

Chapter 9

Gender in New Guinea

Erik Svärd

earlier Stockholm University

The present study classifies gender systems of 20 languages in the New Guinea region, an often neglected area in typological research, according to five criteria used by Di Garbo (2014) for African languages. The results show that gender in New Guinea is diverse, although around half of the languages have two-gendered sex-based systems with semantic assignment, more than four gender-indexing targets, and no gender marking on nouns. The gender systems of New Guinea are remarkably representative of the world, although formal assignment is underrepresented. However, the gender systems of New Guinea and Africa are very different. The most significant difference is the prevalence of non-sex-based gender systems and gender marking on nouns in Africa, whereas the opposite is true in New Guinea. Finally, four typologically rare characteristics are singled out: (1) size and shape as important criteria of gender assignment, with large/long being masculine and small/short feminine, (2) the co-existence of two separate nominal classification systems, (3) no gender distinctions in pronouns, and (4) verbs as the most common indexing target.

Keywords: agreement, grammatical gender, indexation, New Guinea, Papuan, typology.

1 Introduction and background

Most typological research on gender has focused on languages in Eurasia, Africa, Australia, and the Americas. Less research has been conducted in the region of New Guinea, which contains as many as one sixth of all languages of the world. In recent descriptions, languages of New Guinea of highly variable genealogical affiliation have been shown to exhibit many unusual gender systems. This is important for the study of gender as gender systems are often very stable and not prone to borrowing. However, little has been done to survey the diversity of



gender in New Guinea. The purpose of this paper is to counteract this issue by investigating 20 New Guinean languages, both Papuan and non-Papuan, for which gender has been described and to compare their gender systems in an areal and a typological perspective. Specifically, the research questions are:

- How is grammatical gender expressed in a diverse sample of 20 languages of New Guinea?
- How do the gender systems of New Guinea compare with other geographical areas (notably Africa) and the world as a whole?
- Are there any phenomena in gender which are unique to or surprisingly common in the languages of New Guinea?

In order to investigate this, five criteria are used to classify the gender systems of New Guinea. The distribution of values of these criteria are then compared with the world in general and Africa in particular.

1.1 Defining gender

Hockett (1958: 231) defines gender as “classes of nouns reflected in the behavior of associated words”. In other words, gender is conceived of as noun classes triggering agreement. The idea of gender as based on the behavior of associated words is reflected in the focus on agreement, which Corbett (1991: 4) calls the determining criterion of gender. In order to define gender, Corbett presents Steele’s (1978) description of agreement:

The term *agreement* commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another. For example, adjectives may take some formal indication of the number and gender of the noun they modify.
(Steele 1978: 610 as cited in Corbett 1991: 105)

According to Corbett, agreement is an asymmetric relationship between the controller (i.e., the element determining agreement, e.g., subject noun phrase) and the target (i.e., the element whose form is determined by agreement) (Corbett 2006: 4). Importantly, Corbett adopts a ‘canonical approach’: that is, the basis for Corbett’s discussion are those ‘canonical’ instances which are best and clearest but not necessarily the most frequent (Corbett 2006: 9). Canonical agreement can be summarized as follows (adapted from Corbett 2006: 9):

- the controller is present, has overt expressions of features, and is consistent in agreement, and its part of speech is not relevant;
- the target has bound expression of agreement, obligatory and regular marking which is doubling the marking of the noun, has a single controller, and its part of speech is not relevant;
- the domain in which agreement occurs is local, and it is one of multiple domains.

More recently, Di Garbo (2014: 8) gives a few examples illustrating the fact that in many languages both pronouns and noun phrase-internal targets do not presuppose a syntactic antecedent or controller. In order to counter this, Di Garbo (2014) uses the term *indexation* instead, following Croft (2001; 2003; 2013) and Iemmolo (2011). In this definition, indexation is used to refer to grammatical strategies signaling (i) lexical and grammatical properties of nouns, and (ii) semantic properties of NP referents, which are independent of the presence of any overt syntactic antecedent (Di Garbo 2014: 8). Following Di Garbo, the following terms are used in this study (adapted from Di Garbo 2014: 8):

- *indexing target* or *index* refers to entities with inflectional morphology signaling gender;
- *syntactic antecedent* refers to the NP indexed by the pronominal target;
- *indexation trigger* or *trigger* refers to the entities that activate the use of a certain indexation pattern in a given discourse domain.

Despite the difference in terminology, the end result of both agreement in Corbett (1991) and indexation in Di Garbo (2014) is the same, with both being cover terms for the same linguistic feature. Since this is mainly a typological study, its purpose is to be comparable with earlier and future typological research on gender without relying on theoretical concepts that are as yet not widely accepted. However, since indexation is gaining ground, it is embraced in this chapter.

1.2 Gender research on New Guinea

Although gender has not been extensively researched in New Guinea, the region shows much promise for exhibiting a high variety of gender systems. The New Guinea region is home to approximately 1,200 languages belonging to around three dozen language families spoken in an area smaller than 900,000 km², which

makes it the most linguistically diverse region in the world (Foley 2000: 357). Nevertheless, there are two dominating language families: the Austronesian family, spoken in the coastal areas, and the Trans-New Guinean (TNG) family, which is concentrated to the mountainous inland. The Austronesian and the TNG languages comprise around 300 languages each and typically do not show gender, although there are some important exceptions (Foley 2000: 358–363). Thus, gender is lacking at least in approximately half of the languages of New Guinea. As for the remaining languages, gender is found in the West Papuan, Sko, and Sepik languages, as well as several isolates such as Yava, Burmeso, and Kuot (Foley 2000: 371).¹ Gender is also present in Torricelli and Lower Sepik-Ramu languages, but as parts of larger and more complex systems of noun classification (Foley 2000: 371). It also occurs in some isolated cases in the TNG family, such as Nalca (Mek) (Svärd 2013) and the Ok languages, e.g., Mian (Fedden 2011), and in very few Austronesian languages, including Teop (Oceanic) (Mosel & Spriggs 2000). By counting these gendered languages based on the numbers given by Foley (2000), gender in New Guinea can be estimated to occur in at least 120 languages of different families and isolates. The genealogical diversity suggests that gender may be highly diverse in New Guinea.

However, Foley suggests that gendered languages of New Guinea have some features in common, including the presence of gender assignment based on specific criteria of size and shape, as well as the presence of languages with two separate systems (Foley 2000; Svärd 2015: 8–9). Combined with the observation that gender in New Guinea is concentrated in languages with high genealogical diversity, this suggests that gender may be highly diverse in New Guinea.

2 Method and data

The sampling method used in this study is a *variety sample* (Bakker 2012). Rather than trying to represent the real population of languages as would be achieved by a probability sample, the sample is designed to achieve the largest variety of results in regard to the chosen feature, while entirely omitting languages lacking the feature.

In this study, the sample is restricted to New Guinea as delimited by Foley (2000: 357), including New Guinea proper as well as surrounding islands. First and foremost, the sample includes only languages with gender. Secondly, the languages were chosen from as many families as possible, as far as the availability

¹Foley (2000: 371) also mentions the Sulka language of New Britain, but there are no indications of a gender system in the grammar by Tharp (1996).

of material permitted, while still accounting for variation within families if there were reasons to do so. This was primarily based on the information by Foley (2000) and others.

Table 1 lists the languages of the sample together with family, genus, ISO code, and source, along with a map of the languages shown in Figure 1. The names primarily follow Glottolog, except for Motuna (Glottolog: Siwai), where I follow Onishi (1994). Also, the Glottolog form Warapu is used despite Barupu occurring in Corris (2005). Furthermore, language families and genera are based on Glottolog, so that a genus in the table below does not always agree with the genus for the same language in *WALS*.

Table 1: The language sample. The en dash indicates no grouping or that the language is itself the closest node to the family node.

Family	Genus	ISO	Language	Source
Austronesian, Oceanic	Nehan-North Bougainville	tio	Teop	Mosel & Spriggs (2000)
Isolate	–	gpn	Taiap	Kulick & Stroud (1992)
Isolate	–	bzu	Burmeso	Donohue (2001)
Isolate	–	kto	Kuot	Lindström (2002)
Left May	–	amm	Ama	Årsjö (1999)
Lower Sepik-Ramu	Lower Sepik	yee	Yimas	Foley (1991)
Ndu	–	mle	Manambu	Aikhenvald (2008)
North Bougainville	–	roo	Rotokas	Robinson (2011)
Sepik	–	aau	Abau	Lock (2011)
		sim	Mende	Hoel et al. (1994)
Sko	–	skv	Skou	Donohue (2004)
		wra	Warapu/Barupu	Corris (2005)
South Bougainville	–	siw	Motuna/Siwai	Onishi (1994)
Torricelli	–	avt	Au	Scorza (1985)
	Arapesh	ape	Bukiyip	Conrad & Wogiga (1991)
	West Palai	van	Walman	Brown & Dryer (2008)
Trans-New Guinea	Mek	nlc	Nalca	Svärd (2013); Wälchli (2018)
	Ok-Oksapmin	mpt	Mian	Fedden (2011)
		opm	Oksapmin	Loughnane (2009)
West Papuan ²	–	ayz	Maybrat	Dol (2007)

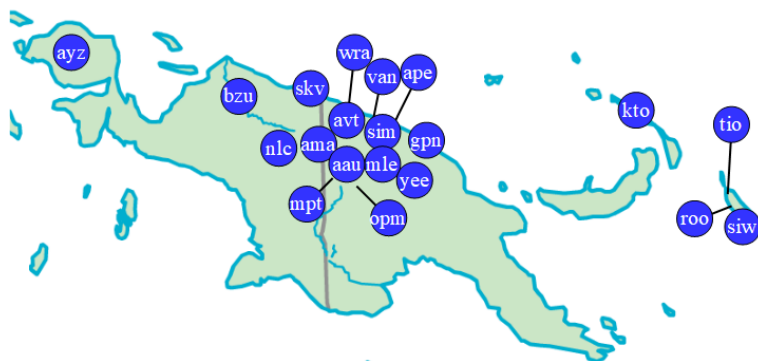


Figure 1: The geographical locations of the languages in the sample labeled with ISO codes

The main sources of data used in this study are reference grammars, which are listed for each language in Table 1 above. However, many descriptions do not mention the language as having a gender system if gender only occurs in pronouns. Therefore it was also necessary to examine the sections on pronouns. If the available descriptions for a language neither mentioned gender nor showed it directly in the section(s) about pronouns or in glossed examples, the language was not considered to be eligible for the sample.

In order to make the languages of the study typologically comparable, the study employs five classificatory criteria used by Di Garbo (2014) to classify the gender systems of Africa, viz.,

- Sex-based and non-sex-based gender systems.
- Number of genders.
- Gender assignment.
- Number of gender-indexing targets.
- Occurrence of gender marking on nouns.

Di Garbo also uses other classificatory criteria in order to investigate the interactions of gender and number, and gender and evaluative morphology. However,

²More recent studies suggest that the traditional West Papuan Phylum is probably not an accurate genealogical grouping, but instead consists of as many as seven unrelated language groups (Dol 2007: 5). Since the exact position of Maybrat in such a regrouping is unknown to the present author, West Papuan is kept here as proxy to a genealogical family.

this study is not aimed to directly investigate these interactions, and thus only the criteria above were chosen.

An important advantage of adopting Di Garbo's approach is that this makes the results for New Guinea directly comparable with Africa as for the selected criteria. In addition, since the first three criteria are the same as the ones used in the *WALS* chapters by Corbett (Corbett 2013a,b,c), most of the results are comparable to a worldwide sample. In order to illustrate the distributions, maps were created using the Interactive Reference Tool of the *World atlas of language structures (WALS)*³ using ISO codes and coordinates from Glottolog.

3 Overview of gender characteristics

In the following sections, the distribution of values of the criteria mentioned in §3 are presented and discussed. Each criterion is discussed with the values shown in a table, followed by some examples of the feature in the sample. In §4, these results are discussed from a typological perspective.

It is important to point out that five languages of the sample were found to have two separate systems of noun classification. As will be discussed in §5.2, only Burmeso exhibits two equivalent gender systems, whereas the other four rather distinguish between gender and noun classifiers. For this reason, the two gender systems of Burmeso will be combined for the purpose of comparison in this chapter, although the values assigned to the separate systems will be given in parenthesis whenever applicable.

3.1 Sex-based and non-sex-based gender systems

Following Di Garbo (2014: 62), each gender system is classified as either sex-based or non-sex-based based according to the typology by Corbett (2013c). Sex-based are those where the gender assignment is based at least partly on natural gender, which often surfaces as masculine-feminine distinctions. Consequently, non-sex-based gender systems are those where gender is not based on natural gender. However, according to Corbett (2013c), all non-sex-based systems are based on some notion of animacy.

As shown in Table 2 and Figure 2, sex-based systems are by far the most common ones, with 19 of 20 languages having natural gender as their semantic core. Only the Austronesian language Teop exhibits a non-sex-based system.

Sex-based gender systems present some difficulty in assigning nouns denoting inanimate referents. Non-sex-based systems, i.e., systems based on animacy, can

³See <http://www.eva.mpg.de/lingua/research/tool.php>.

Table 2: Sex-based and non-sex-based gender systems in the sample

Sex-based or non-sex-based	No. of lgs.	%	Languages
Sex-based	19	95%	Abau Ama Au Bukiyip Burmese Kuot Manambu Maybrat Mende Mian Motuna Nalca Oksapmin Rotokas Skou Taiap Walman Warapu Yimas
Non-sex-based	1	5%	Teop
Total:	20	100%	



Figure 2: Sex-based and non-sex-based systems. Colors indicate: sex-based (blue) and non-sex-based (red).

potentially assign every noun according to animacy alone. However, sex-based systems do not by definition have any specific way of assigning nouns that refer to objects without natural gender. Thus, based on how inanimate nouns are assigned gender, the sex-based gender systems in the sample can be further divided into three groups where inanimates are assigned to

1. one of the sex-based genders,
2. both of the sex-based genders based on other criteria, or
3. one or more other non-sex-based genders.

As will be discussed in §3.2, almost half of the languages in the sample (9 of 20) have only two genders, both of which are sex-based. Thus, since option 3 is only available in languages with more than two genders, almost half of the languages in the sample assign inanimate nouns to one of the sex-based genders.

Assigning inanimates to only one of the two genders occurs e.g., in Mende (Sepik), where gender is distinguished only in second and third person singular pronouns. For animate referents, the form of the pronoun is determined by the sex of the referent, while inanimates are usually referred to with the feminine forms (Hoel et al. 1994: 17). An example of this is shown in (1), where *Max* (male name) (1a) and *Lusi* (female name) (1b) occur with the masculine and feminine pronoun forms respectively, and the inanimate *masiji* ‘hair’ (1c) is referred to with the feminine form. Mende thus distinguishes masculine vs. other.

- (1) Mende (Sepik) (Hoel et al. 1994: 19, 31, 46)
- a. *Max wasilaka ri-a*
M. big 3SG.M-INTEN
‘Max is big.’
 - b. *Lusi kava awu-n u-nda sir-a*
L. bad⁴ fight-OBJ do-HAB 3SG.F-INTEN
‘Lusi is a good fighter.’
 - c. *masiji-n tivi unak si horngo-ku-a*
hair-OBJ tie so.that.not 3SG.F loosen-FUT-INTEN
‘Tie the hair so that it won’t loosen.’

⁴When used with the habitual *-nda*, *kava* ‘bad’ functions as an intensifier (Hoel et al. 1994: 31).

Assigning inanimates to both sex-based genders based on other criteria is more common in the sample. In most languages, the assignment of inanimates is based on semantic criteria, most commonly on the criteria shape and size (see also §5.1 below). One such a language is Abau (Sepik), where three-dimensional or long or extended objects, as well as liquids are masculine, whereas two-dimensional, flat or round objects with little height as well as abstract entities are feminine (Lock 2011: 47). Thus, *su* ‘coconut’ (round), *now* ‘tree’ (long), and *hu* ‘water’ (liquid) are masculine, while *iha* ‘hand’ (flat) and *hne* ‘bird’s nest’ (round with little height) are feminine (Lock 2011: 48–50). In a language such as Abau, this is very much based on the speaker’s perception. This can be seen in (2); when referring to the tree from which he makes the paddle (2a), *youk* ‘paddle’ is masculine, since the tree is long and not at all round or flat. However, when referring to the actual paddle (2a), which has the salient features of flat and round, the feminine form is used.

(2) Abau (Sepik) (Lock 2011: 50)

- a. *Ha-kwe youk se seyr.*
 1SG.SBJ-TOP paddle 3SG.M.OBJ cut
 ‘I cut the ‘paddle’ tree.’
- b. *Ha-kwe youk ke lira.*
 1SG.SBJ-TOP paddle 3SG.F.OBJ see
 ‘I see the paddle.’

The third type of sex-based systems is one where inanimates are assigned to genders other than sex-based ones. Naturally, this can only occur in languages with more than two genders. An example of a language with such a system is Nalca (TNG, Mek) (Svärd 2013; Wälchli 2018). Nalca has five main genders: masculine, feminine, neuter, default, and non-noun. As shown in (3), these are apparent in a set of case marker hosts following the NP, which constitute the only indexing target in Nalca. The masculine and feminine genders are used exclusively for nouns denoting male and female humans respectively. Inanimates are divided between the neuter and default genders: the neuter contains all nouns of the phonological structure (C)V (including at least one noun denoting humans, *me* ‘son, child’), while most inanimate nouns belong to the residual default gender. The default gender also contains some gender-neutral nouns denoting humans, most of which are plural, e.g., *nang* ‘people’. The non-noun gender is used e.g., with adverbs, locatives, and despite its name the nominalizer *-a*. It is also used when gender is switched off, in which case nouns still trigger agreement but

due to syntactic phenomena agree with the non-noun gender.⁵ In the examples below, both the neuter *si* ‘name’ and the masculine name *Zakheus* ‘Zacchaeus’ are shown in (3a), the feminine *genong* ‘mother’ in (3b), the default (DEFAULT) *pik* ‘way’ in (3c), and the two non-noun (NNOUN) constructions in (3d). The first instance of non-noun gender in (3d) is due to the intervention of the quantifier *nauba* ‘many’ between *nimi* ‘men’, which belongs to the default gender, and the case marker host, whereas the second is due to the nominalizer *-a*’.

(3) Nalca (TNG, Mek) (own examples)

- a. *alja si ne-ra Zakheus be-k u-lu-m-ok*
 3SG.GEN name N-TOP Z. M-ABS be-IPFV-PST.3SG
 ‘a man called by name Zacchaeus’ (Lk 19:2)⁶
 lit. ‘his name was Zacchaeus’
- b. *Nadya genong ge-ra heknya do?*
 1SG.GEN mother F-TOP who Q
 ‘Who is my mother?’ (Mk 12:48)
- c. *Na bi-nim-na pik e-ra ugun-da ella*
 1SG go-FUT-PRS.1SG way DEFAULT-TOP 2PL-TOP knowledge
u-lu-lum ...
 be-IPFV-PRS.2PL
 ‘And you know the way where I am going.’ (Jn 14:4)
- d. *... nimi nauba a-ra seleb longo-m-ek-a’*
 men many NNOUN-TOP already assemble-PRF-PST.3PL-NMLZ
a-k eib-ok
 NNOUN-ABS see[PFV]-PST.3SG
 ‘... he saw the large crowds...’ (Mt 6:34; lit. ‘he saw that many men had assembled’)

Finally, the only non-sex-based gender system in the sample occurs in the Austronesian language Teop, which has two genders (I and II) with two subgenera for the first gender (I-E and I-A), reflecting the form of the singular article preceding nouns. The genders and the nouns that belong to them are:

⁵The concept of switching gender on and off is an extremely rare phenomenon and goes well beyond the bounds of this study. For a comprehensive description of the Nalca gender system and discussion on switching gender on and off, see Wälchli (2018).

⁶The overwhelming majority of data available in Nalca consists of a translation of the New Testament. The English translation used is the American Standard Version, whereas the glossings and literal translations were devised by the present author. For a description and discussion of the methodology, see Svärd (2013) and Wälchli (2018).

- *Gender I-E*: Contains all proper names, kinship terms, and nouns denoting pets or humans with a particular communal or important social status (Mosel & Spriggs 2000: 334–335).
- *Gender I-A*: Contains most nouns and can be considered the unmarked gender (Mosel & Spriggs 2000: 336–338).
- *Gender II*: Contains names of plants and their parts (but not fruits), objects made of plant material, invertebrates without legs, and many mass and abstract nouns (Mosel & Spriggs 2000: 338).

This is strikingly similar to the noun classification system found in Siar (Frowein 2011; not in the sample), spoken on the opposite coast. Siar does not have a true gender system, since it only shows gender on articles preceding nouns and thus does not exhibit indexation.⁷ However, nouns are still assigned according to a system of nominal classification similar to Teop:

- *Proper*: Contains mostly names, kinship terms and other nouns closely related to humans and culture such as professions (Frowein 2011: 104–105).
- *Common 1*: A very heterogenous residual class, consisting of all nouns not in the proper or common 2 genders (Frowein 2011: 108).
- *Common 2*: Contains semantically marked nouns, including entities that are smallish or individuated from a greater mass, but also other semantic types; some examples are insect, birds, other smallish animals, plants and parts of plants, tools, loanwords, geographic locations, some meteorological phenomena, groups and sets, and ordinals (Frowein 2011: 105–107).

Teop and Siar thus clearly display the differences between a gender system and a simpler noun classification system according to the criteria of gender used in this paper.

⁷In this study, a word is only considered an indexing target if it has a functional load other than expressing gender and number. The reason for this is that otherwise languages such as Siar, which has a set of markers preceding only nouns, would be considered as having gender. Such a system would be difficult to separate from a system showing noun classification only on the noun itself, i.e., without indexation.

3.2 Number of genders

The second criteria concerns the number of genders in a language, based on Corbett (2013a). Each language is assigned the value two, three, four, or five or more genders (see Table 3 and Figure 3). The majority of the languages have only two genders, in all cases sex-based. Only one, Mian, has four genders. Of the remaining languages, three languages have three genders, whereas the remaining five languages have five or more genders, viz., Nalca (5), Motuna (6),⁸ Burmeso (9 [3+6]),⁹ Yimas (around 12), and Bukiyip (18 genders).

In contrast to the previous criterion, it is more difficult to identify subgroups based on values of the number of genders; e.g., the languages with three genders are very different from each other. Nevertheless, some of the languages have the following specific characteristics of

1. two genders where one is unmarked,
2. three genders consisting of masculine, feminine, and neuter, or
3. very large systems.

More than half of the languages with two genders have one which is unmarked, all of which are sex-based. Consequently, in these languages, either the feminine or the masculine gender is unmarked. An example of such a language is Maybrat (West Papuan), which has the conveniently named genders masculine and unmarked (i.e., non-masculine) (Dol 2007: 89). Thus, nouns denoting male humans (or in some cases other male animates) are masculine, whereas all others (including those denoting females) belong to the unmarked gender. This is shown in (4). In (4a) ‘old’ indexes ‘his father’, in (4b) ‘his mother’, and in (4c) ‘big’ indexes ‘house’.

(4) Maybrat (West Papuan) (Dol 2007: 90)

- a. *y-atia* *y-anes*
 3M-father 3M-old
 ‘His father is old.’/‘his old father’

⁸Onishi (1994) states that Motuna has six genders: masculine, feminine, diminutive, local, manner, and dual-paucal. However, the author does not elaborate on gender assignment, and I have been unable to satisfactorily conclude that the dual-paucal is truly a gender, which Onishi states. However, all form a complementary and mutually exclusive system, with separate identifiable markers and where a word may take only one gender.

⁹Burmeso has two gender systems, with three genders belonging to the first system and the other six belonging to the second system (see §5.2).

Table 3: Number of genders in the languages of the sample

Number of genders	No. of lgs.	%	Languages
Two	11	55%	Abau Kuot Manambu Maybrat Mende Oksapmin Skou Taiap Teop Walman Warapu
Three	3	15%	Ama Au Rotokas
Four	1	5%	Mian
Five or more	5	25%	Nalca (5) Motuna (6)
Burmeso (9 [3+6])			Yimas (~12) Bukiyip (18)
Total:	20	100%	

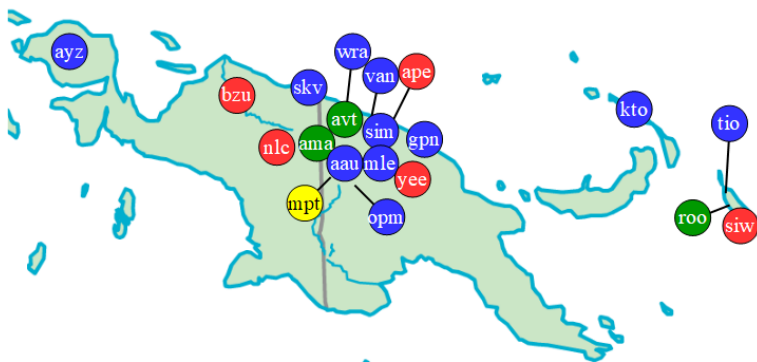


Figure 3: Number of genders. Colors indicate: two (blue), three (green), four (yellow), and five or more (red).

- b. *y-me* *m-anes*
 3M-mother 3U-old
 ‘His mother is old.’/‘his old mother’
- c. *amah* *m-api*
 house 3U-big
 ‘The house is big.’/‘the big house’

However, not all such languages use the masculine gender as the marked one. Languages where the masculine is marked are Warapu (Sko), Maybrat (West Papuan), Mende (Sepik), and Taiap (isolate), whereas the feminine is marked in Skou (Sko). It is also marked in Ama (Left May), which has three genders: masculine, feminine, and compound. However, the situation is more complex in Ama, both because there are three genders, and because the feminine also includes e.g., some non-female animates (Årsjö 1999: 68).

Except for Ama, which is mentioned above, the three-gendered systems belong to the second type, since all have masculine, feminine, and neuter. While this implies that inanimates are found only in the neuter gender, all languages assign some inanimates to the masculine and feminine genders as well, with or without sex-based motivation. For example, in Rotokas (North Bougainville), inanimate objects associated with male culture (such as hunting or warfare) and long, thin objects are masculine (see also §5.1), whereas most inanimates are assigned either to the feminine or to the neuter genders (Robinson 2011: 46–48).

The third and final type is languages with very large gender systems, viz., Bukiyip (Torricelli, Arapesh) and Yimas (Lower Sepik-Ramu, Lower Sepik). These are markedly different from all other languages in the sample. The most immediate difference is of course the vastly larger number of genders. Bukiyip has as many as 18 genders (Conrad & Wogiga 1991: 8–10), while Yimas has around a dozen genders, with Foley (1991: 119) distinguishing 10 and Phillips (1993: 175) as many as 16. All other languages in the sample have six genders or fewer. The Bukiyip genders and their indexing forms are shown in Table 11 in §3.5. The most important feature of these two gender systems is that both have semantic-formal assignment and gender marking on nouns. These two factors, which are uncommon in the sample, are undoubtedly related to the subsistence of their large systems.

Finally, a highly interesting case is Burmeso, which is the only language in the sample with two gender systems. The first system has three genders (masculine, feminine, and neuter), each with an additional subgender for inanimates, whereas the second system has six genders (I–VI). The exact nature of the gender systems and their interaction will be discussed further in §5.2.

3.3 Gender assignment

The third criterion concerns gender assignment and contains two values (see Table 4 and Figure 4), viz., semantic, or semantic and formal.

As can be seen in Figure 4, the majority of languages in the sample have semantic assignment. However, there are major differences between the various semantic systems as to their complexity. As mentioned in §3.1, Mende (Sepik) has an extremely simple system of gender assignment, where all nouns denoting human or sometimes animate males are masculine while all other nouns are feminine.

In Rotokas (North Bougainville), however, the situation is more complex. Rotokas has three genders: masculine, feminine, and neuter. Both the masculine and the feminine gender contain nouns denoting male and female referents respectively, but complexity arises for inanimates. The masculine gender contains many inanimate objects, which are often associated with male culture or which are long or thin (Robinson 2011: 46). The feminine gender also contains many inanimate objects, some of which are tools or related to water, but many which have no apparent semantic or formal criteria at all (Robinson 2011: 47). Finally, as expected, many inanimate nouns belong to the neuter gender (Robinson 2011: 48).

Thus, while a learner of Mende is easily able to guess the correct gender of any noun, a learner of Rotokas is hard-pressed to guess the correct gender of an inanimate object. Even if there are rules, many of these are probably not tacitly known. Furthermore, even if the rules for gender assignment can be explicitly stated, the system may still be opaque if the rules are not general or have numerous exceptions. One example is Manambu (described further below), where gender assignment sometimes carries the notion of large size, so that larger animals are masculine and smaller animals feminine. However, insects are masculine despite their small size.¹⁰

It is therefore possible to further split the systems with semantic assignment into two: transparent semantic vs. semantic + opaque (Figure 5), where opacity signals the inability of the researcher to find any clear semantic or formal criteria for gender assignment. It is possible that a language may have semantic + formal + opaque assignment, but no such system was clearly identified in the sample.

¹⁰It is of course possible to imagine various explanations why insects are not feminine, e.g., perhaps are they are not regarded as animals. However, this only further illustrates the reason for not regarding Manambu gender assignment as transparent. Although there certainly is a general pattern of size distinctions for gender assignment in Manambu, it is merely a pattern and not a rule.

Table 4: Systems of gender assignment in the sample

Gender assignment	No. of lgs.	%	Languages
Semantic	16	80%	Abau Ama Au Burmeso Manambu Maybrat Mende Mian Motuna Oksapmin Rotokas Skou Taiap Teop Walman Warapu
Semantic and formal	4	20%	Bukiyip Kuot Nalca Yimas
Total:	20	100%	

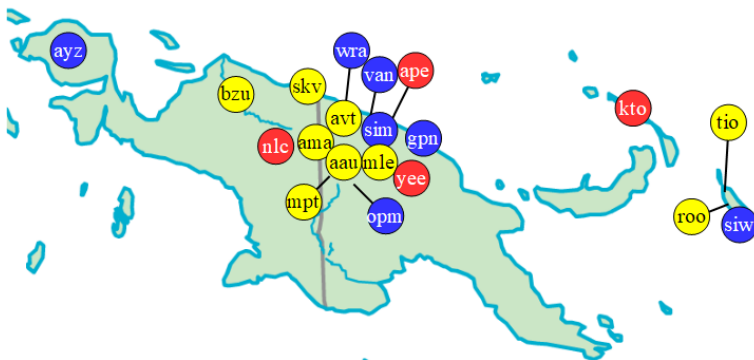


Figure 4: Systems of gender assignment. Colors indicate transparent semantic (blue), semantic and formal (red), and semantic and opaque (yellow).

Table 5: Types of semantic assignment of gender assignment in the sample

Gender assignment	No. of lgs.	%	Languages
Transparent semantic	8	40%	Au Maybrat Mende Motuna Oksapmin Taiap Walman Warapu
Semantic and formal	4	20%	Bukiyip Kuot Nalca Yimas
Semantic and opaque	8	40	Abau Ama Burmeseo Manambu Mian Rotokas Skou Teop
Total:	20	100%	

This thus gives rise to three types of gender assignment: transparent semantic, semantic and formal, and semantic and opaque (Table 5).¹¹

Since all languages have some form of semantic assignment, the most basic system is necessarily one where all nouns are assigned their genders based on few and clear semantic criteria. Mende has already been mentioned above and is exemplified in (1) in §3.1. However, semantic systems can be more complex while still retaining transparent semantic criteria, e.g., via a larger number of

¹¹It is not explicitly stated, but Au (Scorza 1985) appears to have a simple semantic system where nouns denoting human males are masculine, human females are feminine, and the rest are neuter. However, this is complicated somewhat by masculine and neuter agreement being homophonous in the singular.

Table 6: Gender indexation forms in Motuna (adapted from Onishi 1994: 70)

	Demonstrative	Article	Adjective/ classifier/ kinship term endings	Possessor/ local NP endings	Verbal endings
Masculine	<i>ong</i>	<i>hoo/shoo</i>	<i>-ng</i>	<i>-ng</i>	<i>-ng</i>
Feminine	<i>ana</i>	<i>tii</i>	<i>-na</i>	<i>-na</i>	<i>-na</i>
Diminutive	<i>oi</i>	<i>tii</i>	<i>-ni</i>	<i>-ni</i>	<i>-ni</i>
Local	<i>owo</i>	<i>ti</i>	–	<i>-no</i>	<i>-no</i>
Manner	–	<i>tiwo</i>	–	–	<i>-nowo</i>
Dual-paucal	<i>oi</i>	<i>tii</i>	–	<i>-ni</i>	<i>-(n)i</i>

gender distinctions. One example is Motuna (South Bougainville), which has six genders: masculine, feminine, diminutive, local, manner, and dual-paucal (Onishi 1994: 68–69). The forms of gender indexation in Motuna are shown in Table 6.

In Motuna, animate referents are assigned gender based on their natural gender; this also includes nouns associated with mythical characters such as *raa* ‘the sun’ and *hingjoo* ‘the moon’, which are assigned the gender of their character (Onishi 1994: 70). Animals are most commonly masculine, but can be assigned the feminine gender if emphasizing that the referent is a female. On the other hand, the majority of inanimate nouns are masculine, but can be treated as diminutive when emphasis is placed on their small size. This includes nouns which signify smallish things, e.g., *irihwa* ‘finger’ or *kaa* ‘young tree’ (Onishi 1994: 71). Nouns with spatial or temporal meaning are inherently local gender. The manner gender contains only two nouns. Finally, the dual-paucal gender can be used also when the speaker does not want to specify the gender of a sentential topic (Onishi 1994: 71).

In contrast to the transparent semantic criteria in Mende and Motuna mentioned above, many languages have much more complex systems. If gender assignment is neither semantically transparent nor apparently formal, it is classified as being opaque, with Rotokas having already been mentioned at the beginning of this section. Another example of such a language is Manambu (Ndu), which exhibits the fairly common feature of gender assignment based on size and shape (see §5.1). Manambu has two genders, masculine and feminine, and in general gender assignment appears to follow semantic criteria. However, these are far from transparent:

1. Humans are assigned gender based on their sex, except nouns denoting small children, which can be assigned gender based on size (Aikhenvald 2008: 116–117).
2. Higher animates are assigned gender based on their size and natural gender: larger animals are masculine, whereas smaller animals are feminine, except when the sex of the referent is known. Furthermore, nouns denoting young animals are feminine (Aikhenvald 2008: 117).
3. Lower animates such as insects are masculine. However, if the lower animate has a certain shape, it is assigned gender based on shape; thus, *gwa:s* ‘turtle’ is feminine, since it is round, while *mu* ‘crocodile’ is masculine since it is long (Aikhenvald 2008: 117).
4. Inanimates are assigned gender based on their size and shape: long and/or large objects are masculine, whereas small and/or round objects are feminine. Thus, *vø:y* ‘spear’ is masculine, since it is large, but it is feminine if referring to small spears or shotguns (Aikhenvald 2008: 117).
5. Natural phenomena are assigned gender based on whether they are complete or not: if they are uncomplete or if completeness is not emphasized, they are feminine; otherwise, they are masculine (Aikhenvald 2008: 118). Thus, *ga:n* ‘night’ is feminine, unless it implies complete darkness, as is *gəl* ‘cloud’ if there are only a few but masculine if they cover the whole sky. Other natural phenomena are assigned gender based on their shape: e.g., ‘rainbow’ is masculine since it is long, whereas ‘sun’ is feminine since it is round; unless it is really hot, in which case it becomes masculine to reflect its intensity (Aikhenvald 2008: 119).
6. Mass nouns and nouns covering ‘extent’ follow complex patterns. In general, they are assigned gender based on extremity, so that smaller quantities are feminine, whereas larger quantities are masculine (Aikhenvald 2008: 119–120). However, nouns denoting manner, language or voice, or time span are feminine; except for *nabi* ‘year’, which is masculine because it is very long (Aikhenvald 2008: 119).

There are in fact further assignment rules, but the important point is that rules of gender assignment are not semantically transparent. It is especially important to note that it is difficult to ascertain whether there are any rules or merely patterns. That is not to belittle the observations or to claim that the researcher, in

this case Aikhenvald, has done anything wrong. Instead it illustrates the difference to transparent semantic systems, where all gender assignment rules are easily identifiable and apply to all nouns, whereas in opaque systems there are certainly some patterns that can be identified, but exceptions abound.

While it is easy to become amused by the seemingly arbitrary gender assignment rules, one important thing should be noted. In a language such as Manambu, gender has a very important pragmatic function, since it is available as a tool for the speaker to use when emphasizing certain features, not least in jokes:

As a joke, a man can be referred to with feminine gender, and a woman with masculine gender, depending on their ‘shape’ and ‘size’. A smallish fat woman-like man can be treated as feminine, e.g. *numa du* (big.FSG man) ‘fat round man’. And a largish woman can be ironically referred to with a masculine gender form, e.g. *kə-də numa-də ta:kw* (DEM.PROX-M.SG big-M.SG woman) ‘this (unusually) big woman’. (Aikhenvald 2008: 121)

The final category consists of the four languages with both semantic and formal assignment: Nalca is skewed towards semantic assignment, Kuot applies semantic and formal assignment roughly equally, and Bukiyip and Yimas favor formal assignment. For example, among the five genders in Nalca (see §3.1 above), only the neuter is formal, but very much so since it contains only (but all) nouns of the phonological structure (C)V (Wälchli 2018). In comparison, only three of the 18 genders in Bukiyip (see Table 10) are semantic (masculine, feminine, and mixed or unspecified), whereas all others are morphological (Conrad & Wogiga 1991: 8). The same is true for Yimas, where three genders are semantic, while the others are based on phonological criteria (Foley 1991: 119).

3.4 Number of gender-indexing targets

Following Di Garbo (2014: 66), the number of gender-indexing targets is given the value of one, two, three, or four or more. The results are shown in Table 7 and Figure 5, while each type of indexing target is shown in Table 8. The identification and counting of gender-indexing targets was based on the general guidelines used by Di Garbo (2014: 66), where the following general categories were used to identify targets: pronouns, adjectives, demonstratives, verbs, numerals, copulas, complementizers, and adpositions. However, no detailed analysis has been made of different subtypes of these groupings, so the results should be understood only as showing general patterns.

Table 7: Number of gender-indexing targets in the languages in the sample

Number of gender-indexing targets	No. of lgs.	%	Languages
One	4	20%	Ama Mende Nalca Oksapmin
Two	4	20%	Warapu Burmeso Skou Teop
Three	2	10%	Au Taiap
Four or more	10	50%	Abau Bukiyip Kuot Maybrat Manambu Mian Motuna Rotokas Walman Yimas
Total:	20	100%	

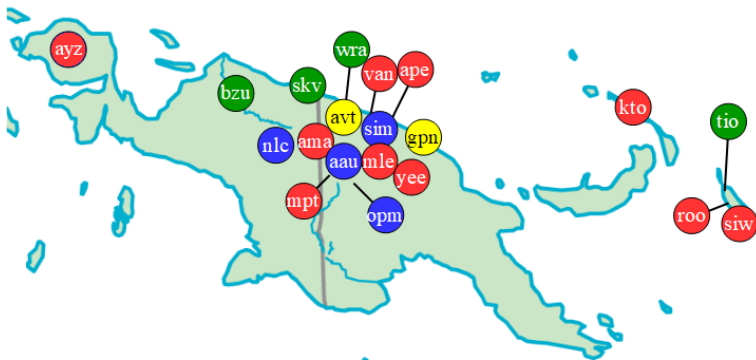


Figure 5: Number of gender-indexing targets. Colors indicate: one (blue), two (green), three (yellow), and four or more (red).

Table 8: Distribution of gender-indexing targets in the languages of the sample

Language	Pronouns	Verbs	Demonstratives	Adjectives	Numerals	Prepositions	Case hosts
Abau	x		x		x		x
Ama		x					
Au	x	x		x			
Bukiyip	x	x	x	x			
Burmeso		x		x			
Kuot	x	x	x	x		x	
Manambu	x	x	x	x			
Maybrat	x	x	x	x			
Mende	x						
Mian	x	x	x	x			
Motuna	x	x	x	x	x		
Nalca							x
Oksapmin	x						
Rotokas	x	x	x	x			
Skou	x	x					
Taiap	x	x	x				
Teop				x	x		
Walman	x	x	x	x	x		
Warapu	x	x					
Yimas	x	x	x	x	x		

As shown in Table 7 and 8, more than half of the languages in the sample have more than four gender-indexing targets.¹² There are also some general patterns to be found in Table 8:¹³

¹²In Burmeso, adjectives are targets in the first gender system whereas verbs are targets in the second system.

¹³'Pronoun' here denotes a word with general pronominal uses (i.e., as constituting an individual noun phrase), whether it belongs to the language-specific category of pronouns or demonstratives. In comparison, 'demonstrative' only refers to attributive forms. Kuot has no independent third person personal pronouns (Eva Lindström, p.c.). However, demonstratives are used with pronominal functions (see also footnote 16 below). 'True pronouns' in Yimas exist only in the first and second person without gender (Foley 1991: 111). The third person is instead expressed with a set of deictics, which show gender and are most commonly used as free pronouns in narrative discourse (Foley 1991: 113). Therefore these forms are considered pronouns for comparative purposes.

- If a language has four gender-indexing targets, they always include pronouns and demonstratives, and almost all such languages include verbs and adjectives, with Abau (Sepik) being the exception.
- If a language has three gender-indexing targets, they include verbs and pronouns.
- If a language has two gender-indexing targets, they mostly include verbs and to a lesser extent pronouns.
- If a language has only one gender-indexing target, the target could be anything (e.g., verbs, pronouns, or even case marker hosts).

Based on the likelihood of a gender-indexing target appearing in a language, it is possible to arrange the distributional tendencies into tentative hierarchies, where the leftmost target is the most typical target while the rightmost target is the least common one. If one target is present in a language, every target to the left is present as well. That is, if a language has only one target, it is likely to be the leftmost one, whereas if a language has five it should include every part of the hierarchy. There are three tendencies:

- pronouns > verbs > demonstratives > adjectives > numerals (holds for 14 out of 20 languages)
- verbs > adjectives > pronouns (3/20)
- other (3/20)

It is also interesting to note that among the ten languages with four or more indexing targets, all except Abau follow the first hierarchy. There is therefore an additional pattern, whereby a gender system with many indexing targets is expected to follow the first hierarchy. In comparison, four of the six languages of the other two categories have two gender-indexing targets or less, with Au having three targets and Abau four.

The languages not describable in terms of the first and second hierarchies are all very different and require some explanation. One example is Nalca (TNG, Mek), which only shows gender on markers functioning as case marker hosts following the NP. These carry the meaning of gender, case, and demonstrative, of which at least the first two mostly occur together. Some of the most common forms are shown in Table 9. Examples were given in (3) in §3.1 above, the first of which is repeated in (5).

Table 9: Some of the most of most frequent forms of case marker hosts words in Nalca.

Case	masc. <i>be-</i>	fem. <i>ge-</i>	neuter <i>ne-</i>	default noun <i>e-</i>	non-noun <i>a-</i>
Topic	<i>bera</i>	<i>gera</i>	<i>nera</i>	<i>era</i>	<i>ara</i>
Topic dem.	<i>benera</i>	<i>genera</i>	<i>nenera</i>	<i>enera</i>	<i>anara/anera</i>
Absolutive	<i>bek</i>	<i>gek</i>	<i>nek</i>	<i>ek</i>	<i>ak</i>
Abs. dem.	<i>benyek</i>	<i>genyek</i>	<i>nenyek</i>	<i>enyek</i>	<i>anyek</i>
Gen./ergative	<i>bedya(')</i>	<i>gedya(')</i>	<i>nedya(')</i>	<i>edy(')</i>	<i>adya(')</i>
Gen./erg. dem.	<i>benedya</i>	<i>genedy</i>	<i>nenedya</i>	<i>enedya</i>	<i>anadya</i>
Comitative	<i>beb</i>	<i>geb</i>	<i>neb</i>	<i>eb</i>	<i>ab</i>
Com. dem.	<i>benyeb</i>	<i>genyeb</i>	<i>nenyeb</i>	<i>enyeb</i>	<i>anyeb</i>
Equative	<i>beneso(')</i>	<i>geneso(')</i>	<i>neneso(')</i>	<i>eneso(')</i>	<i>anaso(')</i>
Benefactive	<i>bemba</i>	<i>gemba</i>	<i>nemba</i>	<i>emba</i>	<i>amba</i>

- (5) Nalca (TNG, Mek) (own example; repeated from 3a)

alja si ne-ra Zakheus be-k u-lum-ok

3SG.GEN name N-TOP Z. M-ABS be-IPFV-PST.3SG

'a man called by name Zacchaeus' (Lk 19:2)

lit. 'his name was Zacchaeus'

Another interesting example is Teop (Austronesian, Oceanic). In Teop, gender is visible on a set of articles preceding nouns, adjectives, and numerals. Two examples of markers preceding adjectives and numerals, respectively, are shown in (6).

- (6) Teop (Austronesian, Oceanic) (Mosel & Spriggs 2000: 330, 328)

a. *a inu a beera*

ART.I.SG house ART.I.SG big

'the big house'

b. *o buaku o hoi*

ART.II.SG two ART.II.SG basket

'the two baskets'

However, since these articles do not carry any other functional load, they do not satisfy the criterion that an indexing target must express something other

than gender and number. Instead, Teop is analyzed as having two targets, viz., adjectives and numerals, which form a unit with the preceding article. On the other hand, the articles preceding nouns are analyzed as overt gender marking (see §3.5).

3.5 Occurrence of gender marking on nouns

The final criterion concerns the occurrence of gender marking on nouns (see Table 10 and Figure 6), following Di Garbo (2014: 69). Gender marking on nouns is of course not considered indexation, but it is a common feature e.g. in African languages and most certainly a characteristic trait of many gender systems.

Most languages of the sample (17 of 20) do not have overt gender marking, with Bukiyip, Teop, and Yimas being the only exceptions. In both Bukiyip and Yimas, gender is shown on nouns via suffixes; the Bukiyip noun suffixes are given in Table 11. Both languages are unusual in the sample by their having many noun classes (18 in Bukiyip, around a dozen in Yimas), many gender-indexing targets (both four or more), and semantic-formal assignment. In fact, these features are probably tightly interconnected with the overtness of gender. The combination of many genders and morphological gender assignment appears more common when noun classes are overtly distinct.

On the other hand, Teop (Austronesian, Oceanic) has a very different kind of marking. As mentioned above, Teop has a set of articles that obligatorily precede nouns, adjectives, and numerals. Thus, the latter two are indexation, while the articles preceding nouns are considered overt marking. The forms of the markers are shown in Table 12.

Note that Teop has two genders, one of which is divided into two subgenders. The reason for them not being separate gender is that the distinction is kept only on the articles preceding nouns, and never on the articles preceding adjectives and numerals. Thus, since overt gender marking cannot constitute gender as it is not indexation, Teop only has two genders.

This is very similar to the related Austronesian language Siar (not in the sample), which also has articles preceding nouns (Frowein 2011). However, the Siar articles are not used in other contexts, so the absence of indexation renders Siar genderless. Nevertheless, a pronoun can be placed before e.g., an adjective, which is similar to the use of the Teop article. However, pronouns in Siar do not show any gender distinctions. The difference between Teop and Siar in this regard is shown in (7) and (8), respectively.

Table 10: Occurrence of gender marking on nouns in the sample

Gender marking on nouns	No. of lgs.	%	Languages
Yes	3	15%	Bukiyip Teop Yimas
No	17	85%	Abau Ama Au Burmeso Kuot Manambu Maybrat Mende Mian Motuna Nalca Oksapmin Rotokas Skou Taiap Walman Warapu
Total:	20	100%	

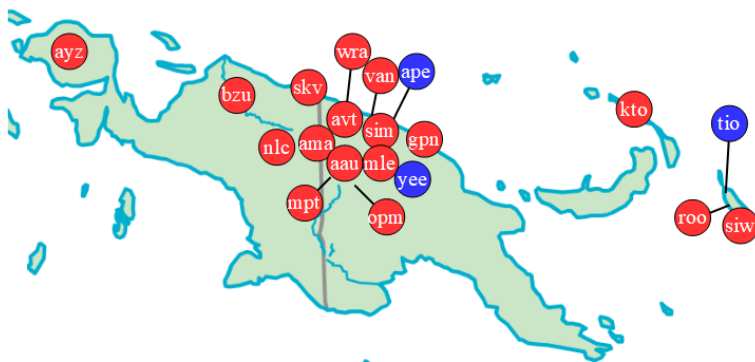


Figure 6: Occurrence of gender marking on nouns. Colors indicate: yes (blue), and no (red).

Table 11: Bukiyip noun classes and noun class suffixes (adapted from Conrad & Wogiga 1991: 10)

Class	Glossing	Example		Noun suffix	
		singular	plural	singular	plural
1	betel nut	<i>búb</i>	<i>búbús</i>	<i>-b/n</i>	<i>-bús</i>
2	village	<i>wabél</i>	<i>walúb</i>	<i>-bél</i>	<i>-lúb</i>
3	feces	<i>dewag</i>	<i>dewas</i>	<i>-g/-gú</i>	<i>-s/-as</i>
4	woman	<i>élmatok</i>	<i>élmagou</i>	<i>-k</i>	<i>-ou/-eb</i>
5	banana	<i>apam</i>	<i>apas</i>	<i>-m/-bal</i>	<i>-s/-ipi/-bal</i>
6	moon	<i>aun</i>	<i>aub</i>	<i>-n/-nú</i>	<i>-b</i>
7	man	<i>éلمان</i>	<i>élmom</i>	<i>-n/-nú</i>	<i>-m</i>
8	child	<i>batawiny</i>	<i>batawich</i>	<i>-ny/-l</i>	<i>-ch/-has</i>
9	leaf	<i>chuwup</i>	<i>chuwus</i>	<i>-p</i>	<i>-s</i>
10	mosquito	<i>aul</i>	<i>auguh</i>	<i>-l/-ny</i>	<i>-guh</i>
11	dog	<i>nobat</i>	<i>nobagw</i>	<i>-t/-tú</i>	
12	sago leaves	<i>lohuhw</i>	<i>lohulúh</i>	<i>-hw</i>	
13	road	<i>yah</i>	<i>yeh/yegwih</i>	<i>-V₁h</i>	<i>-V₂h</i>
14	box		<i>kes</i>	<i>-s</i>	<i>-s</i>
15	small pig		<i>buligún</i>	<i>-gún</i>	<i>-gún</i>
16	garden		<i>yawihas</i>	<i>-has</i>	<i>-has</i>
17	personal names			-	-
18	place names			-	<i>-gún</i>

Table 12: Gender marking in Teop on articles preceding nouns (Mosel & Spriggs 2000: 322)

	head (SG)	head (PL)	target (SG)	target (PL)
Gender I-E	<i>e</i>	<i>o</i>	<i>a</i>	<i>o</i>
Gender I-A	<i>a</i>	<i>o</i>	<i>a</i>	<i>o</i>
Gender II	<i>o</i>	<i>a</i>	<i>o</i>	<i>a</i>

- (7) Teop (Austronesian, Oceanic) (Mosel & Spriggs 2000: 326)

a inu a ruta

ART.I house ART.I small

'the small house / the house is small'

- (8) Siar (Austronesian, Oceanic) (Frowein 2011: 206)

Ép rumai i mètèk.

ART.CO1 house 3SG new

'The house is new.'

Finally, some languages have overt marking in some cases or at least something resembling it. One example is Kuot (isolate), where some nouns belong to various declension classes (as defined by noun endings), which in turn belong to a certain gender (Lindström 2002: 176). Another example is Rotokas (North Bougainville), which has noun suffixes expressing both number and gender (Robinson 2011: 41). However, these are not always present: in (9a), *aveke* 'stone' has a feminine singular suffix, but in (9b) it remains unmarked.

- (9) Rotokas (Robinson 2011: 42)

a. *riako-va aveke-va peka-e-vo uva rakoru*
 woman-SG.F stone-SG.F turn.over-3SG.F-IPST and snake
keke-e-vo uva kea-o-e oisio uo-va
 look.at-3SG.F-IPST and mistake.for-3SG.F-IPST as eel-SG.F

'The woman turned over to the stone and saw a snake but mistook it for an eel.'

b. *kaveakapie-vira aveke tovo-i-vo uva kove-o-e*
 insecure-ADV stone place-3PL-IPST and fall-3SG.F-IPST

'They placed the stone insecurely and it fell down.'

Since gender marking on nouns is not always present, Rotokas cannot be said to have obligatory overt marking.

4 Typological comparison

This section compares the results of this study with previous research on Africa and the world as a whole. The data on Africa is from Di Garbo (2014), which used the same five criteria of this study to investigate a variety sample of 100

languages. The data on the world as a whole is based on the three *WALS* chapters on gender by Corbett (2013a,b,c). These three *WALS* chapters correspond to the first three classification criteria of this study. Unfortunately, the remaining two have no corresponding *WALS* data, rendering the final two criteria comparable only for New Guinea and Africa.

Some care had to be taken when comparing the results, since the samples are of different types. Whereas this study employs a variety sample, Corbett uses a proportional sample (of 257 languages) (see §2). Di Garbo also uses a variety sample (of 100 languages) although with some differences, most importantly the inclusion of 16 non-gendered languages as well as being intentionally genealogically skewed. To make the data comparable, languages without gender have been omitted from Corbett’s and Di Garbo’s samples in this section, leaving 112 languages for Corbett and 84 for Di Garbo.

Classification criterion 1: Sex-based and non-sex-based gender systems (§3.1). In the sample of this study, sex-based systems are by far more common, with only Teop (Austronesian, Oceanic) having a non sex-based system. In comparison, in Di Garbo’s (2014: 63) sample, 48 languages (57%) had sex-based gender systems and 36 languages (43%) non-sex-based gender systems. In Corbett’s (2013c) sample, 84 languages (75%) have sex-based systems and 28 (25%) non-sex-based. A comparison of the percentage distributions is shown in Figure 7.

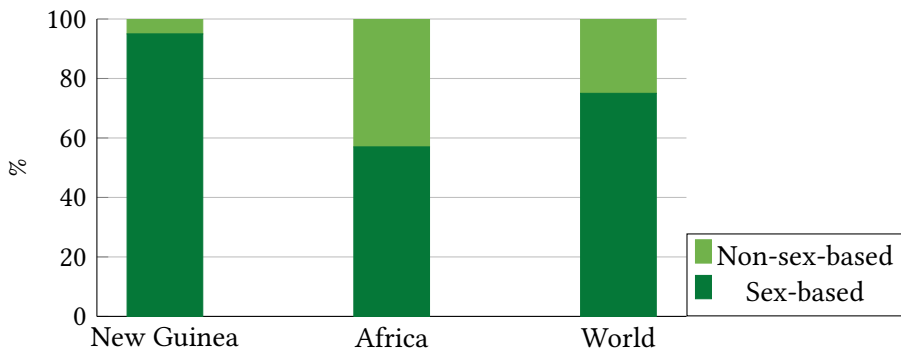


Figure 7: Sex-based and non-sex-based systems in New Guinea, Africa and the world

Sex-based systems are more common in all samples, although even more so in the sample from New Guinea. According to Corbett’s (2013b) data, non-sex-gender systems are actually uncommon in most regions, being found primarily in the Niger-Congo languages of Africa, which account for the vast majority of non-sex-based systems in the sample. More specifically for Africa, in most cases

only one system occurs in an entire family: this is true e.g., for the Bantu, Mel, and North-Central Atlantic families, which together account for 33 of the 36 non-sex-based gender systems in Di Garbo's sample. It is therefore not surprising that the non-sex-gender systems are relatively common, since 31% of the gendered languages (26/84) in Di Garbo's sample are Bantu languages.

An interesting discussion about the differences between sex-based and non-sex-based systems is presented by Luraghi (2011), who argues that they have different diachronic origins. Non-sex-based systems originate from the grammaticalization of classifiers, whereas sex-based systems originate from agreement with groups of nouns that show different morphosyntactic behavior. Since sex-based systems are more common, it is thus not surprising that they are the primary ones in New Guinea. It is likely not a coincidence that the only non-sex-based gender system of the sample is found in an Austronesian language, a family remarkably devoid of gender but abounding with classifiers.

Classification criterion 2: Number of genders (§3.2). In the sample of this study, eleven languages (55%) have only two genders, three languages (15%) three genders, one language (Mian; TNG, Ok-Oksapmin) (5%) four genders, and the final five languages (25%) five genders or more. In Di Garbo's (2014: 65) sample, 42 languages (50%) have only two genders, seven languages (8%) three genders, one (Jul'hoan; Kx'a) (1%) four genders, and the final 34 languages (40%) five genders or more. In Corbett's (2013a) sample, 50 languages (45%) have only two genders, 26 languages (23%) three genders, 12 languages (11%) four genders, and the final 24 (21%) five genders or more. A comparison between the percentage distributions is shown in Figure 8.

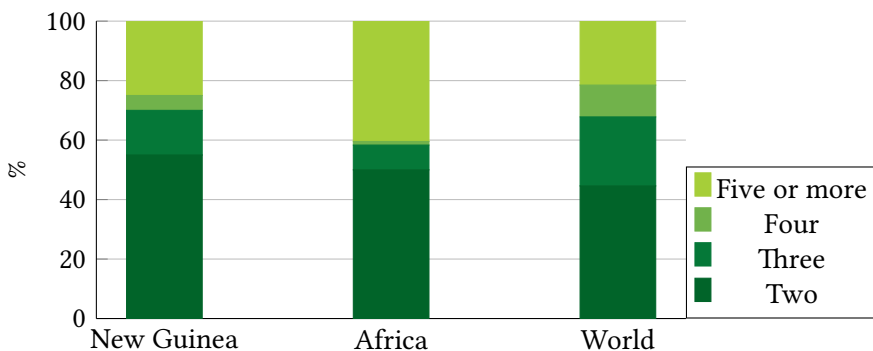


Figure 8: Number of genders in New Guinea, Africa and the world

The distributions in all three samples are similar to a large extent, with two-gender systems being present in around half of the languages. In Africa, large

systems are much more common than in New Guinea or the world as a whole. However, this may once again be because of the sample. As mentioned before, 31% of the languages present in Di Garbo's (2014) sample are Bantu languages, all of which have very large gender systems. In the sample of this study, however, the rather large Torricelli and Lower Sepik-Ramu families, which according to Foley (2000: 372) have large systems, are represented only by Bukiyip and Yimas respectively (i.e., 10% of the sample). It is thus very probable that the similarities between the distribution numbers of genders in New Guinea and Africa actually are greater than indicated here.

Classification criterion 3: Gender assignment (§3.3). This criterion is less straightforward to compare, since this study uses three values (transparent semantic, semantic and formal, and opaque), whereas Di Garbo (2014) and Corbett (2013c) use only two (semantic, and semantic and formal). For the purpose of this comparison, the languages of the purely semantic and semantic + opaque groups are added somewhat tentatively into a semantic group. While this may appear misleading, it is important to note that the researchers investigating these languages considered them as having semantic gender assignment and no traces of formal assignment rules have been identified by the present author. Indeed, both languages exemplified in Corbett (2013c), Bininj Gun-Wok (Gunwinyic; northern Australia) and Russian, would be considered opaque using the values of this study.

In the sample of this study, 16 languages (80%) exhibit semantic gender assignment, whereas only four languages (20%) show semantic and formal assignment. In comparison, in Di Garbo's (2014: 67) sample, six languages (7%) have semantic assignment, 76 languages (90%) semantic and formal assignment, while the remaining two languages (2%) have unknown assignment (disregarded in Figure 9). In Corbett's (2013c) sample, 53 languages (47%) exhibit semantic assignment, and 59 languages (53%) semantic and formal assignment. A comparison between the percentage distributions is shown in Figure 9.

As can be clearly seen in Figure 9, in New Guinea, semantic assignment is by far more common, while it is by far the most uncommon form of gender assignment in Africa, including of course the Bantu languages. In the world as a whole, the ratio is more or less equal. Thus, New Guinea and Africa both represent two extremes while the world as a whole is more average. However, according to Corbett (2013c), semantic and formal assignment is mostly found in the Indo-European, Afro-Asiatic, and Niger-Congo families, which together represent a large amount of the languages of the world.

It is not surprising that semantic and formal assignment appears more often in Di Garbo's and Corbett's samples than in New Guinea, since no family is rep-

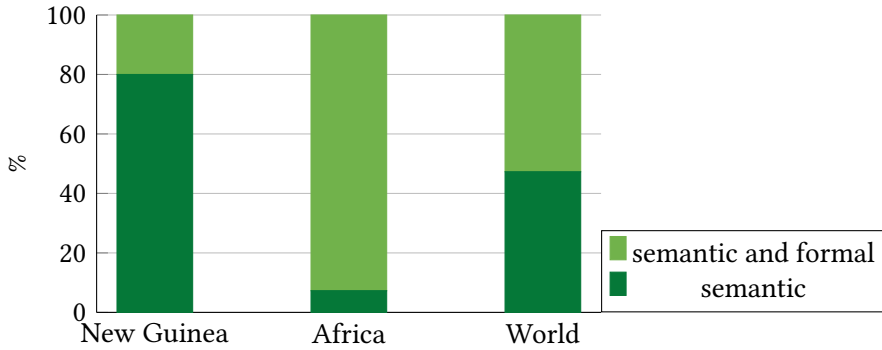


Figure 9: Gender assignment in New Guinea, Africa and the world

resented with more than three members in this study. Bukiyip (Torricelli, Arapesh) and Yimas (Lower Sepik-Ramu, Lower Sepik) both belong to rather large families, so it is possible that a proportional sample would show that semantic and formal assignment indeed is more common than it appears here. Nevertheless, it is interesting that it occurs in few families, both in New Guinea and the world, which Corbett (2013c) relates to these systems necessarily being older. As argued by Luraghi (2011), this implies that most gender systems of Africa are old. Exclusive semantic assignment is however found in both older and younger systems, and thus it cannot be claimed that the predominance of semantic assignment indicates that those systems are young. Interestingly, semantic and formal assignment is found in Nalca (TNG, Mek), which has a very young gender system (Wälchli 2018).

Classification criterion 4: Number of gender-indexing targets (§3.4). In the sample of this study, four languages (20%) have only one gender-indexing target, another four languages (20%) two targets, two languages (10%) three targets, and the final ten languages (50%) four or more targets. In Di Garbo's (2014: 68) sample, five languages (6%) have only one gender-indexing target, 16 languages (19%) two targets, 28 languages (33%) three targets, and finally 33 languages (39%) four targets or more. No data was available for the remaining two languages. A comparison of the percentual distributions is shown in Figure 10.

Four or more gender-indexing targets is the most common number in both samples, accounting for slightly less than half of all languages. Furthermore, systems of only two targets account for around a fifth of the languages in both samples. As for the two remaining values, the relationships are the opposite: systems of three targets are common in Africa but rare in New Guinea, whereas one-target systems occur in a fifth of the New Guinean languages but only 6% of

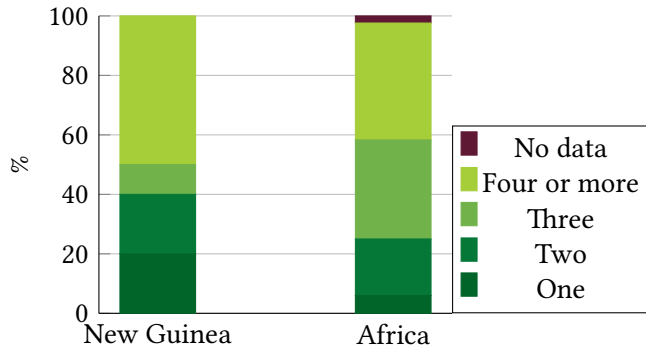


Figure 10: Number of gender-indexing targets in New Guinea vs. Africa

the African languages. However, once again it is probable that these differences are largely due to larger families with more mature gender systems being better represented in Di Garbo’s (2014) sample, while languages from smaller families with possibly less mature gender systems constitute a large part of the sample of this study.

Classification criterion 5: Occurrence of gender marking on nouns (§3.5). In the sample of this study, three languages (15%) have overt gender marking, whereas the remaining 17 (85%) do not. In Di Garbo’s sample, 69 languages (82%) have overt gender marking and 15 (18%) do not. A comparison between the percentage distributions is shown in Figure 11.

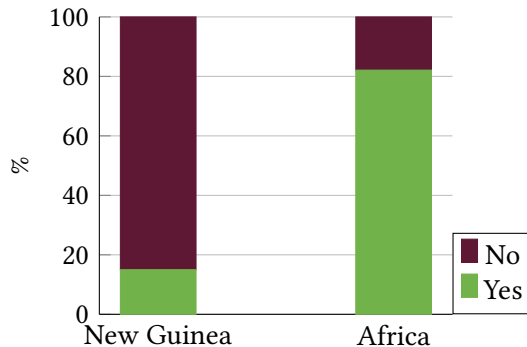


Figure 11: Occurrence of gender marking on nouns in New Guinea vs. Africa

As the figure shows, there is a major disparity between the presence of gender marking on nouns in New Guinea and Africa. In New Guinea, overt gender

marking is rare and occurs in only three languages in the sample, whereas in Africa it occurs in the vast majority of languages.

There is an interesting correlation between this distribution and the one of gender-assignment shown in Figure 4. Thus, semantic assignment without gender marking on nouns is the norm in New Guinea, whereas semantic and formal assignment with gender marking on nouns is the norm in Africa. This correlation is hardly coincidental. A gender system with assignment based on formal criteria benefits greatly from overt gender. In an exclusively semantic system however, obligatory overt gender has no function in gender assignment.

To summarize, it can be confidently stated that the gender systems of New Guinea and Africa are very different. Much of this depends on the hegemony of Bantu languages in Africa (as represented by Di Garbo's sample), which makes the distribution of values much less diverse than in the sample of this study. Nevertheless, the most important differences are (1) the prevalence of semantic and formal assignment and overt gender in Africa, while the exact opposite is true in New Guinea, and (2) as the observation that non-sex-based genders are much more common in Africa. This clearly shows that the two regions have gender systems of very different types. Reasons for this definitely include sample size and technique, but it also suggests that the gender systems of New Guinea may have different diachronic origins.

As for New Guinea in relation to the world as a whole, the above data and figures show that the distribution of values of the three classification criteria is rather similar in New Guinea and the world. In fact, most of the smaller differences can probably be accounted for by sample size. Nevertheless, the main conclusion is that the languages of New Guinea seem to be remarkably representative of the languages of the world, but another study with a proportional sample from New Guinea would elucidate this further.

5 Special characteristics

In this section, four characteristics of the gender systems of New Guinea are highlighted, two of which reflect characteristics mentioned by Foley (2000), viz., gender assignment based on size and shape, and the occurrence of two separate gender systems. The other two, viz., no gender distinctions in pronouns and gender marking on verbs, pertain to two typologically uncommon characteristics. Although these do not occur in all languages of the sample, they are found in geographically and genealogically distant languages and are all characteristic of the region.

5.1 Size and shape

Four languages in the sample (20%) share the property of having size and shape as important criteria for gender assignment. While gender assignment in many languages may carry some form of size- or shape-based rules, the rules discussed here all share the feature that nouns denoting tall, long, or thin objects are considered masculine, whereas nouns denoting short, thick, or round objects are feminine. In addition, they are all core assignment criteria. The languages in the sample exhibiting this feature are: Abau (Sepik), Manambu (Ndu), Skou (Sko), and Taiap (isolate).¹⁴ Their rules based on shape and size are shown in Table 13.

Table 13: Gender assignment rules based on size and shape in the sample

Language	Masculine	Feminine
Abau	- large - three-dimensional - long and extended	- small - two-dimensional (i.e., very thin) - round with little height
Manambu	- large - long	- small - round
Skou	- large - long, thin	- small - round, squat
Taiap	- large - long, high, thin	- small - round, stocky

In these four languages, size and shape are important criteria for gender assignment. One example mentioned in §3.1 above is Abau, which has two genders: masculine and feminine. Humans, along with spirits and domesticated animals, are assigned gender based on their sex, whereas abstract entities are feminine (Lock 2011: 47). However, animals and concrete inanimate objects are assigned their gender based on shape and size. Large, three-dimensional, and/or long and extended objects are masculine, while small, two-dimensional (i.e., very thin), and/or round objects with little height are feminine (Lock 2011: 47). Thus, *su* ‘coconut’ (three-dimensional), *now* ‘tree’ (long), and *hu* ‘water’ (liquid) are masculine, while *iha* ‘hand’ (flat) and *hne* ‘bird’s nest’ (round with little height) are feminine (Lock 2011: 48–50).

¹⁴‘Non-feminine’ in Skou.

It is important to distinguish systems such as the ones above from diminutives. In some languages, diminutives constitute separate genders, such as in Motuna (South Bougainville) (Onishi 1994: 68–69). However, the four languages above show the peculiar characteristics that (1) size and shape function as assignment criteria for the masculine and feminine genders, and (2) they constitute opposing criteria, and (3) they show the same pattern of large/long vs. small/round.

In the sample, size and shape constitute important gender assignment criteria in only these four languages, but similar systems are present in other languages. Rotokas exhibits some similarities with these gender assignment rules in two ways. Firstly, one class of nouns belonging to the masculine gender consists of inanimate objects associated with male culture, but also includes long or thin objects. However, no comparable feminine gender assignment rule has been found. Furthermore, this appears to be only a peripheral gender assignment rule. Secondly, Rotokas has a set of classifiers based on shape and size, classifying nouns based on their being round, narrow, or long. While this is not related to any masculine-feminine opposition, it nonetheless bears some resemblance to these systems.

Another interesting example is Mian (TNG, Ok). Mian has four genders, viz., male, female, neuter 1, and neuter 2, none of which has gender assignment rules resembling those of size and shape (Fedden 2011: 171–176). However, around 50 verbs require the use of a classificatory prefix, which has two functions: firstly, it encodes the direct object of transitive verbs and the subject of intransitive verbs, and secondly it classifies it according to characteristics of the referent, viz., sex, shape, and function (Fedden 2011: 185). This classification system, which is separate from the gender system, includes classes for e.g., long or flat objects, and in some cases overlaps with the gender system (e.g., some neuter 1 nouns are included in the masculine class). A table illustrating the overlap between the two systems is shown in Table 14.

Assigning genders based on shape and size is not very common in the languages of the world (Aikhenvald 2000: chap. 11). Outside of New Guinea, it occurs e.g., in some Afroasiatic languages, such as Oromo and Amharic, Central Khoisan, and Cantabrian Spanish (Aikhenvald 2000: 277; Heine 1982: 191). However, size as an assignment criterion is widespread in Africa, where it e.g., occurs in diminutive and augmentative genders as reported by Di Garbo (2014). An example is in Tonga, where ‘boy’ (noun class 1) can shift to the diminutive noun class 12 to highlight smallness:

Table 14: Overlap between the gender and verb prefix classes of Mian (adapted from Fedden & Corbett 2017: 34). Cells with examples show the attested combinations.

	Masculine	Feminine	Neuter 1	Neuter 2
M-classifier	man, boy, boar	–	sleeping bag, plate, mosquito net	–
F-classifier	–	woman, girl, sow	–	house, steel axe, money
Long	–	–	tobacco, eating implement, bush knife	–
Bundle	–	–	string bag, plastic bag	–
Covering	–	–	blanket, band aid	–
Residue	–	tortoise, scorpion	cassowary egg, plane, hat	–

(10) Tonga (Bantu) (Di Garbo 2014: 147; from Carter 2002: 21)

- a. *mu-sankwa*
CL1-boy
'boy'
- b. *tu-sankwa*
CL12-boy
'small boy'

As for New Guinea, its prevalence specifically in the Sepik area has led Aikhenvald (2008: 113) to suggest that gender assignment based on size and shape may actually be an areal feature of the Sepik area. Indeed, all four languages in this sample found to have such systems are spoken in or near the Sepik area: Abau (Sepik) and Manambu (Ndu) are spoken inside it, while Skou (Sko) and Taiap (isolate) are spoken in relatively adjacent areas. Another oft-cited example is Alamblak (Bruce 1984; not in the sample), also a Sepik language of the same area, which has a system similar to that of Manambu (Aikhenvald 2008: 112).

Thus, gender assignment according to size and shape appears to be an areal feature, since it occurs in a wide area and in languages of different families. This

gives rise to an important question. Why would a system of gender assignment be areal when gender is such a stable and not easily borrowed feature? Although this is far beyond the scope of this study, there are some hints that this may be part of a larger cultural classificatory system (i.e., perceptual, not linguistic). The reason for such a possibility is that besides occurring in and around the Sepik area, there are other New Guinean languages where nouns are grouped based on size and shape with other nouns denoting male or female referents, even when there is no gender system. This is most apparent in the TNG languages of the central highlands; nouns in these languages can be categorized by the type of stance verb they occur with, so that males or large, long, or tall objects occur with ‘stand’, whereas women or small, short, or round objects occur with ‘sit’ (Foley 2000: 372). An example of such a language is Enga (Engan; New Guinea Highlands; not in the sample), which has seven different stance verbs, including *katengé* ‘stand’, which is used for referents considered tall, large, strong, and/or powerful such as ‘men’, ‘house’, and ‘tree’, and *pentengé* ‘sit’, which is used for referents considered small, squat, horizontal, and/or weak such as ‘woman’, ‘possum’, and ‘pond’ (Aikhenvald 2000: 158–159; Rumsey 2002). Thus, it appears that the perception of large, long, or tall objects being related to males and/or masculinity, and small, short, or round objects being related to females and/or femininity is a characteristic of New Guinea that extends beyond gender systems or the Sepik area.

5.2 Two separate systems of noun classification

In most gendered languages, gender constitutes a single system where each noun is assigned to a single class which is reflected in the form of indexation targets. However, there are also languages with two separate systems, both of which appear to constitute or be related to gender systems, but occur with different types of targets. Thus, in such a language each noun is assigned to not just one class, but to two different classes. In the sample of this study, five languages have such systems (see Table 15).

Even in the small sample of this study, the two separate systems range from languages with two more or less equally complex systems (i.e., with similar numbers of forms and uses) to languages where one system is more complex whereas the other is much less so. In order to retain the typological comparability of the results, a distinction has been made between systems of gender and systems of noun classifiers. However, it should be stated that there is a thin line between the two and they most certainly constitute two edges of the same continuum.

Table 15: Languages in the sample with separate gender and noun class systems

Separate systems	No. of lgs.	%	Languages
Yes	5	25%	Abau Burmeso Mian Motuna Rotokas
No	15	75%	Ama Au Bukiyip Kuot Manambu Maybrat Mende Nalca Oksapmin Skou Taiap Teop Walman Warapu Yimas
Total:	20	100%	

Following these, four of the five languages with two systems of noun classification can be argued to exhibit one gender system and system of noun classifiers, whereas only Burmeso has two systems which both satisfy the conditions for gender systems. In the first system, Burmeso has three genders (masculine, feminine, and neuter), appearing as adjectival agreement suffixes (11a), which are further divided into two subgenders (animate and inanimate), each depending on the plural agreement marker (Donohue 2001: 105–106). However, in the second system (which Donohue calls a noun class system), Burmeso has six genders (I–VI), which occur in verbal agreement prefixes (11b) (Donohue 2001: 101). In addition, there are three words which take both kinds of agreement: *-aysa-* ‘one’, *-akasu-* ‘all’, and *-asna-* ‘white’ (11c).

(11) Burmeso (isolate) (Donohue 2001: 105, 109, 100)

- a. *Da de koya bek-**abo**.*
 1SG 1SG.POSS grandfather good-M.SG
 ‘My grandfather is well.’
- b. *Da mibo j-**ihi**-maru.*
 1SG banana V.SG-see-TPST
 ‘I saw a banana.’
- c. *Sunam **n-asna-b**.*
 axe.SG III.SG-white-M.SG
 ‘(The) axe is white.’

As expected from the number of genders being different, the two systems use different assignment rules. Both systems are sex-based with importance clearly put on sex and animacy, but none of them have only transparent semantic rules: e.g., ‘wind’ is neuter/III, ‘rain’ masculine/IV, and ‘star’ masculine/III (Donohue 2001: 103–107). A comparison of the overlap of the two systems is exemplified in Table 16, showing how members are assigned to both systems.

Near the other end of the spectrum lies Rotokas (North Bougainville). Rotokas has three genders, viz., masculine, feminine, and neuter, which appear e.g., in pronouns, demonstratives, adjectives, and verbs (12a) (Robinson 2011). However, Rotokas also has noun classifiers, which consist of two different sets. The first set consists of four classifiers; these distinguish between shape and size, and importantly occur on both attributive (12b) and predicative modifiers of the classified noun (Robinson 2011: 50).

(12) Rotokas (North Bougainville) (Robinson 2011: 149, 50)

- a. *Pita vaio ora Kariri **ava-si-ei** voka-sia*
 P. DL.ANIM and K. go-3DL.M-PRS walk-DEP.SEQ
 ‘Peter and Kariri are going for a walk.’
- b. *gorupasi **isi** rutu karuvera **isi** aio-a-voi*
 strong CL.round very Singapore CL.round eat-1SG-PRS
 ‘I am eating a really strong Singapore fruit.’

The other set of classifiers, which has more members and have collective meanings, occurs following, or instead of, the classified noun (Robinson 2011: 51). Interesting to note is that classified nouns become neuter in regards to gender agreement (Robinson 2011: 53).

Table 16: Comparison between genders and noun classes in Burmese (adapted from Donohue 2001: 108). As in the source, the subgenders of the masculine, feminine, and neuter genders are shown.

Class	Masculine	Feminine	Neuter	M	F	N
				inanimate	inanimate	animate
I	male humans (most birds, animals etc.) 2SG PRO 1SG PRO	(birds of paradise) pigeons sago garden female humans	sea	neck	-	sea wound
II			nose ear eye	-	small goanna	string shapes sago rinser (lower)
III	axe papaya ground bird	-	(some small animals) bench	papaya rattan mountain lake (all tubers) upper sago trough female child	goanna	-
IV	head, flesh, feces, finger, elbow, sun, cloud, rain, sand, mud	-	-	-	-	-
V	-	-	(arrows)	banana	-	-
VI	-	-	-	coconut	-	-

Table 17: Numeral classifiers in Abau (adapted from Lock 2011:57)

Class	Characteristics	One	Two	Three
1	Human beings; spirits	<i>pru-eyn</i>	<i>pru-eyns</i>	<i>pru-ompri</i>
2	Non-human	<i>ka-mon</i>	<i>k-reys</i>	<i>k-rompri</i>
3	Small objects with some volume	<i>na-mon</i>	<i>na-reys</i>	<i>na-rompri</i>
4	Flat surface objects; experience nouns	<i>si-rom</i>	<i>s-eyns</i>	<i>s-ompri</i>
5	Long, relatively thin objects	<i>pi-ron</i>	<i>pi-reys</i>	<i>pi-rompri</i>
6	Geographical locations	<i>u-mon</i>	<i>u-reys</i>	<i>u-rompri</i>
7	Flat objects with hardly any volume	<i>i-mon</i>	<i>i-reys</i>	<i>i-rompri</i>
8	Certain type trees	<i>li-mon</i>	<i>li-reys</i>	<i>li-rompri</i>
9	Bundles of long non-cut items	<i>ein-mon</i>	<i>ein-deys</i>	<i>ein-rompri</i>
10	Temporal	<i>leik-mon</i>	<i>leik-reys</i>	<i>leik-rompri</i>
11	Bundles of long cut items	<i>hnaw-mon</i>	<i>hnaw-reys</i>	<i>hnaw-rompri</i>
12	Part of a long object	<i>houk-mon</i>	<i>houk-reys</i>	<i>houk-rompri</i>

Abau also exhibits a clear noun classifier system (Table 17). There are two genders in Abau, masculine and feminine, which follow opaque gender assignment rules and appear in e.g., pronouns and demonstratives. However, the numerals ‘one’, ‘two’, and ‘three’ do not agree with this system, but instead take one of twelve prefixes based on semantic criteria of the referent. However, the same noun can be used with different numeral classifiers in order to indicate a specific referent, so that e.g., *su piron* ‘one coconut’ refers to the whole coconut palm and not just the fruit, since class 5 signals long objects, while *su kamon* ‘one coconut’ is used when referring to just the fruit, since class 2 does not carry the semantic feature of length. It is thus evident that this system of noun classifiers is not lexically determined by the noun itself and thus not a gender system.

Mian has a similar albeit different system. In Mian, there is a set of verbal classificatory prefixes which are divided into six classes (Table 18). These prefixes are used only for around 50 verbs, the vast majority of which refer to forms of object manipulation, movement, and handling (Fedden 2011: 172). Once again, this is clearly not a full-fledged gender system, but rather a classifier system.

Table 18: Classifiers in Mian (adapted from Fedden 2011: 172)

Class	Characteristics	Verbal classificatory prefixes	
		Singular	Plural
1	Masculine	<i>do(b)-</i>	<i>do(l)-, dl-</i>
2	Feminine	<i>om-</i>	
3	Long object	<i>to(b)-</i>	<i>tebe(l)-</i>
4	Bundle-like object	<i>go(l)-</i>	<i>gule(l)-</i>
5	Flat object	<i>gam-</i>	<i>geme(l)-</i>
6	Residue class	<i>o(b)-</i>	<i>o(l)-</i>

Finally, Motuna is a particularly interesting case since its secondary system lies near the boundary between genders and noun classifiers. Besides its gender system (described in §3.3), Motuna has another noun classification system consisting of 51 different classifiers, which are visible in the forms of adjectives, verbs, participial clauses, articles, demonstratives, possessive pronouns, and numerals (Onishi 1994: 162–163). Thus, as for indexation, the system is very reminiscent of a gender system. However, the classes are not lexically determined, meaning that the same noun may occur with various classifiers depending on the referent. Furthermore, as expected for a noun classifier system, the classifiers refer properties such as size, shape, type of vegetable, and collectives (e.g., ‘bundle’, ‘packet’). Thus, *moo* ‘coconut’ can occur with classes 4 *-mung* ‘plant/fruit/nut/egg/things made of plant/coin’ (> ‘coconut (nut/tree)'), 5 *-ri* ‘nut with hard shell’ (> ‘coconut’), 6 *-mo* ‘bunch of nuts’ (> ‘coconut’), 13 *-ri* ‘round object’ (> ‘coconut’), and 30 *-ita* ‘half/side’ (> ‘half coconut shell’) (Onishi 1994: 166–167). Therefore, this system in Motuna is a system of noun classifiers, not genders.

Despite the small size of the sample used in this study, the proportion and the geographic and genealogical spread of languages with two separate systems of nominal classification indicate that the phenomenon is rather common and widespread in New Guinea. Besides the languages of this study, two of which are mentioned by Foley (2000: 373), viz., Burmeso and Motuna, similar systems have been noted in the Sepik languages Iwam, Wogamusin, and Chenapien, which together with their relative Abau (which is included in this sample) suggest that this is a feature of the Sepik family (Lock 2011: 46). However, it does not appear to be common outside of New Guinea, as systems of this type only occur in a few Indic, Dravidian, Iranian, and some Arawak languages (Aikhenvald 2008: 185).

5.3 No gender distinctions in pronouns

According to Greenberg's (1963: 90) 43rd Universal, "[if] a language has gender categories in the noun, it has gender categories in the pronoun".¹⁵ However, this generalization is not reflected in the languages sampled for this study, where four languages do not exhibit gender in pronouns (see Table 19).¹⁶

Table 19: Occurrence of gender distinctions in independent pronouns in the sample

Gender in pronouns	No. of lgs.	%	Languages
Yes	16	80%	Abau Au Bukiyip Kuot Manambu Maybrat Mende Mian Motuna Oksapmin Rotokas Skou Taiap Walman Warapu Yimas
No	4	20%	Ama Burmeseo Nalca Teop
Total:	20	100%	

¹⁵'Pronoun' is here understood as 'independent pronoun'.

¹⁶ As in §3.4, the demonstratives in Kuot and Yimas with pronominal functions are here understood as pronouns for the purpose of typological comparison, just as the present author would do for the Latin *is*, *ea*, and *id*, regardless of the proper language-internal analysis. Nevertheless, if they should rather not be regarded as pronouns, the point of this section would be even stronger.

As seen in the above table, almost a quarter of the languages in the sample have no gender distinctions in independent pronouns. In comparison, only two languages (Mende and Menya) have gender distinctions solely in pronouns.

While these results are interesting, the phenomenon can be found in other languages as well. This can be investigated by comparing two *WALS* chapters, viz., Corbett's (2013a) chapter on number of genders and Siewierska's (2013) chapter on gender distinctions in independent pronouns. These chapters do not share the same sample: Corbett's sample consists of 257 languages, whereas Siewierska's contains 378 languages. Of these languages, 188 occur in both samples, 74 of which have gender systems. Of these remaining 74 gendered languages (which of course should not be assumed to be representative of anything), surprisingly, 15 languages (20%) do not show gender distinctions in independent pronouns. Coincidentally, this is the same ratio as in New Guinea as shown in Table 18 above. Thus, it is clear that Greenberg's statement is not universal, although it certainly is a common pattern.

5.4 Gender indexation on verbs

According to Greenberg's 31st Universal, "if either the subject or object noun agrees with the verb in gender, then the adjective always agrees with the noun in gender." That is, if the verbs are indexing targets, so are adjectives. However, this generalization is not reflected in the distribution of values of the fourth classification criteria in the languages sampled for this study (see Table 8). Three of the 15 languages with gender marking on verbs show no indexation on adjectives.

The results are even more striking when compared with Bybee (1985). In her survey of fifty languages, only 16% of the languages showed gender in verbs (Bybee 1985: 18). However, in the sample of this survey, 75% of the languages have gender marking on verbs, with Ama even having it as the only indexing target. Verbs thus seem to be more prototypical indexing targets than adjectives in the sample of this study, and it would be interesting to conduct further studies on this with a larger and worldwide sample.

6 Conclusions and further studies

The languages of New Guinea show remarkable diversity in grammatical gender, but there are still common patterns. Except Teop (Austronesian, Oceanic), all languages in the sample have sex-based gender systems. More than half of the languages have only two genders, and only Bukiyip (Torricelli) and Yimas (Lower

Sepik) have very large systems, with 18 and around a dozen genders respectively. In the vast majority of the languages, gender assignment is semantic. Half of the languages have four or more indexing targets, most commonly pronouns and verbs. Gender marking on nouns is rare and occurs in only three languages in the sample. The typological comparison suggests that the genders systems of New Guinea are remarkably representative of the world. Sex-based gender systems are more common in both New Guinea and the world, and the ratio of numbers of genders are very similar, with the rate of occurrence of the values being two > three ≥ five or more > four genders. Semantic and formal gender assignment occurs in slightly more than half of the languages of the world, while it is much more uncommon in New Guinea. The gender systems of New Guinea and Africa are very different. This depends largely on the numerous Bantu languages, which make the languages of Africa whole less diverse than the sample of this study. The most significant difference is the prevalence of non-sex-based gender systems and gender marking on nouns in Africa, whereas the opposite is true in New Guinea. This suggests that they may have different diachronical origins.

Four special characteristics have been found in the gender systems of New Guinea, none of which are typologically common. Firstly, four languages of the sample share the property of size and shape as important criteria for gender assignment. In these languages, nouns denoting large and/or long objects are masculine, whereas small and/or short items are feminine. This characteristic is also shared with many African languages. Secondly, five languages of the sample have two separate nominal classification systems. In these languages, each noun is assigned to two classes which are reflected in different indexing targets, although only Burmeso exhibits two equivalent gender systems whereas the others rather distinguish between genders and noun classifiers. Thirdly, four languages in the sample have no gender distinctions in pronouns, which is unexpected according to Greenberg's 43rd Universal. Finally, verbs are the most common gender-indexing targets in the languages of the sample, which is uncommon. In three languages of the sample, verbs are indexing targets while adjectives are not, which contradicts Greenberg's 31st Universal.

Future studies should consider more languages and be proportional, as well as aim at investigating how the gender systems of New Guinea may affect the theory of gender. There are also more specific areas of study that would benefit from further research. Firstly, the special characteristics discussed in this study could benefit from more research. One example is gender assignment based on size and shape, which appears to be a feature of the Sepik area. However, Skou (Sko) and Taiap (isolate) are spoken outside of the immediate area, and similar distinctions

have been found in non-gendered languages of New Guinea. It would thus be interesting to investigate the actual geographical distribution of such systems. Also, the inclusion of the criterion of manipulability of gender assignment as used in Di Garbo (2014) would probably further improve the comparison between gender in New Guinea with Africa.

It would also be interesting to investigate features not discussed in this study. One such feature is pluralia tantum, i.e., plural nouns with no or only an unusual singular form (Koptjevskaja-Tamm & Wälchli 2001: 629), for which there are indications that it may be relevant for gender. This can be seen in Ama (Left May), which has a separate compound gender containing nouns denoting referents with many parts, e.g., heaps, piles, and mass nouns (Årsjö 1999: 68). For a discussion of pluralia tantum in languages of New Guinea see also Olsson (2019 [this volume]) and Dryer (2019 [this volume]).

Future studies could also investigate the diachrony of gender in New Guinea. Some languages of New Guinea have been found to have diachronically young gender systems, including Nalca (TNG, Mek) of the sample of the present study, and the prevalence of sex-based systems suggest that many gender systems in New Guinea have diachronic origins different from e.g., the non-sex-based gender systems of Africa.

Special abbreviations

The following abbreviations are not found in the Leipzig Glossing Rules:

I, II, III etc.	gender I, II, III etc.	INTEN	intensifier
ANIM	animate	N-	non-
C	common gender	NNOUN	non-noun gender
CL	classifier	PRO	pronoun
CO1	common gender 1	RED	reduplication
DEFAULT	default gender	SEQ	sequential
DEP	dependent (verb)	TPST	today's past/hodiernal past
DL	dual	U	unmarked gender
HAB	habitual		

References

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

- Aikhenvald, Alexandra Y. 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.
- Årsjö, Britten. 1999. *Words in Ama*. Uppsala: Uppsala University MA thesis.
- Bakker, Dik. 2012. Language sampling. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology* (Oxford Handbooks Online). Oxford: Oxford University Press. <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199281251.001.0001/oxfordhb-9780199281251>, accessed 2015-5-15.
- Brown, Lea & Matthew S. Dryer. 2008. Diminutive as an inflectional category in Walman. Unpublished manuscript.
- Bruce, Leslie P. 1984. *The Alambalak language of Papua New Guinea (East Sepik)* (Pacific Linguistics C 81). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Bybee, Joan. 1985. Diagrammatic iconicity in stem-inflection relations. In John Haiman (ed.), *Iconicity in syntax*, 11–48. Amsterdam: John Benjamins.
- Carter, Hazel. 2002. *An outline of Chitonga grammar*. Lousaka: Bookwords Publishers.
- Conrad, Robert J. & Kepas Wogiga. 1991. *An outline of Bukiyip grammar* (Pacific Linguistics C 113). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013a. Number of genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/30>.
- Corbett, Greville G. 2013b. Sex-based and non-sex-based gender systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/31>.
- Corbett, Greville G. 2013c. Systems of gender assignment. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/32>.
- Corris, Miriam. 2005. *A grammar of Barupu: A language of Papua New Guinea*. Sydney: University of Sydney dissertation.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William. 2003. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press.

- Croft, William. 2013. Agreement as anaphora, anaphora as coreference. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 95–118. Berlin: De Gruyter.
- Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa*. Stockholm University dissertation.
- Dol, Philomena Hedwig. 2007. *A grammar of Maybrat: A language of the Bird's Head Peninsula, Papua Province, Indonesia*. Canberra: Pacific Linguistics.
- Donohue, Mark. 2001. Animacy, class and gender in Burmeso. In Andrew Pawley, Malcolm Ross & Darrell Tryon (eds.), *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton* (Pacific Linguistics 514), 97–115. Canberra: Australian National University.
- Donohue, Mark. 2004. A grammar of the Skou language of New Guinea. http://pubman.mpdl.mpg.de/pubman/item/escidoc:402710/component/escidoc:402709/skou%5C_donohue2004%5C_s.pdf, accessed 2015-2-6. Unpublished manuscript.
- Dryer, Matthew S. 2019. Gender in Walman. In Francesca Di Garbo, Bruno Olson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 171–196. Berlin: Language Science Press. DOI:10.5281/zenodo.3462766
- Fedden, Sebastian. 2011. *A grammar of Mian*. Berlin: Mouton de Gruyter.
- Fedden, Sebastian & 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. DOI:10.5334/gjgl.177
- Foley, William A. 1991. *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- Foley, William A. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29. 357–404.
- Frowein, Friedel Martin. 2011. *A grammar of Siar: An Oceanic language of New Ireland province, Papua New Guinea*. Victoria: La Trobe University dissertation.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge: MIT Press.
- Heine, Bernd. 1982. African noun class systems. In Hansjakob Seiler & Christian Lehmann (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen, Teil I: Bereich und Ordnung der Phänomene*, 189–216. Tübingen: Narr.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.

- Hoel, Hanna Marie, Tarja Ikäheimonen & Michiyo Nozawa. 1994. Mende grammar essentials. http://www-01.sil.org/pacific/png/pubs/928474531144/Mende_Grammar_Essentials.pdf, accessed 2015-3-20. Unpublished manuscript.
- Iemmolo, Giorgio. 2011. *Towards a typological study of differential object marking and differential object indexation*. Pavia: Università degli studi di Pavia dissertation.
- Koptjevskaja-Tamm, Maria & Bernhard Wälchli. 2001. The Circum-Baltic languages: An areal-typological approach. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic languages: Typology and contact*, vol. 2, 615–750. Amsterdam: John Benjamins.
- Kulick, Don & Christopher Stroud. 1992. The structure of the Taiap (Gapun) language. In Thomas E. Dutton, Malcolm Ross & Darrell T. Tryon (eds.), *The language game: Papers in memory of Donald C. Laycock* (Pacific Linguistics C 110), 203–226. Canberra: Australian National University.
- Lindström, Eva. 2002. *Topics in the grammar of Kuot: A non-Austronesian language of New Ireland, Papua New Guinea*. Stockholm: Stockholm University dissertation.
- Lock, Arjen. 2011. *Abau grammar* (Data Papers on Papua New Guinea Languages 57). Ukarumpa: SIL-PNG Academic Publications.
- Loughnane, Robyn. 2009. *A grammar of Oksapmin*. University of Melbourne dissertation.
- Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45(2). 435–464.
- Mosel, Ulrike & Ruth Spriggs. 2000. Gender in Teop (Bougainville, Papua New Guinea). In Barbara Unterbeck (ed.), *Gender in grammar and cognition I: Approaches to gender* (Trends in Linguistics. Studies and Monographs 124), 321–349. Berlin: Mouton de Gruyter.
- Olsson, Bruno. 2019. The gender system of Coastal Marind. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 197–223. Berlin: Language Science Press. DOI:10.5281/zenodo.3462768
- Onishi, Masayuki. 1994. *A grammar of Motuna (Bougainville, Papua New Guinea)*. Canberra: Australian National University dissertation.
- Phillips, Colin. 1993. Conditions on agreement in Yimas. In Jonathan D. Bobaljik & Colin Phillips (eds.), *Papers on case & agreement I* (MIT Working Papers in Linguistics 18), 173–213. Cambridge: MIT.
- Robinson, Stuart Payton. 2011. *Split intransitivity in Rotokas, a Papuan language of Bougainville*. Nijmegen: Radboud Universiteit dissertation.

- Rumsey, Alan. 2002. On the grammaticalization of posture verbs in Papuan languages, its bodily basis and cultural correlates. In John Newman (ed.), *The linguistics of sitting, standing and lying*, 179–211. Amsterdam: John Benjamins.
- Scorza, David. 1985. A sketch of Au morphology and syntax. In *Papers in New Guinea linguistics 22* (Pacific Linguistics A 63), 215–273. Canberra: Australian National University.
- Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/44>.
- Steele, Susan. 1978. Word order variation: A typological study. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language*, vol. 4: *Syntax*, 585–624. Stanford: Stanford University Press.
- Svärd, Erik. 2013. *Selected topics in the grammar of Nalca*. Stockholm University BA-thesis. <http://www.diva-portal.org/smash/get/diva2:631006/FULLTEXT02.pdf>, accessed 2013-11-9.
- Svärd, Erik. 2015. *Grammatical gender in New Guinea*. Stockholm: Stockholm University MA thesis.
- Tharp, Doug. 1996. Sulka grammar essentials. In John M. Clifton (ed.), *Two non-Austronesian grammars from the islands*, 77–179. Ukarumpa: Summer Institute of Linguistics.
- Wälchli, Bernhard. 2018. The rise of gender in Nalca (Mek, Tanah Papua): The drift towards the canonical gender attractor. In Sebastian Fedden, Jenny Audring & Greville G. Corbett (eds.), *Non-canonical gender systems*, 68–99. Oxford: Oxford University Press.