-kʉ	-kɨ	-gɨ	-ɨ -gɨ -kɨ	-~hukí -~húkí	-gɨ -kɨ
		*-jõ?	'CL:PALM'	-~jo	-jũ		-~jo	-jɨ̃	-wõ?	-dõ		

‘flat’	*-*p’a* *-*~páia*	*-*p’a* *-*~páia*	‘CL:FLAT’	-*~pái*	-*pa*	-*wa* -*pha*		-*wa*		-*baja* -*hãi*	-*~ɸáí*	-*hãi*
‘river’	**tj’ia*	*-*ja*	‘CL:RIVER’	-*ja*	-*ja*		N/A	-*ja*		-*sa* -*ja*	-*riá* -*ríá*	
‘lake’	**tjitta* *-*t’a*	**-t’a-t’o?*	‘CL:LAKE’	N/A		-*ra*	-*taro*			-*ra*		

Table B2: WT classifier cognates, reconstructions, and etymologies

PT Gloss	PT	PWT	PWT Gloss	E. Sio	C. Sio	E. Sek	P. Sek	Kor	Mai
‘M’	*-*k’ɨ* *-*kɨ*	*-*k’ɨ* *-*kɨ*	‘M’	-*ɨ* -*kɨ*	-*gɨ*	-*kɨ* -*i*	-*kɨ* -*i*	-*ɨ*	-*ɨ* -*kɨ* -*i*
‘F’	*-*k’o* *-*ko*	*-*k’o* *-*ko*	‘F’	-*o* -*ko*	-*go*	-*o* -*ko*	-*o* -*ko*	-*o*	-*o* -*ko*
‘INAN'	*-*k’a* *-*ka*	*-*k’a* *-*ka*	‘CL:3D’	-*ka*	-*ga*	-*kã*	-*ka*	-*a* (SG) -*kʰa-a-ã* (PL)	-*ga* (SG) -*gaña* (PL)
‘tree’	**tjũkkɨ*	*-*jɨ̃*	‘CL:TREE’	-*jɨ̃*	-*jɨ*	-*ñɨ*	-*jɨ*	-*ñɨ*	-*ñi*
‘LOC’	*-*t’o*	*-*t’o*	‘CL:CONCAVE’	-*do*	-*do*	-*ro*	-*ro*	-*ro*	-*ro* -*to*
‘flat’	*-*p’a*	*-*p’a*	‘CL:FLAT’	-*bã*	-*ba*	-*pa*	-*pa*	-*pa*	
?	N/A	*-*wɨ*	‘CL:CONTAIN’	-*wɨ*	-*wɨ*	-*wɨ*	-*wɨ*	-*wɨ*	-*bɨ*
‘to weave’	**tɨõ-*	**t’ɨ*	‘CL:WOVEN’	-*dɨ*	-*dɨ*	-*rɨ*	-*rɨ*	-*rɨ*	-*rɨ* -*tɨ*
‘river’	**tj’ia*	*-*ja*	‘CL:RIVER’	-*ja*	-*ja*	-*ja*	-*ja*	-*d͡za*	-*ja*
‘lake’	**tjitta* *-*t’a*	*-*t’a*	‘CL:LAKE’	-*da*	-*da*	-*ra*	-*ra*	-*ra* -*kʰa*	-*ra*

Appendix C: Relevant sound correspondences in bound and free forms

In Table C, I summarise the reflexes of consonants in the Tukanoan languages, both in bound forms, i.e. suffixes, and in free forms, i.e. roots. For ease of exposition, I do not specify the phonological contexts that trigger the variations within each language's free forms, and refer the reader to Chacon (2014) for a fuller account of these contexts. For now, it suffices to note that it is not possible to judge whether the bound forms developed from the intervocalic reflexes or word-initial reflexes, since there are not enough known cognates to make a sound judgement, and since in a number of instances, the bound forms display a different reflex altogether. This was already mentioned in Chacon (ibid.), and further discussed for bound **p'* and **p* in Chacon and Michael (2018), who argue that PT had a lenis/fortis distinction where the lenis form **p'* occurred in suffix groups that had an odd number of morae, and fortis **p* in suffix groups that had an even number of morae, although this distinction is now lexically conditioned (id., 72). As demonstrated below, a fairly similar distinction can be found with regard to: PT **k'* with bound reflexes *g*, *k*, ∅; PT **k* with bound reflexes *g*, *k*, *kʰ*; PT **tj'* with bound reflexes *s*, *r*, *j*; PT **tj* –or more specifically **~tj*, as I only found this sound in bound forms that are nasalised– with bound reflexes *s*, *ñ*, *~j*, *~d*, *~h*, and perhaps also *~w* (see Section 6.2.1); PT **t'* with bound reflexes *t'*, *d*, *ɖ*, *r*, *ɾ*, *t*, *tʰ*, *ʔd*; and PT **t* with bound reflexes *d*, *r*, *t*. Since classifiers with a single source (see Appendix B) can have multiple reflexes in a single language, it is possible that a similar fortis/lenis distinction played a role in the development of the reflexes listed here.

Furthermore, I have grouped together TAN and LET, and E. SEK and P. SEK respectively, since the reflexes listed below are very similar or identical within these two language pairs. There are a number of differences within these language pairs, including phonological differences, but none that are of particular relevance here (see Strom 1991; Eraso 2015; Vallejos 2013).

Table C: Sound correspondences and reconstructions of free and bound consonants.

	PT		Context	PET	TAT	DES	TUK	KOT	KUB	PIS	BSA	TAN / LET	MAK	PWT	E. SIO	C. SIO	SEK	KOR	MAI
L	*p'	FREE	[_	*p'	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*b*	*p'	*p'*	*p'*	*p*	*p*	*ʔb*
			V _ V]STEM		*p*		V'*p*	V'*p*	*b*	*p*		V'*p*			*β*	*h*	*h*	*h*	*h*
			~V _ V]STEM		*b*		V'*b*	V'*b*	*b*	*b*		V'*b*			*β*	*h*	*h*	*h*	*h*
			[_ V*p*		*b*		*b*	*b*	*p*	*b*		*p*			*p'*	*h*	*h*	*h*	*h*
		BOUND		*w, *b	*w*	*b*	*w*	*b?*	*w*	?	*b*	?	*b*	*p', *w	*p', w*	*p', w*	*p, w*	?	*b*
F	*p	FREE	[_	*p	*p*	*p*	*p*	*pʰ*	*p*	*p*	*h*	*p* (*ɸ*)	*h*	*h, *ø	*h*	*h*	*h*	*h*	*h*
			V _ V					*p*											
		BOUND		*p, *b	?	*m*	*p, m*	?	*b*	?	*m*	?	*m*	*h, *ø	*h, ø*	*h, ø*	*h, ø*	?	*h*

L	*k’	FREE	[_	*g	∅	*g*	∅	*k*ʰ	*k*	*k*	*g*	∅	*g*	*k, *k	*k’*	*k’*	*k*	*k*	*g, k*
			V _ V]STEM		*k*		V’*k*	V’*k*	∅			*k*			*k*	*k*		*k*ʰ	∅
		BOUND		*g, *k	∅	*g*	*g*	∅	*k*	*g*, ∅	*g*, ∅	∅	*g*	*g, *∅	*g*, ∅	*g*	*g*, ∅	∅	*g*
F	*k	FREE	[_	*k	*k*	*k*	*k*	*k*ʰ	*k*	*k*	*k*	*k*	*k*	*k*	*k*ʰ	*k*	*k*	*k*	*k*
			V _ V]STEM			*g*		*k*	*k*						*k*				
			i,e__]STEM			*k*		*tʃ*	*k*						*k*				
			~V _ V			*g*		*k*	∅						*k*				
		BOUND		*k	*k*	*g?*	*k*ʰ	*k*	*k*	*k*	*k*	*k*	*k*	*k*	*k*	?	*k*	*k*ʰ	?
L	*tj’	FREE	[_	*tj, *d	*j*	*j*	*j*	*j*	*h*	*j*	*j*	*j*	*j*	*tj’	*s’*	*s’*	*s’*	d͡z	*d*
			V _ V]STEM		*r*	*d*	*d*	*d*		*d*	*r*	*r*	*r*						*j*
		BOUND		*j, *dj	*j*	*j*	?	?	*j*	?	*s, j*	*r*	?	*j	*j*	*j*	*j*	d͡z	*j*
F	*~tj	FREE	[_	*~tj	~*j*	~*j*	~*j*	~*j*	*h*	~*j*	~*j*	*h*	~*j*	*j	*s*	*s*	*s*	*ñ*	*d*
			V _ V]STEM		*h*	*h*	*s*	*s*		*tʃ*	*c*		*c*						
		BOUND		*~j, *~dj?	~*j*	~*j*	?	~*j*	~*j*	~*w?*	~*d*	~*h*	?	*~j	~*j*	*j*	*ñ*	*ñ*	*ñ*
L	*t’	FREE	[_	*d	*d>r*	*d*	*d*	*d*	*d*	*d*	*r*	*r*	*r*	*t’	*t’*	*ɖ*	*d*	*r*	*ʔd*
			V _ V]STEM		*r*		V’*t*	V’*t*		*t*	*d*	*t*	*d*		*ɾ*	*t*	*t*	*t*ʰ	*t*
			~V _ V]STEM		*d*		V’*d*	V’*d*		*d*	*d*	V’*r*	*d*		*ɾ*	*t*	*t*	*t*	*t*
		BOUND		*d	*r, ~d*	*r*	*r*	*r*	*d*	*r?*	*r*	*?*	*r*	*t’	*d*	*d*	*r*	*r*	*r*
F	*t	FREE	[_	*t	*t*	*t*	*t*	*t*ʰ	*t*	*t*	*t*	*t*	*t*	*t	*t*ʰ	*t*	*t*	*t*ʰ	*t*
			V _ V]STEM			*d*		*t*	*d*			*t*			*t*				
			~V _ V]STEM			*d*		*t*	*d*			*d*			*t*				
		Bound		*t	*t*	*r*	*r*	*t*	*d*	*r?*	*t*	?	?	*t?	?	?	?	?	*t*

Appendix D: Overview of the functions of noun categorization marking in Tukanoan languages

In order to have the data presented below fit in one table, it has been necessary to make extensive use of abbreviations. The abbreviations differ in meaning per row and are as follows, from left to right:

- **CLASSIFIER-GENDER OPPOSITION**: A) No morphosyntactic difference; B) Gender markers may occur on verbs, but classifiers may not; C) Gender markers may occur on verbs, classifiers may not, and there is some further difference in morphos-syntactic distribution between the two.
- **NOUN CLASSIFIERS**: A) Only gender markers may occur on nouns; B) Only classifiers may occur on nouns; C) Both gender markers and classifiers may occur on nouns.
- **FUNCTION OF NOUN CLASSIFIERS**: A) Derive nouns; B) Both derive and individuate.
- **QUANTIFIER CLASSIFIERS:** A) Only gender may occur on quantifiers; B) Both gender and classifiers may occur on quantifiers; C) Not applicable, or no nominal categorization markers may occur on quantifiers; D) Unknown.
- **NUMERAL, QUALIFIER, DEMONSTRTIVE CLASSIFIERS, CLASSIFIERS IN POSSESSIVE CONSTRUCTIONS, AND OTHER TYPES OF CLASSIFIERS:** A) Only gender may occur on numerals; B) Only classifiers may occur on numerals; C) Both gender and classifiers may occur on numerals; C) Not applicable, or no nominal categorization markers may occur on numerals; D) Unknown.

Table D: Overview of the functions of noun categorization marking in Tukanoan languages

Language	Glottocode	Branch	CL-GENDER opposition	N	Funct. N-CL	QUANT	NUM	QUAL	DEM	POSS	OTHER
C. SIO	sion1247	WT	B	C	B	C	C	C	C	D	D
E. Sio	ecua1247	WT	C	C	B	C	C	C	C	E	A
E. Sek	seco1241	WT	C	C	B	E	C	B	C	D	C
P. SEK	seco1241	WT	B	C	B	D	D	D	D	D	D
MAI	orej1242	WT	B	C	B	E	C	C	C	D	C
KOR	kore1283	WT	C	C	B	E	C	C	C	D	C
TAN	tani1258	ET	C	A	B	D	C	C	C	D	C
LET	ret1239	ET	C	A	A	D	B	D	D	D	D
DES	desa1247	ET	A	B	A	C	C	C	C	D	C
KUB	cube1242	ET	B	C	B	C	C	C	C	D	C
BSA	bara1380	ET	A	C	A	A	C	C	B	C	C
TUY	tuyu1244	ET	B	C	A	C	C	C	C	D	C

PIS	pisa1245	ET	C	C	B	E	B	C	C	D	D
KAR	cara1272	ET	B	C	A	C	C	C	C	C	C
KOT	guan1269	ET	C	C	B	C	A	C	C	D	C
TUK	tuca1252	ET	C	C	B	A	C	C	C	C	D
TAT	tuto1247	ET	C	C	B	B	C	C	C	D	A