

Grammatical gender in New Guinea

Erik Svärd

Department of Linguistics
Degree 30 HE credits
General Linguistics
Independent Project for the Degree of Master (120 credits)
Spring term 2015
Supervisor: Bernhard Wälchli
Examinator: Henrik Liljegren
Expert reviewer: Eva Lindström



Grammatical gender in New Guinea

Erik Svärd

Abstract

The present study investigates the gender systems of 20 languages in the New Guinea region, an often overlooked area in typological research. The languages were classified with five criteria used by Di Garbo (2014) to classify gender systems of African languages. The results showed that the gender systems were 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 much less common. 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. However, gender in Africa is also less diverse largely due to the numerous Bantu languages. Finally, four typologically rare characteristics were found in the sample: (1) size and shape as important criteria of gender assignment, with large/long being masculine and small/short feminine, (2) the presence 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 languages, typology

Sammanfattning

Denna studie undersöker genussystemen hos 20 språk i Nya Guinea-regionen, vilken ofta förbises i typologisk forskning. Språken klassificerades utifrån fem kriterier som användes av Di Garbo (2014) för att klassificera genussystem i Afrika. Resultaten visade att genussystemen var varierade, men ungefär hälften av språken har könsbaserade genussystem med två genus, semantisk genustilldelning, fler än fyra genusindex och ingen genusmarkering på substantiv. Genussystemen är anmärkningsvärt representativa för världen, men formell genustilldelning är mycket mindre vanlig. Jämfört med genussystemen i Afrika är dock Nya Guinea väldigt annorlunda. Den viktigaste skillnaden är den större utbredningen av icke-könsbaserade genussystem och genusmarkering på substantiv i Afrika, medan motsatsen gäller i Nya Guinea. Genus i Afrika är dock till stor del mindre varierat på grund av de talrika bantuspråken. Slutligen hittades fyra typologiskt sällsynta karaktärsdrag i urvalet: (1) storlek och form som viktiga kriterier för genustilldelning, där stort/långt är maskulint och litet/kort feminint, (2) närvaron av två separata nominalklassificeringssystem, (3) inga genusdistinktioner i pronomen och (4) verb som det vanligaste genusindexet.

Nyckelord

grammatiskt genus, indexering, kongruens, Nya Guinea, papuanska språk, typologi

Abbreviations

The abbreviations used for the glossings are based on the Leipzig Glossing Rules¹. Note that no separating character follows person.

I, II, III etc.	gender I, II, III etc.	M	masculine
1	first person	N	neuter
2	second person	N-	non-
3	third person	NMLZ	nominalizer
ABS	absolutive	NNOUN	non-noun gender
ADV	adverb	NOUN	noun
ANIM	animate	OBJ	object
ART	article	PFV	perfective
C	common gender	PL	plural
CL	classifier	POSS	possessive
CO1	common 1 gender	PRF	perfect
DEF	definite (article)	PRO	pronoun
DEFAULT	default gender	PROX	proximal
DEM	demonstrative	PRS	present
DEP	dependent (verb)	PST	past
DET	determiner	Q	question word
DL	dual	RED	reduplication
F	feminine	SBJ	subject
FUT	future	SEQ	sequential
GEN	genitive	SG	singular
HAB	habitual	TOP	topic
INTEN	intensifier	TPST	today's past / hodiernal past
IPFV	imperfective	U	unmarked gender
IPST	immediate past		

¹ See <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>

Table of Contents

1. Introduction	1
1.1 Purpose.....	1
2. Background	2
2.1 Grammatical gender and indexation.....	2
2.1.1 Theoretical background.....	2
2.1.2 Defining gender.....	3
2.2 Typology of gender.....	7
2.2.1 Overview.....	7
2.2.2 New Guinea.....	8
3. Method and data	10
3.1 Sampling and material.....	10
3.2 Procedure.....	12
3.3 Limitations.....	13
4. Results	15
4.1 Sex-based and non-sex-based gender systems.....	15
4.2 Number of genders.....	19
4.3 Gender assignment.....	22
4.4 Number of gender-indexing targets.....	26
4.5 Occurrence of gender marking on nouns.....	31
5. Discussion	35
5.1 Typological comparison.....	35
5.1.1 The World.....	35
5.1.2 Africa.....	38
5.2 Special characteristics.....	43
5.2.1 Size and shape.....	44
5.2.2 Two separate systems of noun classification.....	46
5.2.3 No gender distinctions in pronouns.....	50
5.2.4 Gender in verbs.....	51
5.3 Suggestions for further research.....	51
6. Conclusions	53
Appendix A. The language sample	54
Appendix B. Overview of classification criteria	55
Appendix C. Samples of gender descriptions	56
C.1 Abau (Sepik).....	56
C.2 Teop (Austronesian, Oceanic).....	59
References	62

1. Introduction

Grammatical gender is a well-researched and common linguistic phenomenon. It is widespread in the languages of the world but clusters in families, and although there are geographical patterns, gender systems are often very stable and not prone to borrowing. Traditionally, most typological research has focused on languages in Eurasia and Africa, but important research has also been carried out on other continents such as the Americas and Australia. Less researched is the region of New Guinea, which often has been overlooked in typology despite containing as many as one sixth of all languages of the world. However, in recent years, studies and grammatical descriptions have increasingly shed light on languages with uncommon properties. This includes not least some languages of New Guinea, which have been shown to have exhibit many unusual grammatical phenomena concerning e.g., gender. Despite this, there remains a lack of overviews of New Guinea in general and in particular regarding gender systems in the region. The purpose of this thesis is to counteract this issue by investigating various New Guinean languages which have been observed to have gender. These are then classified by various criteria with the aim to provide an account of gender systems of New Guinea in general. Finally, the results are compared with both the languages of Africa and the world as a whole, and with previous research on gender in New Guinea.

1.1 Purpose

The purpose of this thesis is to investigate 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 intention is to investigate if there are any typologically rare phenomena in gender which are more common in the languages of New Guinea. The aim is thus to provide a classification of gender systems of New Guinea and to put these results in a typological perspective. More specifically, the research questions of the proposed thesis are:

1. How is grammatical gender expressed in a diverse sample of 20 languages of New Guinea?
2. How do the gender systems of New Guinea compare with other geographical areas and the world?
3. 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.

2. Background

2.1 Grammatical gender and indexation

2.1.1 Theoretical background

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. This classical definition, while more than half a century old, is still fundamental to the definition of gender. It is cited e.g., by Corbett (1991:4), who in his now likewise classical book calls gender “the most puzzling of the grammatical categories” (Corbett 1991:1).

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.

However, more recently, Corbett’s notion of agreement has received criticism. Di Garbo (2014:8) gives a few examples, including 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 thesis (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. However, there are reasons for arguing that both terms should be used, based on the perceived differences between agreement targets within the noun phrase, such as adjectives, and pronouns. An example of the surfacing of this can be seen in many African languages. According to Heine (1982:195), most African languages have more than one set of agreement markers, which in most cases can be reduced to two basic sets, which he refers to as *nominal gender* and *pronominal gender*, as represented in Swahili (Niger-Congo, Bantu) in Table 1. Heine further states that nominal agreement sets typically are found with categories like adjective and numeral, whereas pronominal marker sets typically occur with personal or other pronouns.

Table 1: The nominal and pronominal agreement sets in Swahili (adapted from Heine 1982:196).

Class	Nominal	Pronominal
1	<i>m-</i>	<i>a-, yu-, u-, m-</i>
2	<i>wa-</i>	<i>wa-</i>
3	<i>m-</i>	<i>u-</i>
4	<i>mi-</i>	<i>I-</i>
5	<i>(ji-)</i>	<i>li-</i>
6	<i>ma-</i>	<i>ya-</i>
7	<i>ki-</i>	<i>ki-</i>
8	<i>vi</i>	<i>vi-</i>
9	<i>N-</i>	<i>I-</i>
10	<i>N-</i>	<i>zi-</i>
11/14	<i>u-, m-</i>	<i>u-</i>
15	<i>ku-</i>	<i>ku-</i>
16		<i>pa-</i>
17	<i>(-ni)</i>	<i>ku-</i>
18		<i>mu-</i>

These observations by Heine appear to not only be relevant for African languages, and recently the idea has been put forth that what Heine terms pronominal agreement should be considered indexation instead. This is discussed e.g., by Haspelmath (2013), who argues that the concept of indexing is much more useful in typology for bound person forms than the concepts of pronoun or agreement. This usage of indexation for anaphora is further expanded upon by Croft (2013), who questions the supposed contrast between bound person forms and independent pronouns. The resulting distinction is thus one where indexation is used for anaphoric gender, whereas agreement is used for gender agreement within the noun phrase.

2.1.2 Defining gender

The definition of gender used in this thesis is very similar to the one used by Di Garbo (2014). Most importantly, the term indexation is used as a cover term for the same linguistic features as Corbett's (1991) notion of agreement. 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 which are still not widely accepted. However, since indexation is gaining ground, it is embraced in this essay. Thus, in order to facilitate for the reader, the full definition of gender used in this thesis is described and discussed below.

A language is considered as having gender if (1) every noun is assigned to one default noun class, i.e., a *gender*, (2) which is reflected via *indexation* in the form of other words, i.e., (*indexing*) *targets*. Indexation may occur either directly within the same noun phrase or sentence, or without any explicit mention of the noun itself or outside the local domain. In order to separate genders from noun classifiers, further criteria are used, viz., gender must be lexically determined and a closed set of classes. If a noun classification is exclusively related to the referent or a salient feature of a noun, it is not considered gender, nor is it so analyzed if the set of classes is open or at least not obviously closed. Important to note is that *gender* and *noun class* are considered equivalent terms.

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 (Austronesian, Oceanic) (Frowein 2011; not in the sample), 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. Thus, the case marker hosts in Nalca (TNG, Mek) (see 4.4) are indexing targets, since they express gender, case, and demonstrative. In comparison, the articles in Teop (Austronesian, Oceanic) (see also 4.4), which precede adjectives and numerals but encode only gender and number, are not considered targets; instead, they are analyzed as forming wholes with the following adjective or numeral, so that Teop has two indexing targets (adjectives and numerals) rather than one (articles).

The assignment of a noun to a gender is commonly perceived of as exclusive, i.e., a particular noun is used with one and only one gender. However, as will be seen in 4.3, this is not a universal principle. In e.g., Manambu, which has two genders (masculine and feminine), a noun may change gender (i.e., the indexation forms) in order to emphasize specific semantic attributes: e.g., the feminine noun ‘house’ becomes masculine when referring to a particularly large house. Intuitively, it could be argued that nouns in such a language are not assigned to one specific gender. However, in all languages of the sample, including Manambu, each noun always has one default gender to which it prototypically belongs following gender assignment rules. Thus, ‘house’ in Manambu is by default a feminine noun. It only becomes masculine when emphasizing certain aspect, i.e., large size. Therefore, only large houses (naturally depending on the speaker’s perception) can be used with masculine indexation, but all houses (including large ones) can use the feminine forms. Compare this to e.g., the Spanish *hermano* ‘brother’ and *hermana* ‘sister’, which are masculine and feminine respectively. These are mutually exclusive: that is, *hermano* can never be used with feminine indexation forms and vice versa, nor can *hermana* refer to a referent for which *hermano* is used. In contrast, an emphatically large house in Manambu can indeed switch to the masculine gender, but that does not exclude it from being used with its default feminine gender instead. For this reason, the notion of default noun class is employed in the definition above in order to emphasize the fact that nouns by default belong to a certain class, although in some languages nouns may change to another gender, e.g., in order to emphasize specific semantic attributes. Nevertheless, following this definition each noun still is assigned to one and only one gender.

In order to delimit the number of genders in a language, the following criteria are used: (1) genders are defined via the opposition of distinctive forms of indexation targets; (2) if a language exhibits more indexation forms for some targets than others, these will be considered subgenders if they only occur

with a minority of targets or in highly idiomatic or archaic uses; and, (3) a language will be considered as having more than one gender system only if the respective classes cannot be explained as subgenders. Furthermore, homophony does not override the criterion that each noun be assigned to one class; thus, if some nouns appear to belong to one gender in the singular and another in the plural, they are considered as belonging to separate genders.

The most basic distinction to be made is between languages with and without gender. This is fairly straightforward due to the definitions above: if every noun is assigned to a class which is apparent in the form of other words, the language has gender. Thus, a language such as French or Russian has gender, since each noun is assigned a class which is seen in the forms of e.g., adjectives and demonstratives. Even a language such as English has gender, since every noun is assigned to a gender (albeit via very simple semantic rules), which is apparent in the choice of third person singular pronoun forms. In contrast, a good example of a language without gender is Kire (Lower Ramu, Lower Sepik-Ramu). Each noun is assigned to one of four classes and various subclasses, which are apparent in the singular, dual, and plural suffixes of the noun; e.g., nouns of Class 1 follow the pattern singular *-m* and plural *-a*, e.g., *ppum* ‘hand drum’, plural *ppua*, whereas Class 2C follows the pattern singular *-n/-ŋ* and plural *-i*, e.g., *iran* ‘ant’, plural *irani* (Stanhope 1972:54–56). This is not considered gender, since these classes are not reflected in other words.

More complicated than determining the presence of gender is delimiting the number of genders. In many languages, homophony is a major culprit for such difficulties. A commonly mentioned example is Romanian, which has three genders (masculine, feminine, and neuter), which are evident e.g., in adjectives. However, in the singular, the masculine and neuter forms of the adjectives are the same, while in the plural the feminine and the neuter use the same form. This is exemplified by Corbett (1991:150) with the masculine *bărbat* ‘man’, neuter *scaun* ‘chair’, and feminine *fată* ‘girl’, as seen in (1). Following this definition, these are clearly three different genders, since each noun must belong to only one gender.

(1) Romanian (Indo-European, Romance) (Corbett 1991:150–151)

- | | |
|---|--|
| <p>a. <i>bărbatul e bun</i>
man.the is good
‘The man is good.’</p> | <p>d. <i>bărbații sînt buni</i>
men.the are good
‘The men are good.’</p> |
| <p>b. <i>scaunul e bun</i>
chair.the is good
‘The chair is good.’</p> | <p>e. <i>scaunele sînt bune</i>
chairs.the are good
‘The chairs are good.’</p> |
| <p>c. <i>fata e bună</i>
girl.the is good
‘The girl is good.’</p> | <p>f. <i>fetele sînt bune</i>
girls.the are good
‘The girls are good.’</p> |

In some languages, there appear to be more gender distinction for some targets than for others. One such example is Swedish, which traditionally is considered as having two genders: common and neuter. However, in pronouns, and in some mostly idiomatic adjectival forms, more distinctions appear. As for pronouns, Swedish has four third person singular forms: *han* ‘he’, *hon* ‘she’, *den* ‘it’, and *det* ‘it’.² The first three are used only with nouns belonging to the common gender, where the

² More recently a gender-neutral fifth form has become popular, *hen*, similar to the singular use of they.

choice between *han* and *hon* is a matter of natural gender (as in English), whereas *den* is used for the remaining nouns of the common gender (e.g., *stol* ‘chair’, *bok* ‘book’, *hund* ‘dog’). Only *det* is thus used for an entire gender. Finally, a noun denoting a male referent may cause some definite attributive adjectives to take a different form (see (2)), by using the ending *-e* instead of *-a*, which is used by other nouns of the common gender.

(2) Swedish (Indo-European, Germanic) (constructed example)

- a. *de-n glad-e mann-en*
 DET-C happy-DEF.M man-DET.C
 ‘The happy man.’
- b. *de-n glad-a kvinna-n*
 DET-C happy-DEF woman-DET.C
 ‘The happy man.’

The natural question emerging is of course whether Swedish thus has two genders or actually four. Following the definition above, Swedish has two genders since four distinctions only appear in a very small minority of indexing targets (i.e., pronouns), whereas the masculine form of adjectives as in the example above only occurs in idiomatic or archaic usages. However, is it possible that Swedish has two gender systems, with common and neuter as one system, and masculine, feminine, ‘rest common’, and neuter as a second system? According to the definition, this is not the case either. Masculine, feminine, and *den*-nouns always belong to the common gender, whereas *det*-nouns are neuter, so that the first three clearly constitute subgroups of the common gender. If the groupings can be considered as subgroups of higher groupings, these are not considered as part of a separate gender system; thus, the three categories of masculine, feminine, and rest are analyzed as subgenders of the same common gender. This distinction can be seen in Figure 1 and Figure 2, where the relations between the discussed categories in Swedish are compared with the relations between the Mian (TNG, Ok) genders and a system of verb prefix classes (see discussion in 5.2.1). In Mian, not all nouns of the Masculine verb prefix class belong to the Male gender, since some belong to Neuter 1, whereas nouns of the Neuter 1 gender can be assigned to one of five verb prefix classes.

Genders	Common			Neuter
Subgenders	Masculine	Feminine	Rest	

Figure 1: The relation between Swedish genders and subgenders.

Verb prefix classes	Masculine	Flat	Bundle	Long	Residue	Feminine
Genders	Male	Neuter 1			Female	Neuter 2

Figure 2: Overlap between the gender and verb prefix classes of Mian (adapted from Fedden 2011:173).

In conclusion, the definition of gender used in this thesis is quite traditional. Although it is based on the more recent notion of indexation, it is compatible with definitions based on the notion of agreement. As the Examples (1) and (2) indicate, the interpretation of most gender systems remains the same as in traditional analyses, and the results of this study remain true to and comparable with modern research on gender.

2.2 Typology of gender

2.2.1 Overview

Throughout the history of research on grammatical gender, many studies have been on Eurasian or African languages (including Di Garbo 2014), but typological research has shown that gender is by no means restricted to these regions. This is shown in the three main WALS chapters on gender (Corbett 2013a; 2013b; 2013c), where Corbett compares gender systems in the languages in the world using a sample of 257 languages, half of which have no gender. The languages with gender are not evenly spread around the world; in particular, Africa is a major source of gender systems, not least since most of the 1,500 languages of the Niger-Congo family have gender (Corbett 2013a). Other important gender areas are the parts of Eurasia where Indo-European languages are spoken, the Caucasus, Northern Australia, some few parts of the Americas, and (importantly for this study) New Guinea. Thus, areas where gender is not found are the Pacific as well as most of Asia and North America.

Even though Corbett (2013a) includes New Guinea in his sample, typological research on the gender systems of New Guinea and even the region as a whole is lacking. In New Guinea, around a sixth of the world's languages are spoken. The great quantity and diversity of especially the Papuan (i.e., non-Austronesian) languages, coupled with their relative isolation (at least from non-New Guinean languages), make them extremely important for typological research. However, Corbett (1991) includes only one Papuan language in his discussion of gender (viz., Yimas), while the WALS chapters above only include 18 languages from New Guinea proper, which constitute 7% of the sample. Even the literature about New Guinea or Papuan languages lack an overview of gender; e.g., while Foley (1986) includes a chapter about gender in his seminal book on Papuan languages, only a few are discussed (including Yimas).

However, there is growing evidence that some Papuan languages have unusual gender systems. This is noted by Corbett (2013a), who states that “[in] New Guinea, several families show gender, and of widely differing types,” specifically mentioning Mountain Arapesh (Torricelli) which has 13 genders. Another example of a Papuan language with an interesting gender system is Mian (TNG, Ok), which has been documented by Fedden (2011). The Mian gender system includes four gender, masculine, feminine, and two neuters, where the masculine, feminine and neuter 1 are quite homogeneous and encode a contrast in number, whereas neuter 2 is more heterogeneous with no number contrasts. There is also syncretism in the gender and number paradigm. Another unusual gender system has been found in Nalca (TNG, Mek) (Svärd 2013; Wälchli & Svärd 2014), which has three non-default genders (masculine, feminine, and neuter) and two default genders (default noun gender, and non-noun default gender), as well as a special coordination gender on the periphery of the gender system. There is thus much potential in investigating gender and number from a typological perspective in New Guinea.

2.2.2 New Guinea

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. In linguistic literature, the languages of New Guinea are commonly divided into two groupings, viz., Austronesian and Papuan (i.e., non-Austronesian) languages, but this does not presuppose any genetic connection between the Papuan languages (Foley 2000:358).

The occurrence of gender in New Guinea is tightly correlated with the distribution of the 1,200 languages. 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)³. 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 extremely 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 must occur in at least 120 languages of mainly five different families (excluding isolates, TNG and Austronesian). Furthermore, these families show much genealogical diversity, which suggests that gender may be highly diverse in New Guinea.

However, Foley suggest that gendered languages of New Guinea have some features in common. According to him, most languages have binary gender systems with masculine and feminine, but Foley mentions Burmeso (isolate) as an exception of a language with a neuter (Foley 2000:371). Furthermore, Foley states that in all languages where it can be determined, masculine is the marked gender whereas feminine is the unmarked one. It will be argued below that Foley's characterization, while correct as a tendency, downplays the diversity of gender across New Guinea.

As far as gender assignment in New Guinean languages is concerned, Foley argues that the assignment of humans and higher animates is straightforward and based on natural gender, but nouns denoting other animates and inanimate objects are largely assigned gender based on size and shape (Foley 2000:372). This is notably a feature of the Sepik area, where languages such as Manambu (Ndu) assign large or long objects to the masculine gender, whereas small, short or round objects are feminine (Aikhenvald 2008:112, 116). An important point raised by Foley is that even in languages without a gender system, similar semantic contrasts can occur; e.g., in the TNG languages of the central highlands, tall, large or strong nominals referents, prototypically males, occur with 'stand' as a stance-verb, whereas round, small and weak referents, prototypically females, occur with 'sit' (Foley 2000:372).

There are also New Guinean languages with large gender systems (referred to as noun class systems by Foley), especially in the Sepik-Ramu basin in languages of the Torricelli and Lower Sepik-Ramu families (Corbett 2013a; Foley 2000:372). Many of these languages have structurally very similar and typologically unique gender systems, but there are no apparent genealogical links between the two

³ Foley (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).

families. These languages have many genders, such as 11 in Yimas (Sepik-Ramu) and 16 in Arapesh (Torricelli), and gender assignment is both semantic and, to a greater extent, formal (viz., phonological) (Foley 2000:372).

Finally, some languages have gender and noun classification as two separate but crosscutting systems, particularly Motuna (South Bougainville) and Burmeso (isolate) (Foley 2000:373). In Burmeso, there are two noun classes marked on verb prefixes, where male referents are assigned to class I, and female referents to class II. On the other hand, there is a gender system with three genders (masculine, feminine, and neuter), which is marked via indexing suffixes, e.g., on adjectives. Both systems are separate and can disagree: e.g., *akeway* ‘black cockatoo’ is class II, which is prototypically female, but contrary to expectation belongs to the neuter gender (Foley 2000:373). Similar systems can be found among the Sepik languages, which have separate systems of gender and numeral classifiers (Lock 2011:46).

In conclusion, Foley (2000) presents a useful overview of gender in New Guinea and highlights many interesting features, 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. 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.

3. Method and data

3.1 Sampling and material

The sampling method used in this study is what is termed a *variety sample* (Bakker 2012). This means that, rather than trying to represent the real population of languages as would be achieved by a probability sample, the variety sample is maximized (Bakker 2012). Thus, the researcher first has to define which features the sample will be based on (e.g., gender) and then construct a sample in order to achieve the largest variety of results in regard to the chosen feature. Importantly, a variety sample can omit languages which do not have the feature that is investigated. However, its major potential downside is that the researcher must have some knowledge as to where the feature may be most varied; furthermore, it is important avoiding genetic bias, since this might lead to an unrepresentative sample.

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 while still accounting for variation within families if there were reasons to do so, primarily based on the information by Foley (2000) presented in 2.2.2 above. As an example, only one North Bougainville language is included (viz., Rotokas), since it is a small family of only four languages, while Torricelli is represented by three languages since it is a larger family of around 50 languages and mentioned by Foley (2000:365) as having many and varied gender systems. Thus, despite Austronesian and TNG comprising approximately half of all New Guinean languages, only two languages of each family are included in the sample since gender is very rare in these two families.

It should be noted that the availability of material also affected the sample, since New Guinea is still afflicted by inadequate language documentation. Thus, while it was possible to decide which language families to include, the actual languages of these families were sometimes chosen by necessity, such as the Lower Sepik-Ramu family where Yimas (Lower Sepik) is the only well-described language⁴. It was decided that the sample should consist of approximately 20 languages (which became the final number), since it would be difficult to satisfactorily balance a larger sample. Nevertheless, it is possible that the sample consists of as many as one sixth of the approximately 120 expected gendered languages in New Guinea (see 2.2.2).

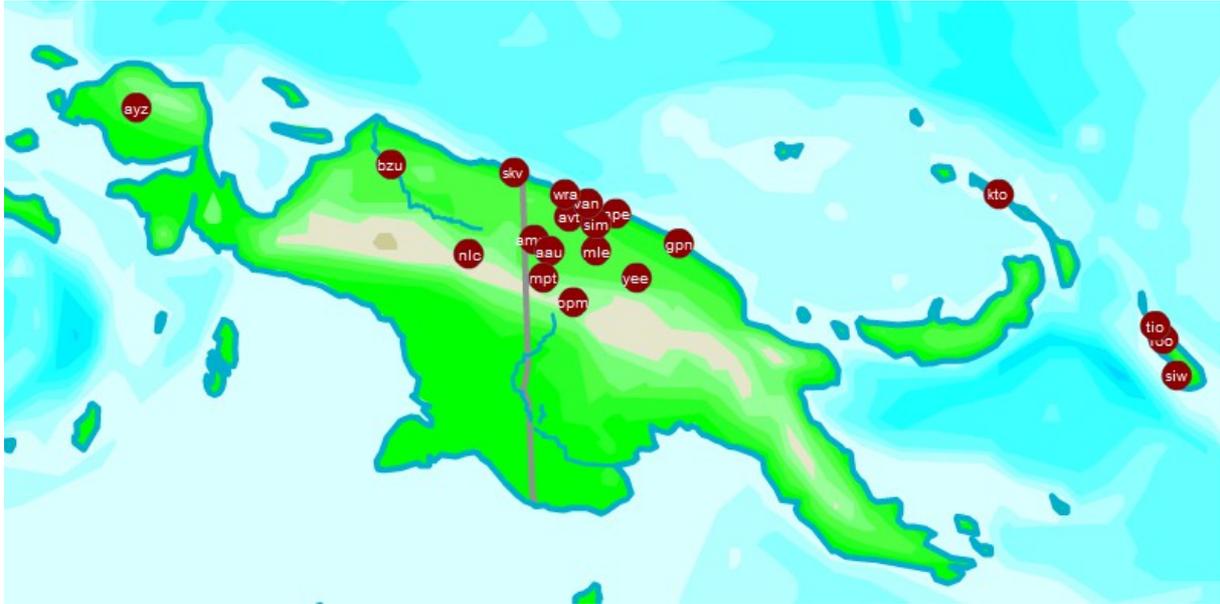
Table 2 (repeated in Appendix A) lists the languages of the sample together with family, genus, ISO code, and source, together with a map of the languages shown in Map 1. The names primarily follow Glottolog, except for Motuna (Glottolog Siwai), where I follow Onishi (1994). Also, the Glottolog form Warapu is used despite Barupu occurring in Corris (2008). 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.

⁴ Also Kire (Ramu) (Stanhope 1972), which unfortunately for this study has no gender system (see 2.1.2).

Table 2: 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 code	Language	Source
Austronesian, Oceanic	Nehan-North Bougainville	tio	Teop	Mosel & Spriggs (2000)
Isolate	–	gpn	Taiap	Kulick and 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 (2008)
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 & Svärd (2014)
	Ok-Oksapmin	mpt	Mian	Fedden (2011)
		opm	Oksapmin	Loughnane (2009)
West Papuan ⁵	–	ayz	Maybrat	Dol (2007)

⁵ The traditional West Papuan Phylum has been shown in more recent studies to probably not be an accurate genealogical grouping, but instead constituting 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.



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

The main sources of data used in this thesis are reference grammars, which are listed for each language in Table 2 above. Since gender is such a common and immediately noticeable phenomenon, it was assumed that if a language had gender, it would at least be mentioned in the reference grammar as either gender or noun class systems. However, many descriptions do not mention the language as having a gender system if gender only occurs in pronouns, so 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 as eligible for the sample. However, even if a language had available descriptions, these were in some cases practically unusable due to reliance on obsolete or highly specialized theories. This is in fact mentioned by Bakker (2012) as a problem in language sampling:

Yet another complication occurs when grammars have been written strictly from the perspective of some theory, most notably Tagmemics or one of the several instantiations of Generative Grammar. For most linguists working in the typological tradition, and even for many linguists working on more upto-date [sic] versions of the theories mentioned, such material is close to inaccessible since the original raw language data are interpreted in terms of theoretical notions such as transformations. So, even if descriptive material is present, it is not always of much use for typological investigations. As a result, the corresponding languages will have to be absent from most samples. (Bakker 2012)

Although the inaccessibility of some grammars is unfortunate, this was only a minor issue in this thesis. In conclusion, the choice of material as well as the problems therein are not exclusive to this study.

3.2 Procedure

The first step of this study was to create a compilation of sketches of the gender systems of the sample. In reference grammars, gender is most often described in multiple sections (e.g., sections on nouns, verbs, pronouns, demonstratives, articles, syntax, or gender). Thus, it was indispensable to create some kind of overview. These descriptions were intended to account for the gender systems of the languages

as far as documented, meaning that they often included more information than necessary for the study, which made them quite bulky (about fifty pages in total). Some examples of these descriptions are included in Appendix C. In order to facilitate the comparison, a digital spreadsheet was used to compile the necessary information from the already created descriptions. A refined and presentable version of this spreadsheet is shown in Appendix B.

The second step consisted of deciding how to make the gender systems in New Guinea typologically comparable. In order to achieve this, the study employs the five classificatory criteria used by Di Garbo (2014) to classify the gender systems of Africa. An important advantage of adopting Di Garbo's approach is that this makes the results for New Guinea directly comparable with Africa (see 5.1.2). In addition, since the first three criteria are the same as the ones used in the WALS chapters by Corbett (2013a; 2013b; 2013c), much of the results is comparable to a worldwide sample (see 5.1.1). In order to illustrate the distributions, maps were created using the Interactive Reference Tool of the World Atlas of Language Structures (WALS)⁶ using ISO codes and coordinates from Glottolog. The five criteria are as follows:

1. *Sex-based and non-sex-based gender systems.* For this criterion, each language is assigned one of the following two values: sex-based or non-sex-based. The distinction differentiates between gender systems where the semantic core is based on natural gender, while non-sex-based gender systems are those where the semantic core is based on animacy.
2. *Number of genders.* This is the number of genders, where each language is given the value point of either two, three, four, or five or more genders.
3. *Gender assignment.* Depending on the patterns of gender assignment, each language is assigned either the value of semantic assignment, or semantic and formal assignment.
4. *Number of gender-indexing targets.* This criterion concerns the number of gender-indexing targets, for which each language is assigned the value of either one, two, three, or four or more indexing targets.
5. *Occurrence of gender marking on nouns.* The final criterion concerns whether gender is overt or covers, and as such the values are either yes or no.

The final step was the actual comparison between the results with those of Africa (Di Garbo 2014) and the world as a whole (Corbett 2013a; 2013a; 2013b). The reason for comparison with Africa were partly because of the presence of the same classification criteria in Di Garbo (2014), but mostly because of the already suspected differences between the two regions. The comparison was a straightforward process and consisted mainly of comparing numbers. However, it also became evident that the results contradicted some claims e.g., by Foley (2000). Because of this, it also became important to further investigate other aspects of the gender systems, such as the presence of two separate gender systems in a language or special patterns in gender assignment based on shape and size. This discussion is found at the end of Chapter 5.

3.3 Limitations

There are two main limitations of this study, both based on the sample. Firstly, the sample is small, and even though the number of gender languages in New Guinea is estimated to be at least 120 languages, this is only a very vaguely defined minimum. Thus, until more documentation is done on the

⁶ See <http://www.eva.mpg.de/lingua/research/tool.php>.

languages of New Guinea, it remains unknown if the 20 languages in the sample constitute one sixth of the gendered languages of New Guinea, or if there are many more unknown gender system unknown to the present author.

Secondly, the results of this study are affected by the use of a variety sample. While this has some important advantages, the results of this study neither inform us about those languages where gender does not occur, nor can it in any way be used for prediction. Since the sample has been devised to result in as much variety as possible, it emphatically does not represent the real distribution of these features in New Guinea. In order to do this, a proportional sample would have to be used, in which TNG and Austronesian languages would of course contribute half of the languages. While this would be interesting, note that this is decidedly not the aim of the present study.

4. Results

In order to gain an overview of the gender systems in the sample, each language has been classified according to the five criteria used by Di Garbo (2014) to classify the gender systems of African language (see 3.2 above) In the following sections, the distribution of values of the criteria 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 Chapter 5, 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 5.2.2, 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 will be given in parenthesis whenever applicable.

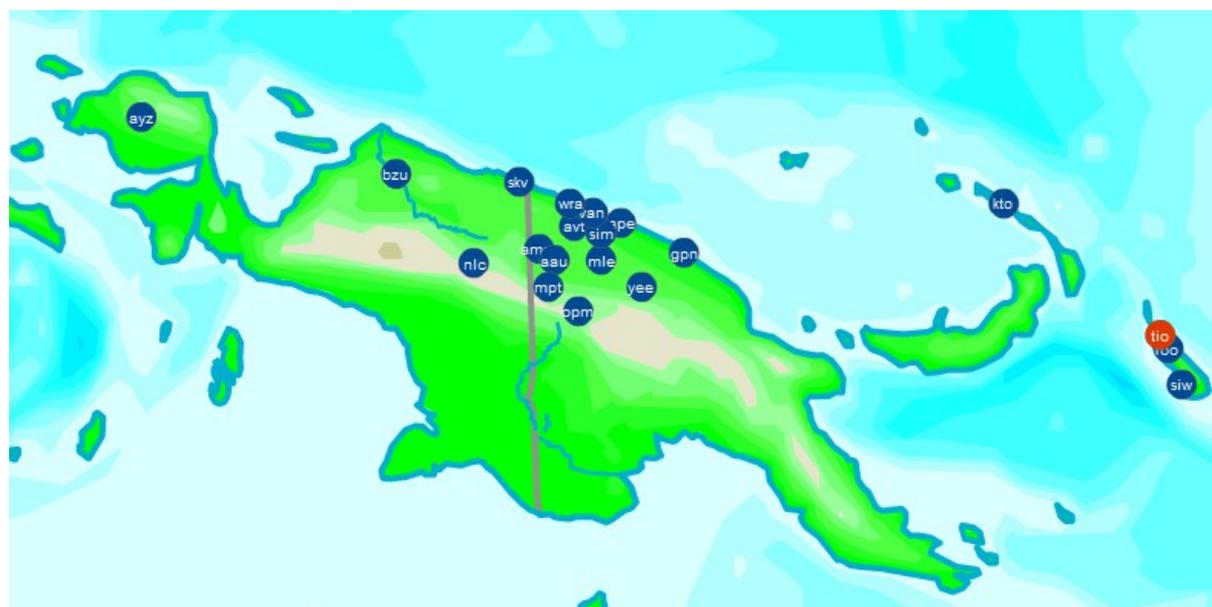
4.1 Sex-based and non-sex-based gender systems

Following Di Garbo (2014:62), each gender system is classified as either sex-based or non-sex-based based on Corbett (2013b). 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 (2013b), all non-sex-based systems are based on some notion of animacy.

As shown in Table 3 and Map 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 sole Austronesian language Teop exhibits a non-sex-based systems.

Table 3: Sex-based and non-sex-based gender systems in the sample.

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



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

Sex-based gender systems present some difficulty in assigning nouns denoting inanimate referents. Non-sex based systems, i.e., systems based on animacy, can potentially assign every noun based on only animacy. However, sex-based systems do not by definition have 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 4.2, almost half of the languages in the sample (9 of 20) have only two genders, which in all cases are sex-based. Thus, since option 3 is only possible in languages with more than two genders, almost half of the languages in the sample have some means of assigning 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 only visible 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 (4), where Max (male name) (3a) and Lusi (female name) (3b) occur with the masculine and feminine pronoun forms respectively, and the inanimate *masiji* ‘hair’ (3c) is referred to with the feminine form.

(3) Mende (Sepik) (Hoel et al. 1994:19, 31, 46 respectively)

a. *Max wasilaka ri-a*

M. big 3SG.M-INTEN

‘Max is big.’

b. *Lusi kava awu-n u-nda sir-a*

L. bad⁷ 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.’

Assigning inanimates to both sex-based genders either 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 5.2.1 below). One such 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 (4); when referring to the tree from which he makes the paddle (4a), *youk* ‘paddle’ is masculine, since the tree is long and not at all round or flat. However, when referring to the actual paddle (4a), which has the salient features of flat and round, the feminine form is used.

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

a. *Ha-kwe youk se seyr.*

1SG.SBJ-TOP paddle 3SG.M.OBJ cut

‘I cut the “paddle” tree.’

⁷ When used with the habitual *-nda*, *kava* ‘bad’ functions as an intensifier (Hoel et al. 1994:31).

- 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 which such a system is Nalca (TNG, Mek) (Svärd 2013; Wälchli & Svärd 2014). Nalca has five main genders: masculine, feminine, neuter, default, and non-noun. As shown in (5), 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 the despite its name the nominalizer *-a*, but also when gender is switched off⁸. In the examples below, both the neuter *si* ‘name’ and the masculine name *Zakheus* ‘Zacchaeus’ are shown in (5a), the feminine *genong* ‘mother’ in (5b), the default (DEFAULT) *pik* ‘way’ in (5c), and the two non-noun (NNOUN) constructions in (5c). The first non-noun gender is due to the intervention of the quantifier *nauba* ‘many’ between *nimi* ‘men’, which belongs to default gender, and the case marker host, whereas the second is due to the nominalizer *-a*.

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

- a. *alja si ne-ra Zakheus be-k ulu-m-ok*
 3SG.GEN name N-TOP Z. M-ABS be-PFV-PST.3SG
 ‘a man called by name Zacchaeus’ (Lk 19:2)⁹
 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)

⁸ The concept of switching gender on and off is an extremely rare phenomenon and goes well beyond the bounds of this thesis. For a comprehensive description of the Nalca gender system and discussion on switching gender on and off, see Wälchli & Svärd (2014).

⁹ The 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 & Svärd (2014).

- c. *Na bi-nim-na pik e-ra ugun-da ella u-lu-lum...*
 1SG go-FUT-PRS.1SG way **DEFAULT-TOP** 2PL-TOP knowledge 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’ a-k eib-ok*
 men many **NNOUN-TOP** PRF assemble-PFV-PST.3PL-NMLZ **NNOUN-ABS** see-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 systems in the sample occurs in the Austronesian language Teop, which has two genders (I and II) with two subgenders 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:

- *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 2000:336–8).
- *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 the noun classification 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 (see discussion on this in 4.5). 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).

Thus, even though Teop has the only non-sex-based system in the sample, the similar system of nominal classification in Siar indicates that this is an Austronesian characteristic.

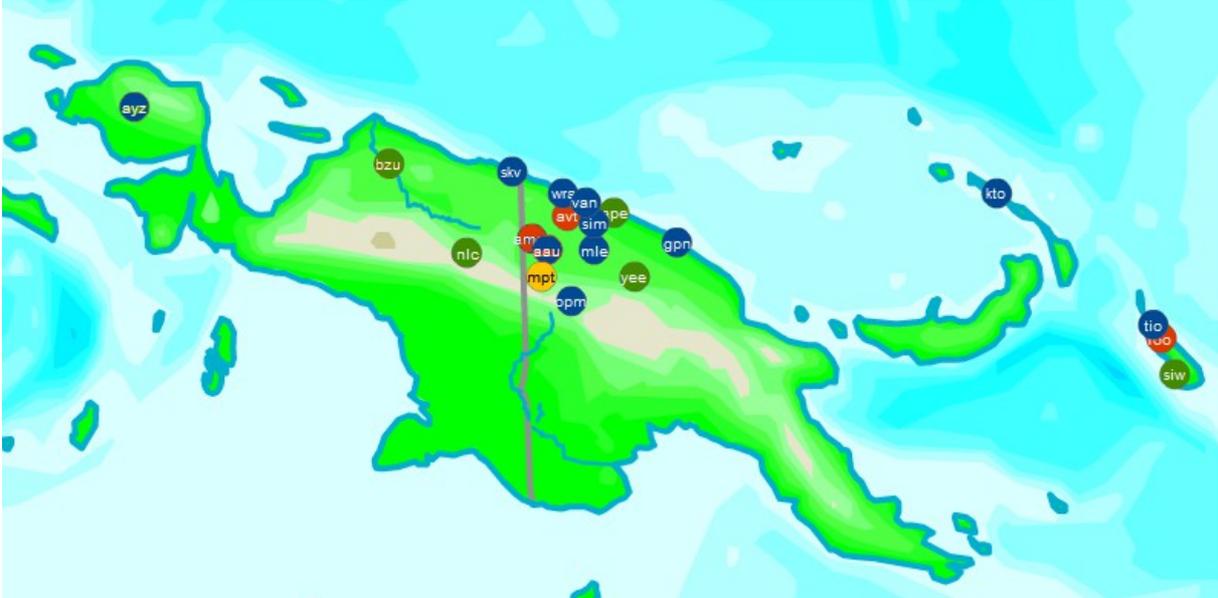
4.2 Number of genders

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 4 and Map 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., Bukiyip (18 genders), Burmeso (9 > 3/6)¹⁰, Motuna (6), Nalca (5), and Yimas (around 12).

Table 4: Number of genders in the languages of the sample.

Number of genders	No. of lgs.	%	Languages (no. of genders)
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%	Bukiyip (18) Burmeso (9 > 3/6) Motuna (6) Nalca (5) Yimas (~12)
Total:	20	100%	



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

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

¹⁰Burmeso has two gender systems, with three genders belonging to the first system and the other six belonging to the second system (see 5.2.2).

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

Most languages do not belong to either of these types and cannot meaningfully be compared in this regard. Therefore, the following paragraphs focus on describing the features of the above three types.

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 (6): in (6a) ‘old’ indexes ‘his father’, in (6b) ‘his mother’, and in (6c) ‘big’ indexes ‘house’.

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

- a. *y-atia* *y-anes*
 3M-father 3M-old
 ‘His father is old.’/‘his old father’
- 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 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. E.g., in Rotokas (North Bougainville), inanimate objects associated with male culture (such as hunting or warfare) and long, thin objects are masculine (see also 5.2.1), whereas most inanimates are assigned to either the feminine or 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, including the others with five or more genders. The immediately noticeable 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 gender or fewer. A table of the Bukiyip genders and their indexing forms are shown in Table 12 in 4.5. The most important features of these two gender systems is that both have semantic-formal agreement 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 be discussed further in 5.2.2.

4.3 Gender assignment

The third criterion concerns gender assignment and contains three values (see Table 5 and Map 4):

1. *Transparent semantic*: The gender assignment rules are exclusively semantic and transparent. Gender would in many cases not have to be indicated in the dictionary of such a language.
2. *Semantic and formal*: The system of gender assignment contains at least one rule pertaining to formal criteria.
3. *Opaque*: Gender does not follow transparent semantic criteria, nor is it identifiably based on any formal criteria. In some cases it is probable that gender is based on e.g., phonology, although most languages probably have arbitrary and/or semantically complex rules. Note that even a system for which the rules can be explicitly stated can be opaque, such as for German (Zubin & Köpcke 1986). Gender in such a language must be included in a dictionary.

The distinction between transparent semantic and opaque systems is not present in either Di Garbo (2014:66–67) or Corbett (2013c). However, a comparison between a transparent semantic system such as Mende (Sepik) and an opaque semantic system such as Rotokas (North Bougainville) shows that such a distinction is helpful. As mentioned in 4.1 above, Mende 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.

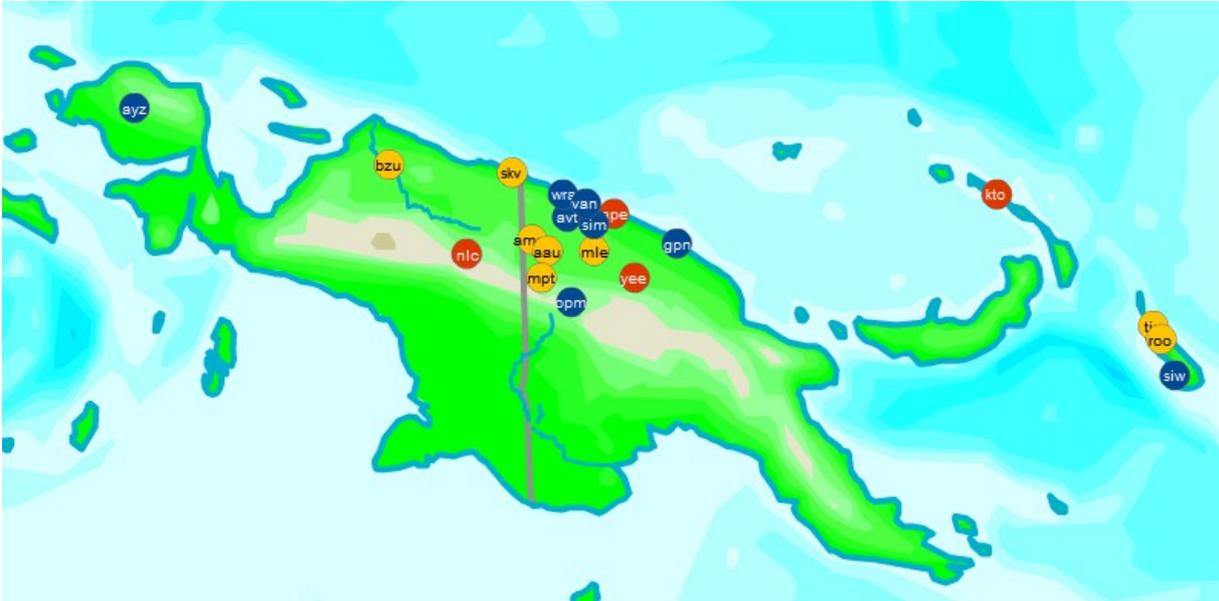
In Rotokas however, the situation is more complex. Rotokas has three genders: masculine, feminine, and neuter. Both the masculine and feminine gender contain nouns denoting male and female referents respectively, but complexity arises for inanimates. The masculine gender contains many inanimate objects, which are often associated with male culture of 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, many inanimate nouns belong as expected to the neuter gender (Robinson 2011:48).

Thus, while a learner of Mende would be able to easily guess the correct gender of any noun, a learner of Rotokas would be hard-pressed to guess the correct gender of an inanimate object. This is very similar to the situation found in many European languages, such as German or French, where second language learners regularly fight with the seemingly arbitrary rules of gender assignment (such as the German names for the three main pieces of cutlery, *Löffel* ‘spoon’, *Gabel* ‘fork’, and *Messer* ‘knife’, which are masculine, neuter, and feminine respectively). Even if there are rules, many of these are not learnable and 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 non-general, have numerous exceptions, or follow conflicting semantics. 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 size¹¹.

Table 5: Systems of gender assignment in the sample.

Gender assignment	No. of lgs.	%	Language
Transparent semantic	8	40%	Au ¹²
			Maybrat
			Mende
			Motuna
			Oksapmin
			Taiap
			Walman
			Warapu
Semantic and formal	4	20%	Bukiyip
			Kuot
			Nalca
			Yimas
Opaque	8	40%	Abau
			Ama
			Burmeso
			Manambu
			Mian
			Rotokas
			Skou
			Teop
Total:	20	100%	



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

¹¹ It is of course possible to imagine various explanations for why insects are not feminine, e.g., perhaps are they 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.

¹² It 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.

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 exemplified in (3) in 4.1. However, semantic systems can be more complex while still retaining transparent semantic criteria. 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 are shown in Table 6.

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>

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 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 above, many languages have much more complex systems. If they are neither semantically transparent nor contain any apparent formal criteria, they are 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 5.2.1). Manambu has two genders, masculine and feminine, and in general gender assignment appears to follow semantic criteria. However, these are far from transparent:

1. Human are assigned based on their natural gender, except nouns denoting small children, which can be assigned gender based on size (Aikhenvald 2008:116–117).
2. Higher animates are assigned 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 it; 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, *væy* ‘spear’ is masculine, since it is large, but it is feminine if referring to small spears or shutguns (Aikhenvald 2008:117).

5. Natural phenomena are assigned gender based on whether they are complete or not: if they are uncompleted 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. 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 *nabi* ‘year’, which is masculine due to it being very long (Aikhenvald 2008:119).

There are in fact further assignment rules, but the point has been made; i.e., the rules of gender assignment are not semantically transparent. Especially important to note that it is difficult to ascertain whether they are rules or merely patterns. That is not to belittle the observations or claim that the researcher, in this case Aikhenvald, has done anything wrong; instead it illustrates the difference between transparent semantic systems, where all gender assignment rules are easily identifiable and apply to all nouns, and opaque systems, where patterns most certainly can be found but in which exceptions are abound.

While it is easy to become amused by the seemingly 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-d-ə ta:kw* (DEM.PROX-M.SG big-M.SG woman) ‘this (unusually) big woman’. (Aikhenvald 2008:121)

The last type is the languages with semantic and formal, of which there are four: Nalca is skewed towards semantic assignment, Kuot favors both semantic and formal assignment roughly equally, and Bukiyip and Yimas favor formal assignment. E.g., among the five genders in Nalca (see 4.1 above), only the neuter is formal, but very much so since it contains only (but all) nouns of the phonological structure (C)V (Wälchli & Svärd 2014). In comparison, only three of the 18 genders in Bukiyip (see Table 12) 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).

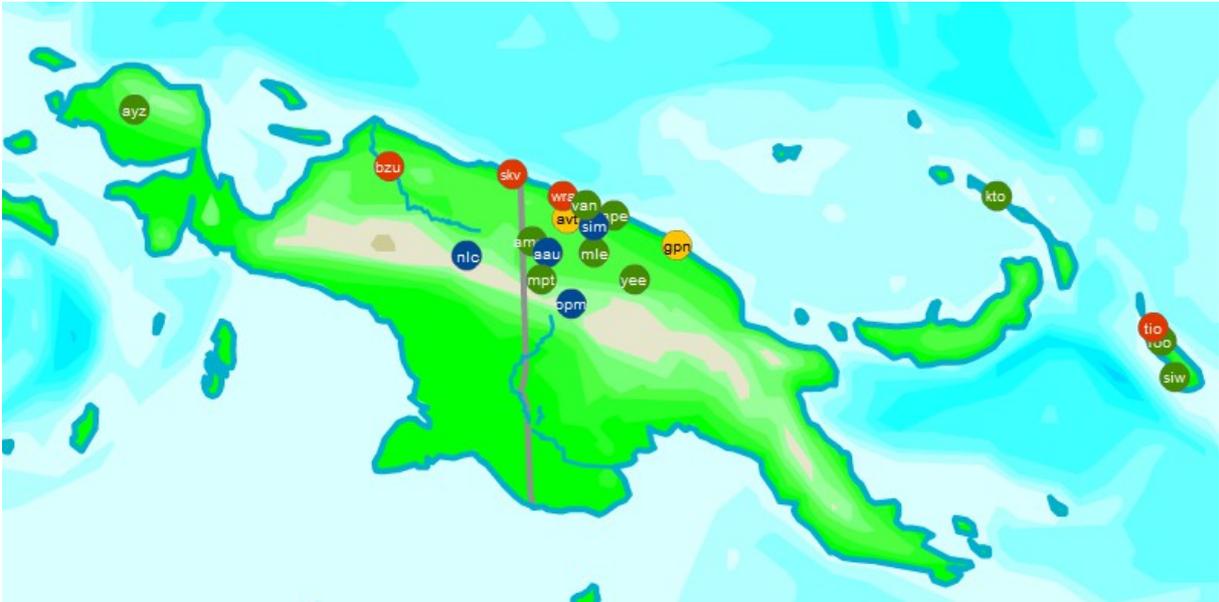
4.4 Number of gender-indexing targets

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 Map 6, while the 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: adjectives, demonstratives, verbs, numerals, copulas, complementizers, and

adpositions. In this thesis, pronouns were also considered a basic category and differentiated from demonstratives unless they formed part of the same paradigm. However, no detailed analysis was done on 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%	Barapu Burmeso Skou Teop
Three	2	10%	Au Taiap
Four or more	10	50%	Abau Bukiyip Kuot Maybrat Manambuu Mian Motuna Rotokas Walman Yimas
Total:	20	100%	



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

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

Language	Pronouns ¹³	Verbs	Demonstratives	Adjectives	Numerals	Prepositions	Case marker hosts
Abau	X		X		X		X
Ama		X					
Au	X	X		X			
Bukiyip	X	X	X	X			
Burmeso		X ₂ ¹⁴		X ₁ ¹⁴			
Kuot	X ¹⁵	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 ¹⁶	X	X	X	X		

As the tables above show, more than half of the languages in the sample have more than four gender-indexing targets. There are also some interesting patterns to be found in Table 8:

- 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.

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.

¹⁴ Burmeso adjectives are targets in the first gender system whereas verbs are targets in the second system.

¹⁵ Kuot has no independent third person personal pronouns (Eva Lindström, p.c.). However, demonstratives are used with pronominal functions (see also Footnote 15 below)

¹⁶ '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). Because of this, these forms are considered pronouns for comparative purposes.

- 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 hierarchies, where the leftmost is a the most typical target while the rightmost is less common. It also implies that 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:

1. pronouns > verbs > demonstratives > adjectives > numerals (14/20)
2. verbs > adjectives > pronouns (3/20)
3. other (3/20)

The first hierarchy is common, accounting for 70% of the languages. The remaining six languages do not show such convincing pattern, but three of them appear to follow the second hierarchy. The last three languages are all wholly different from each other. The distribution is shown in Table 9.

Table 9: Distribution of gender-indexing hierarchies in the languages

Gender-indexing hierarchy	No. of lgs.	%	Languages
pronouns > verbs > demonstratives > adjectives > numerals	14	70%	Bukiyip Kuot Manambu Maybrat Mende Mian Motuna Oksapmin Rotokas Skou Taiap Walman Warapu Yimas
verbs > adjectives > pronouns	3	15%	Ama Au Burmese ¹⁷
other	3	15%	Abau Nalca Teop
Total:	20	100%	

The languages following the first two hierarchies do not require much explanation. However, note that they are similar, except the fact that pronouns are leftmost in the first hierarchy and rightmost in the second. Since the other indexing targets are unaccounted for in the second hierarchy, it is unknown where they would appear. However, it is important to note that since the sample is small, the second hierarchy may not represent a common pattern but only a haphazard coincidence in these language. Nevertheless, the first pattern does seem to be significantly common and comprehensive.

It is also interesting to note that among the ten languages with four or more indexing target, all except Abau belongs to the first hierarchy. There is therefore an additional pattern, whereby a gender system

¹⁷*Burmese poses some problems since adjectives and verbs are targets in two separate systems. However, since one of them as well as the language as a whole follow the second hierarchy, gender in Burmese is considered to follow this hierarchy while still acknowledging the difficulties of the analysis.*

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 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 the first two are obligatory. Some of the most common forms are shown in Table 10. Examples were given in (5) in 4.1 above, the first of which is repeated in (7).

Table 10: Some of the most of most frequent forms of case marker hosts words in Nalca.

	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>edya(')</i>	<i>adya(')</i>
Gen./erg. dem.	<i>benedyā</i>	<i>genedyā</i>	<i>nenedyā</i>	<i>enedyā</i>	<i>anadyā</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>

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

alja si ne-ra Zakheus be-k ulu-m-ok
 3SG.GEN name N-TOP Z. M-ABS be-PFV-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 example of markers preceding adjectives and numerals, respectively, are shown in (8).

(8) Teop (Austronesian, Oceanic) (Mosel & Spriggs 2000:330, 328, respectively)

- 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’

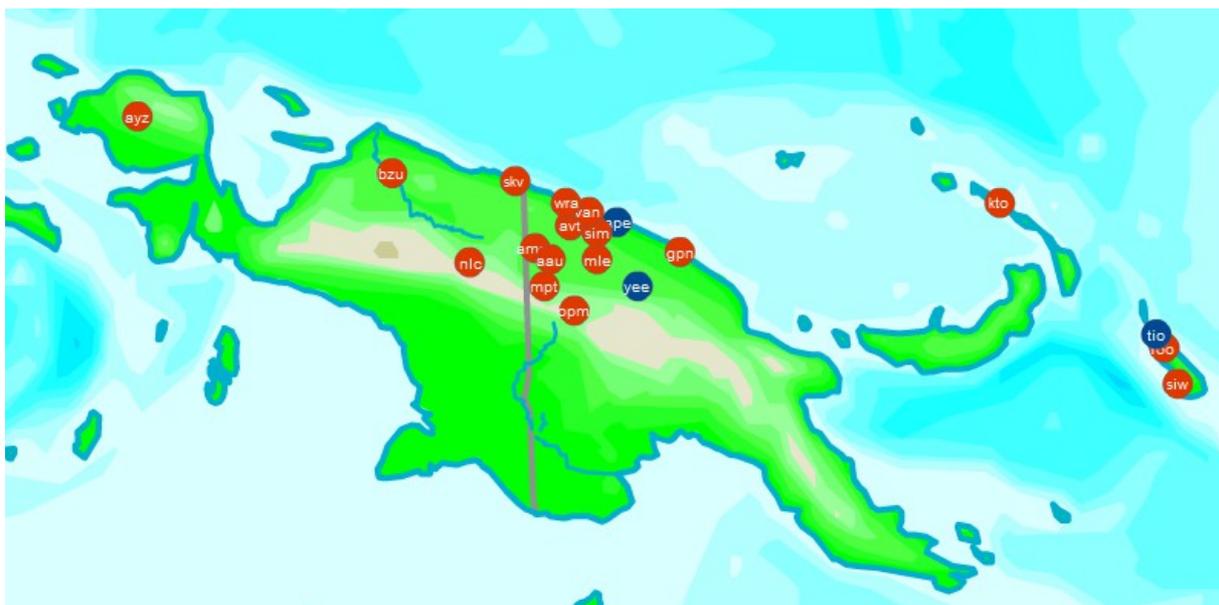
However, since these do not carry any other functional load, they do not satisfy the criterion that an indexing target must express something other than gender and number. Instead, Teop is analyzed as having two targets, viz., adjectives and numerals, which form a unit with the preceding article. On the other hand, the articles preceding nouns are analyzed as overt gender marking (see 4.5).

4.5 Occurrence of gender marking on nouns

The final criterion concerns the occurrence of gender marking on nouns (see Table 11 and Map 6), following Di Garbo (2014:69). Gender marking on noun is of course not considered indexation, but it is a common feature e.g., in African languages and most certainly an essential part of many gender systems.

Table 11: Occurrence of gender marking on nouns in the sample.

Gender marking on nouns	No. of lgs.	%	Language
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%	



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

Most languages of the sample (17 of 20) do not have obligatory 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 12. 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: many genders and morphological gender assignment must prevail more easily when noun classes are overtly distinct.

Table 12: Bukiyip noun classes and noun class suffixes (adapted from Conrad & Wogiga 1991:10).

Noun 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>yawihás</i>		<i>-has</i>	<i>-has</i>
17	personal names			-	-
18	place names			-	<i>-gún</i>

On the other hand, Teop (Austronesian, Oceanic) has a very different kind of marking. As mentioned above, Teop has a set of articles which 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 13.

Table 13: 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>

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 (9) and (10), respectively.

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

a inu a rutaa
 ART.I house ART.I small
 ‘the small house / the house is small’

- (10) 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 or at least something resembling it in some cases. 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 (11a), *aveke* ‘stone’ has a feminine singular suffix, but in (11b) it remains unmarked.

- (11) Rotokas (Robinson 2011:42)

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

5. Discussion

This chapter consists of three sections. The first section discusses the distribution of values of the classification criteria in the present study in a typological perspective, compared firstly with the languages of the world (Corbett 2013a; 2013b; 2013c) and secondly with Africa (Di Garbo 2014). It aims to answer the second research question (repeated from 1.1):

2. How do the gender systems of New Guinea compare with other geographical areas and the world?

The second section discusses particular characteristics found in the gender systems of New Guinea, such as peculiar size and shape distinctions, and two separate systems of nominal classification. It also discusses how these features, in combination with the results, relate to previous research on gender in New Guinea. Thus, the second sections aims to answer the third research question (repeated from 1.1):

3. Are there any phenomena in gender which are unique to or surprisingly common in the languages of New Guinea?

Finally, the third section consists of suggestions for further research. This section also highlights the limitations of the study.

5.1 Typological comparison

5.1.1 The World

This section compares the results of this study with the three WALS chapters on gender by Corbett (2013a; 2013b; 2013c). These three WALS chapters correspond to the first three classification criteria of this thesis; unfortunately, the remaining two have no corresponding WALS data. However, the samples are of different types: Corbett uses a proportional sample (of 257 languages), whereas this study uses a variety sample (see 3.1). The consequence of this is that comparing the actual percentages is not without problems, since the numbers for New Guinea do not represent the real distribution of the values. However, the sample is still somewhat representative, since it includes languages from different families, and it does indicate which features are widespread and which are not. To make the data comparable, languages without gender have been omitted from Corbett's sample in this section, leaving 112 languages.

Classification criterion 1: Sex-based and non-sex-based gender systems (4.1). 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 Corbett's (2013b) sample, 84 languages (75%) have sex-based systems and 28 (25%) non-sex-based. A comparison of the percentual distributions is shown in Figure 3.

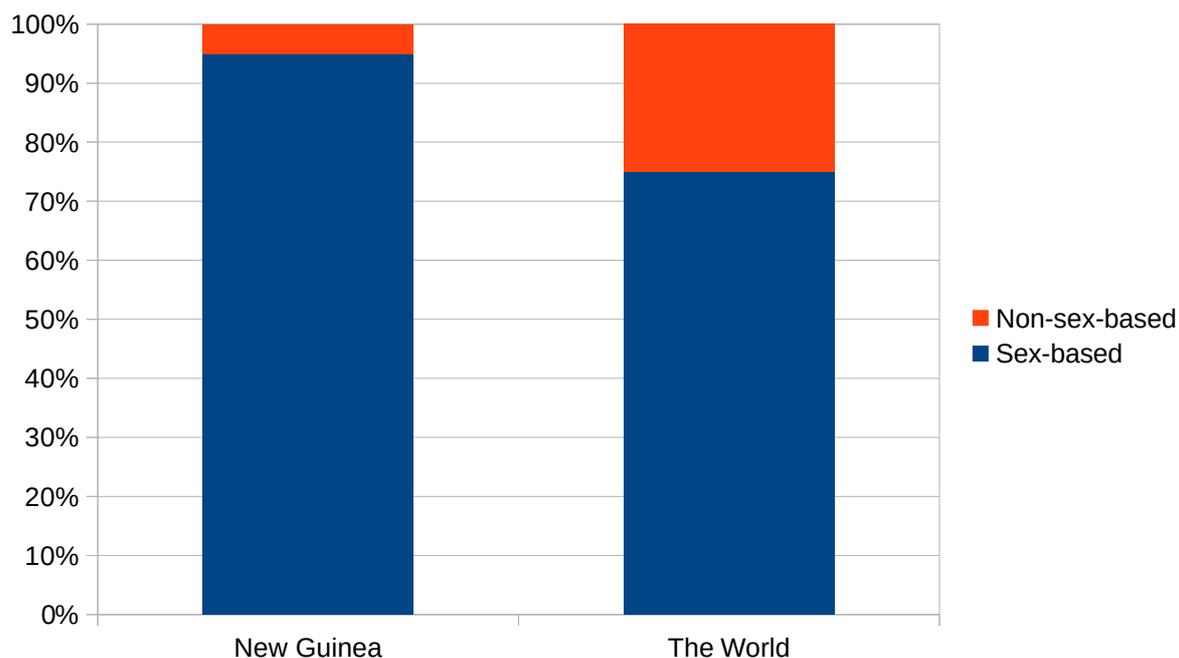


Figure 3: Sex-based and non-sex-based systems in New Guinea vs. the world.

Sex-based systems are more common in both 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. An interesting discussion about the differences between and sex-based and non-sex-based systems is presented by Luraghi (2011), who argues that they have different diachronical origins, with non-sex-based systems originating from the grammaticalization of classifiers and sex-based systems from agreement from different morphosyntactic behaviors of groups of nouns. Since sex-based systems are more common, it is thus not surprising that they are the primary ones in New Guinea, nor is likely a coincidence that the only non-sex-based gender system is found in an Austronesian language, a family remarkably devoid of gender but abound with classifiers.

Classification criterion 2: Number of genders (4.2). 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 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 percentual distributions is shown in Figure 4.

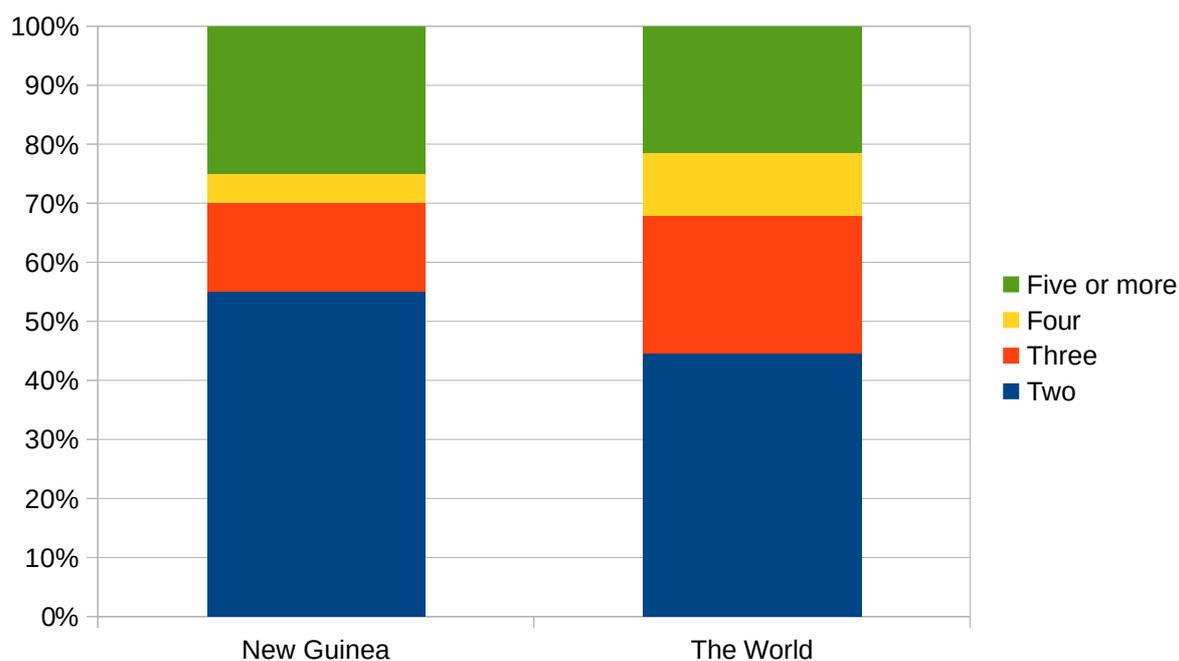


Figure 4: Number of genders in New Guinea vs. the world.

The distributions of New Guinea and the world overlap to a great extent. The major difference is that four-gender systems are less common in New Guinea, whereas two-gendered systems are less common in the world as a whole. However, the distributions should not be compared in detail; e.g., since only one language in the sample of this study has four genders, it is impossible to judge the actual distributions of such systems in the population. Despite this, it appears that both New Guinea and the world follow the same pattern, where the rate of occurrence of the values is two > three ≥ five or more > four.

Classification criterion 3: Gender assignment (4.3). This criterion is less straightforward to compare, since this study uses three values (transparent semantic, semantic and formal, and opaque), whereas Corbett (2013c) uses only two (semantic, and semantic and formal). For the purpose of this comparison, the languages of the transparent semantic and 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 (Gunwinygic; 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 Corbett's (2013c) sample, 53 languages (47%) exhibit semantic assignment, and 59 languages (53%) semantic and formal assignment. A comparison between the percentual distributions is shown in Figure 5.

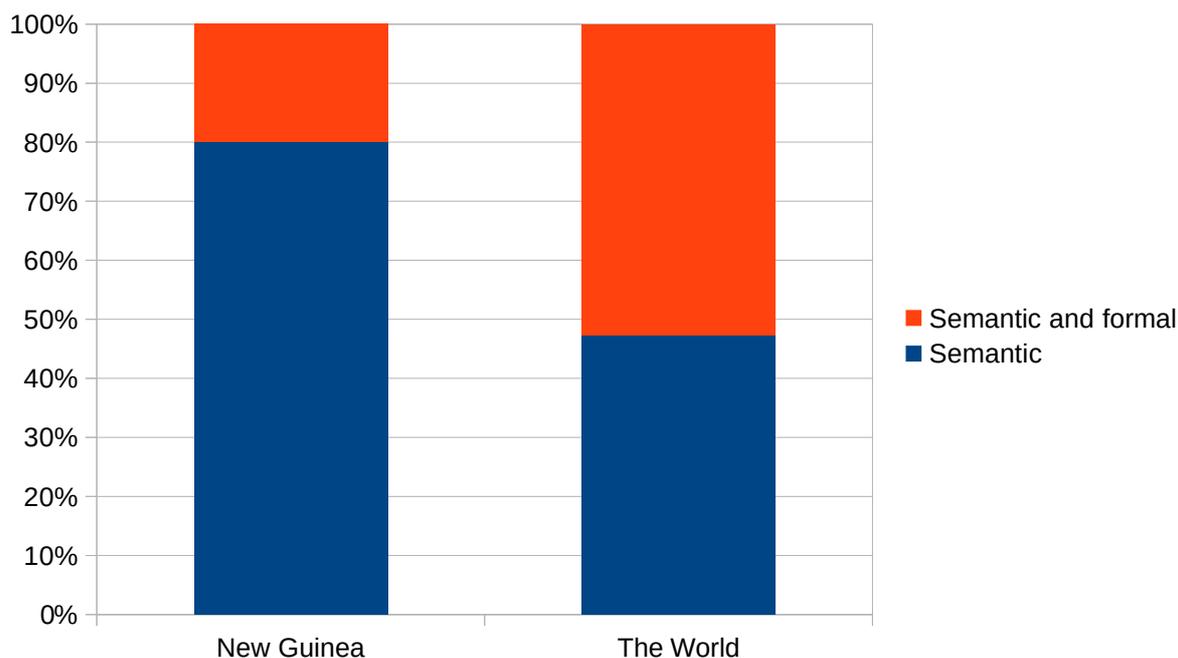


Figure 5: Gender assignment in New Guinea vs. the world.

There is a major discrepancy between New Guinea and the world as a whole. In New Guinea, semantic assignment is by far more common, whereas the ratio in the world is more or less equal. 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 therefore not surprising that semantic and formal assignment appears more common in the world than in New Guinea, since no family is represented with more than three members. 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.

To summarize, the above data and figures show that the distribution of values of the three classification criteria are rather similar in New Guinea and the world. In fact, most of the smaller differences can probably be accounted for by sample size. The only major difference is the distribution of the two systems of gender assignment; however, as discussed above, this is largely because of the prevalence of semantic and formal agreement e.g., in the Niger-Congo languages. This, among other observations, will become clearer in the following section comparing the sample of this study with Africa (Di Garbo 2014). 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.

5.1.2 Africa

This section compares the results of this study with the Di Garbo's (2014) study on gender in Africa. Both this thesis and Di Garbo's use the same five classification criteria. Furthermore, both studies employ variety samples, although Di Garbo makes a few changes, most important of which is the inclusion of non-gendered languages. Di Garbo's sample consists of 100 languages, 16 of which have no gender systems, rendering 84 of the languages comparable with the sample of this study.

Classification criterion 1: Sex-based and non-sex-based gender systems (4.1). In the sample of this study, sex-based systems are by far more common, with only Teop 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. A comparison of the between the percentual distributions is shown in Figure 6.

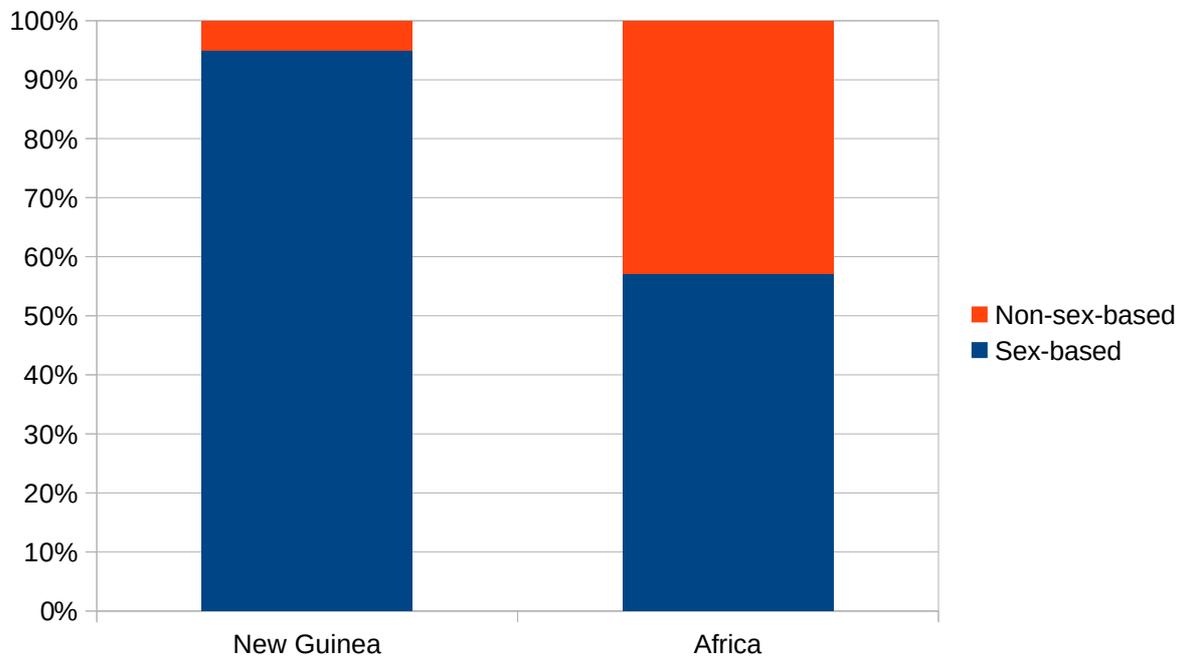


Figure 6: Sex-based and non-sex-based gender systems in New Guinea vs. Africa.

Non-sex-based systems are much more common in Africa than in New Guinea, accounting for slightly less than half of all languages. Interestingly, 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. It therefore also 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.

Classification criterion 2: Number of genders (4.2). In the sample of this study, eleven languages (55%) have only two genders, three languages (20%) three genders, one language (Mian; TNG, Ok-Oksapmin) (5%) four genders, and the final five languages (20%) five genders or more. In Di Garbo's (2014:65) sample, 42 languages (50%) have only two genders, seven languages (8%) three genders, one (Ju|'hoan; Kxa) (1%) four genders, and the final 34 languages (40%) five genders or more. A comparison between the percentual distributions is shown in Figure 7.

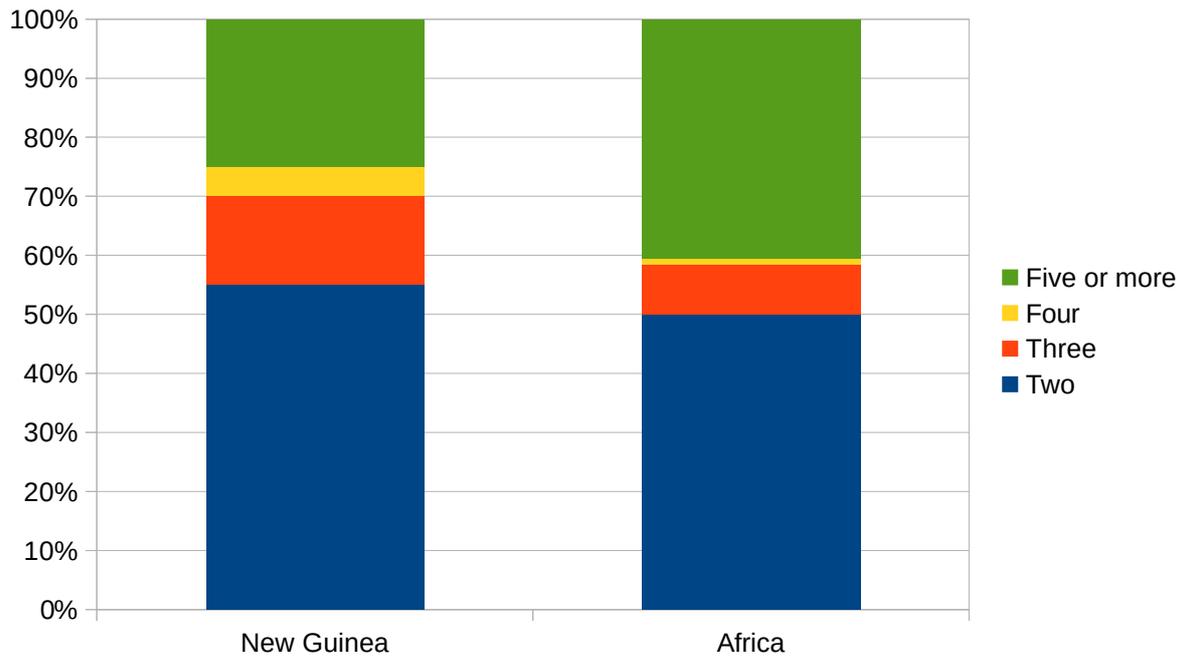


Figure 7: Number of genders in New Guinea vs. Africa.

In both samples, two-gender systems are present in around half of the languages. Furthermore, four-gender systems are apparently very uncommon in both regions. However, this is where the similarities end. In Africa, large systems are much more common than in New Guinea, leaving three-gendered systems a minor group. In comparison, systems with three or five and more genders are equally common in the sample of this study.

Once again, however, this may 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 (4.3). As in the comparison with Corbett's (2013c) sample (see 5.1.1 above), the values of transparent semantic and opaque assignment have been merged as semantic. Thus, in the sample of this study, 16 languages (80%) exhibit semantic gender assignment, whereas only four languages (20%) show semantic and formal assignment. 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. A comparison between the percentual distributions is shown in Figure 8.

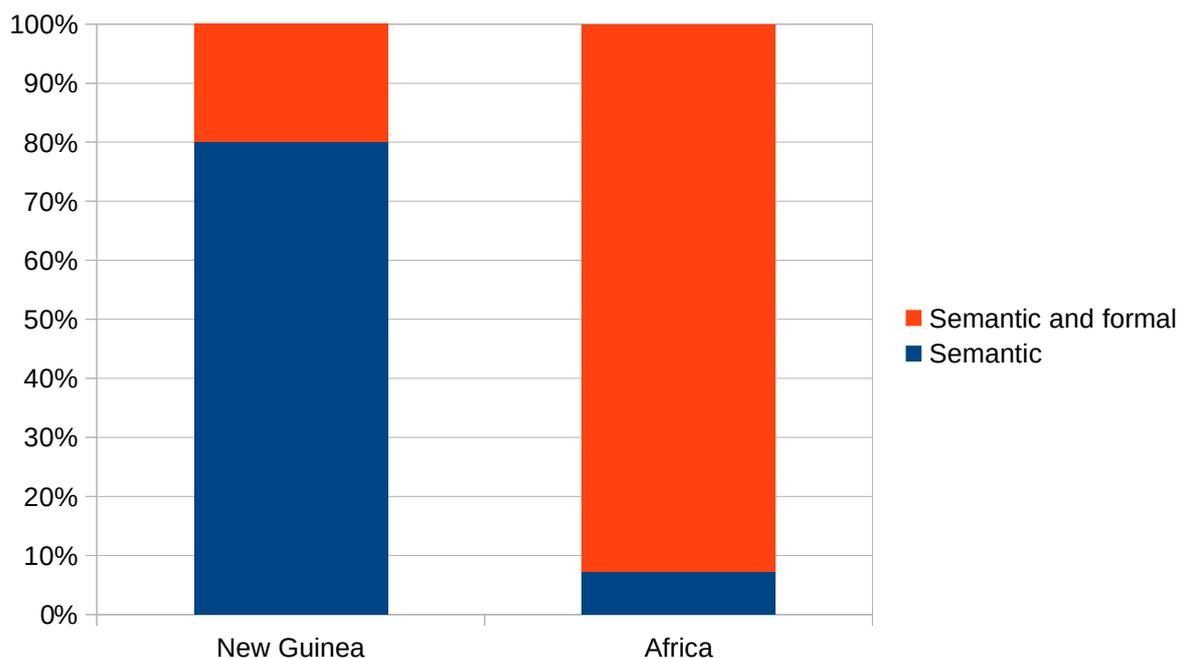


Figure 8: Gender assignment in New Guinea vs. Africa.

As shown in the above figure, the distribution of values for the criterion is widely different. In New Guinea, semantic and formal assignment is uncommon (although see the discussion of the previous criterion above), while it is by far the most common form of gender assignment in Africa, including of course the Bantu languages. As argued by Luraghi (2011), this implies that most gender systems of Africa are old. 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 & Svärd 2014).

Classification criterion 4: Number of gender-indexing targets (4.4). 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 9.

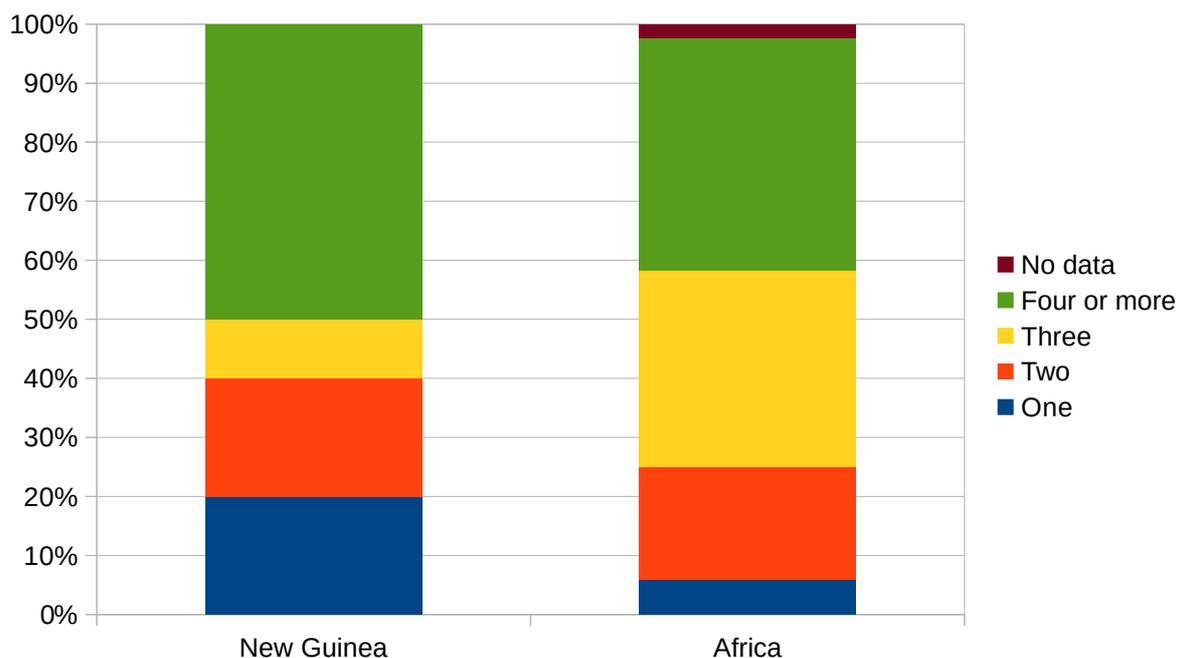


Figure 9: Number of gender-indexing targets in New Guinea vs. Africa.

Four or more gender-indexing targets is the most common number in both sample, 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 the African languages. However, once again it is probably that these differences are largely due to larger families with more established gender systems being better represented in Di Garbo's (2014) sample, while languages from smaller families with possibly less well-established gender systems constitute a large part of the sample of this study.

Classification criterion 5: Occurrence of gender marking on nouns (4.5). 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 percentual distributions is shown in Figure 10.

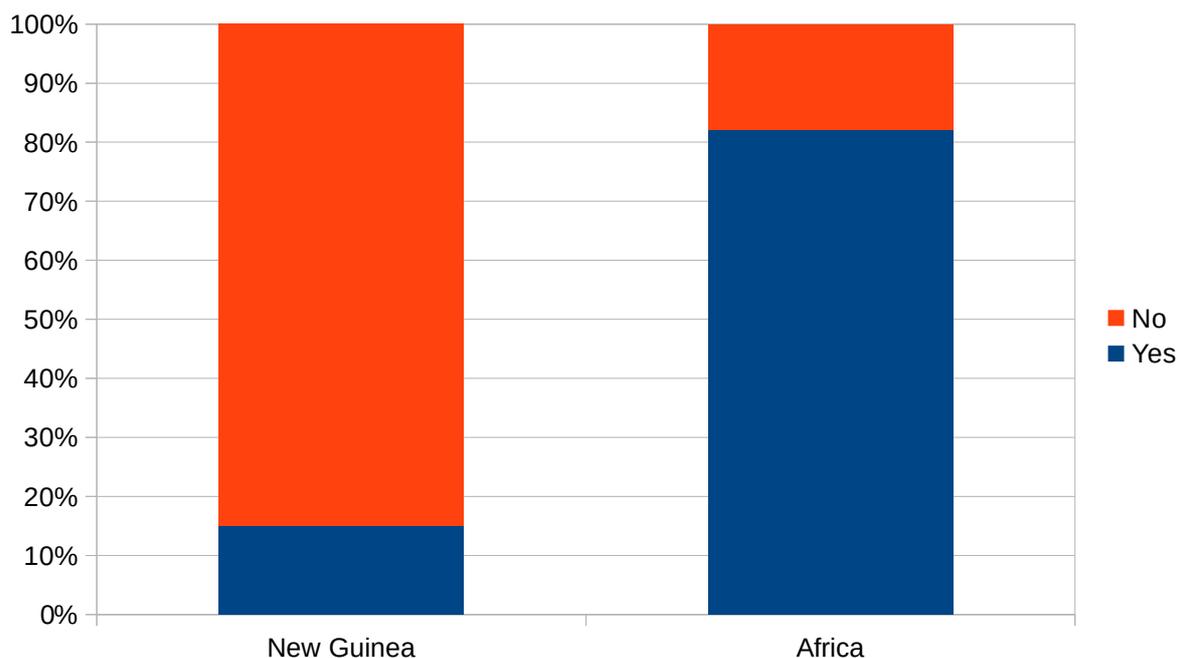


Figure 10: Occurrence of gender marking on nouns in New Guinea vs. Africa.

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, 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 8. 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 would benefit greatly from overt gender, which in fact would be the very targets of the criteria. In an exclusively semantic system however, obligatory overt gender would have 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, 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 implies that the gender systems of New Guinea may have different diachronical origins.

5.2 Special characteristics

In this section, four of the 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, 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.

5.2.1 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). Their rules based on shape and size are shown in Table 14.

Table 14: Gender assignment rules based on size and shape in the sample.

Language	Masculine ¹⁸	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 4.1 above is Abau, which has two genders: masculine and feminine. Humans, along with spirits and domesticated animals, are assigned gender based on their natural gender, 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).

It is important to distinguish systems such as the ones above from diminutives. In some languages, diminutives constitute separate genders, e.g., Motuna (South Bougainville) (Onishi 1994:68–69) or in many African languages (Di Garbo 2014:92), which are based (possibly among other factors) on size and/or shape. 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.

¹⁸ ‘Non-feminine’ in Skou.

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 2011:155–159). 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 2011:172). 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 figure illustrating the overlap between the two systems is shown in Figure 11 (repeated from Figure 2).

Verb prefix classes	Masculine	Flat	Bundle	Long	Residue	Feminine	
Genders	Male	Neuter 1			Female		Neuter 2

Figure 11: Overlap between the gender and verb prefix classes of Mian (adapted from Fedden 2011:173).

An interesting point to note is that Mian can be analyzed as having only two genders, viz., masculine and feminine, which indeed is the traditional analysis, based on the homophony of much of the gender agreement. This two-gender analysis is discussed and ultimately rejected by Fedden (2011:163–171; 2007). However, using this two-gender analysis, which is done by e.g., Foley (2000:372; Foley 1986:81), it is possible to interpret Mian as having gender assignment rules based on size that are reversed compared to the ones above, so that small objects are masculine, whereas large objects are feminine. Shape, however, is not reversed, so that round objects are feminine, whereas long objects are masculine. Foley exemplifies this with *imen* ‘taro’, where *imen-e* ‘small/one taro’ is masculine, whereas *imen-o* ‘large taro/large quantity of taro’ is feminine, where the agreement marker *-e* corresponds to the masculine singular and *-o* to the feminine singular. In the modern analysis by Fedden, *imen* is considered as belonging to the neuter 1 gender, where *-e* and *-o* are the singular and plural markers respectively, with the correspondences to the masculine and feminine being merely a matter of homophony (Fedden 2007:164). Thus, according to Fedden’s analysis, size and shape are not related to a noun’s belonging to different genders, but instead related to number marking.

Assigning genders based on shape and size is uncommon 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, most focus has been placed on New Guinean languages, specifically in the Sepik area, and Aikhenvald (2008:113) suggests 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 Alamlak (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, 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 system and not easily borrowed? Although this is far beyond the scope of this thesis, 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; as mentioned in 2.2.2 above, 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.

5.2.2 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 which occur with different types of targets. Thus, in such a language each noun is assigned not just to one class, but to two different classes. In the sample of this study, five languages have such systems (see Table 15).

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

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

Even in the small sample of this study, the two separate systems range from languages two more or less equally complex systems 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 (see also 2.1.2); however, it should be stated that there is a thin line between the two and most certainly constitute two edges of the same continuum. 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 equivalent gender system. In the first system, Burmeso has three genders (masculine, feminine, and neuter), appearing as adjectival agreement suffixes (12a), which are further divided into two subgenera 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 (12b) (Donohue 2001:101). In addition, there are three words which take both kinds of agreement: *-aysa-* ‘one’, *-akasu-* ‘all’, and *-asna-* ‘white’ (12c).

(12) Burmeso (isolate) (Donohue 2001:105, 109, 100 respectively)

- a. *Da de koya bek-abo.*
 1SG 1SG.POSS grandfather good-M.SG
 ‘My grandfather is well.’
- b. *Da mibo j-ih-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 natural gender 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 members of both systems is shown in Table 16.

Table 16: Comparison between genders and noun classes in Burmeso (adapted from Donohue 2001:108).

Class	Masculine	Feminine	Neuter	M inanimate	F inanimate	N animate
I	male humans (most birds, animals etc.) 2SG PRO	(birds of paradise) pigeons sago garden	sea	neck	–	sea wound
II	1SG PRO	female humans	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	–	–

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 (13a) (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 (13b) and predicative modifiers of the classified noun (Robinson 2011:50).

(13) Rotokas (North Bougainville) (Robinson 2011:149, 50 respectively)

- 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).

Abau also exhibits a clear noun classifier system. There are two genders in Abau, masculine and feminine, which follow opaque gender assignment rules and appear in e.g., pronouns and demonstratives (see Appendix C.1). 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 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.

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-eyn</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-eyn</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>

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 noun classifier system.

Table 18: Third person verbal classificatory prefixes 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 4.3), 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 somewhat 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, with similar systems occurring only in a few Indic, Dravidian, Iranian, and some Arawak languages (Aikhenvald 2008:185).

5.2.3 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 pronoun¹⁹." However, this is clearly proven wrong by the results of this study, where five languages do not exhibit gender in pronouns (see Table 19).

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

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

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.

¹⁹ 'Pronoun' is here understood as 'independent pronoun'.

²⁰ As in 4.4 above, 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.

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) on gender distinctions in independent pronouns. These chapters do not share the same sample, so 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 not should be assumed to be representative of anything), a surprising 15 languages (20%) do not show gender distinctions in independent pronouns. Coincidentally, this is the same ratio as in New Guinea as shown in the Table 19 above. Thus, it is clear that Greenberg's statement is not universal, although it certainly is a common pattern.

5.2.4 Gender in verbs

According to Greenberg's 31st Universal, "if either the subject or object noun agrees with the verb in gender, then adjective always agrees with the noun in gender." That is, if the verbs are indexing targets, so are adjectives. However, this is proven wrong by the distribution of values of the fourth classification criteria (see 4.4): 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.

5.3 Suggestions for further research

The most crucial limitation of this study is the sample. The sample is very small, and it is cannot be considered a satisfactorily accurate representation of gender in New Guinea. It would thus be crucial for samples in future studies to contain many more languages. Future samples would also need to be proportional. Additionally, the present study does not directly aim to investigate how its results may affect the theory of gender, which naturally would be topic of much importance in further studies.

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.

It would also be interesting to investigate features which are 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). Another topic for further research could be complexity. Complexity in the gender systems of Africa is investigated by Di Garbo (2014), and it would be interesting to compare complexity in Africa vs. New Guinea.

Future studies could also investigate the diachrony of gender in New Guinea. Firstly, 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. Secondly, the prevalence of sex-based systems suggest that many gender systems in New Guinea have different diachronical origins from e.g., those of Africa.

6. Conclusions

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, while a quarter of the languages have five or more genders. Only Bukiyip (Torricelli) and Yimas (Lower Sepik) have very large systems, with 18 and around a dozen genders respectively, although Burmeso has two separate gender systems with three and nine genders respectively. In the vast majority of the languages, gender assignment is semantic. Half of these have transparent semantic assignment, whereas the other half have opaque gender gender assignment. Half of the languages have four or more indexing targets, the most common of which are pronouns and verbs. Finally, gender marking on nouns is rare and occurs in only three genders in the sample.

The typological comparison suggest 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. Non-sex-based systems are prevalent in only a few language families in the world, and in the sample of this study it only occurs in the Austronesian language Teop. Secondly, the distribution of the four values for the number of genders are very similar, with the rate of occurrence of the values being $two > three \geq five \text{ or more} > four$ genders. Finally, 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. However, in both New Guinea and the world as a whole, it occurs in few families.

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 suggest 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. Secondly, five languages of the sample have two separate nominal classification systems. In these language, each noun is assigned to two classes which are reflected in different indexing targets, although only Burmeso exhibits two equivalent gender systems whereas the other rather distinguish between genders and noun classifiers. Thirdly, four languages in the sample have no gender distinctions in pronouns, which according to Greenberg's 43rd Universal should not occur. 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.

Appendix A. The language sample

For convenience, the full sample of languages is repeated in Table 20, together with family, genus, ISO code, and source. It is also shown in Table 2 and discussed in 3.1. The names primarily follow Glottolog, except for Motuna, where the Glottolog name Siwai appears less common. Also, the Glottolog form Warapu is used despite Barupu occurring in Corris (2008). Furthermore, language families and genera are based on Glottolog, so that a genus in the table below can conflict with the genus for the same language as given in WALS.

Table 20 (repeated from Table 2): The language sample. The en dash indicates no grouping (i.e., isolate) or that the language is itself the closest node to the family node.

Family	Genus	ISO code	Language	Source
Austronesian	Oceanic	tio	Teop	Mosel & Spriggs (2000)
Isolate	–	gpn	Taiap	Kulick and 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 (2008)
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 & Svärd (2014)
	Ok-Oksapmin	mpt	Mian	Fedden (2011)
		opm	Oksapmin	Loughnane (2009)
West Papuan	–	ayz	Maybrat	Dol (2007)

Appendix B. Overview of classification criteria

Table 21 presents an overview of the distribution of the values of the five classification criteria discussed in Chapter Error: Reference source not found. Note that for the second criteria on number of genders, the actual number of genders is shown in parenthesis for languages with five or more genders.

Table 21: Overview of the distribution of values of the classification criteria in the sample.

Language	Sex-based or non-sex-based	No. of genders	Gender assignment	No. of gender-indexing targets	Gender marking on nouns
Abau	sex-based	2	opaque	≥ 4	no
Ama	sex-based	3	opaque	1	no
Au	sex-based	3	trans. semantic	3	no
Bukiyip	sex-based	≥ 5 (18)	semantic-formal	≥ 4	yes
Burmeso	sex-based	3	opaque	2	no
Kuot	sex-based	2	semantic-formal	≥ 4	no
Manambu	sex-based	2	opaque	≥ 4	no
Maybrat	sex-based	2	trans. semantic	≥ 4	no
Mende	sex-based	2	trans. semantic	1	no
Mian	sex-based	4	opaque	≥ 4	no
Motuna	sex-based	≥ 5 (6)	trans. semantic	≥ 4	no
Nalca	sex-based	≥ 5 (5)	semantic-formal	1	no
Oksapmin	sex-based	2	trans. semantic	1	no
Rotokas	sex-based	3	opaque	≥ 4	no
Skou	sex-based	2	opaque	2	no
Taiap	sex-based	2	trans. semantic	3	no
Teop	non-sex-based	3	opaque	≥ 4	no
Walman	sex-based	2	trans. semantic	≥ 4	no
Warapu	sex-based	2	trans. semantic	2	no
Yimas	sex-based	≥ 5 (~12)	semantic-formal	≥ 4	yes

Appendix C. Samples of gender descriptions

In order to illustrate the compilation of sketches of the gender systems of the sample, this appendix contains the sketches of the two particularly interesting gendered languages, viz., Abau (Sepik) and Teop (Austronesian, Oceanic). Abau is included since it exemplifies gender assignment based on size and shape as well as two separate nominal classification system (one of genders and the other of classifiers). It is also interesting due to its system of noun phrase markers. On the other hand, Teop is included since it is the only Austronesian language in the sample and exhibits some interesting features, including a non-sex-based gender system and gender marking on nouns.

C.1 Abau (Sepik)

Abau has two genders (masculine and feminine) and a cross-cutting system of numerals classifiers (Lock 2011:46). Gender and number are not marked on the nouns themselves, but are indicated by pronouns, demonstratives, and other grammatical markers that follow the noun (Lock 2011:46). Interestingly, Abau has a system of noun phrase markers, which combine the expression of gender, case, topic, demonstrative, and number (see Table 22).

Table 22: Overview of all possible syntactic NP markers in Abau (adapted from Lock 2011:91).

Syntactic case	Syntactic NP marker		Proximal demonstratives	Distal demonstratives
	Case	+Topic	DEM	DEM
SBJ.MSG	<i>hiy</i>	<i>hiy-kwe</i>	<i>o-hiy</i>	<i>s-o-hiy</i>
SBJ.FSG	<i>hok</i>	<i>ho-kwe</i>	<i>o-hok</i>	<i>s-o-hok</i>
SBJ.DU	<i>hoh</i>	<i>hoh-kwe</i>	<i>o-hoh</i>	<i>s-o-hoh</i>
SBJ.PL	<i>hom</i>	<i>hom-kwe</i>	<i>o-hom</i>	<i>s-o-hom</i>
OBJ.MSG	<i>se</i>	<i>se kokwe</i>	<i>e-h-e</i>	<i>s-e-h-e</i>
OBJ.FSG	<i>ke</i>	<i>ke kokwe</i>	<i>o-kw-e</i>	<i>s-o-kw-e</i>
OBJ.PL	<i>me</i>	<i>me kokwe</i>	<i>e-m-e</i>	<i>s-e-m-e</i>
GEN.MSG	<i>so</i>		<i>o-hiy so</i>	<i>s-o-hiy so</i>
GEN.FSG	<i>ko</i>		<i>o-hok ko</i>	<i>s-o-hok ko</i>
GEN.PL	<i>mo</i>		<i>o-hom mo</i>	<i>s-o-hom mo</i>

Gender assignment in Abau is semantic and largely determined by the size and shape of the referent (Lock 2011:47). The general principles of gender assignment are shown in Table 23 below.

Table 23: General principles of gender assignment in Abau (adapted from Lock 2011:47).

Semantic criteria	Masculine	Feminine
Spirits and humans Domesticated animals	- marked according to inherent gender	- marked according to inherent gender
Animals	- large animals	- small animals
Concrete entities	- three-dimensional - liquids - long and extended	- two-dimensional - flat surface with little height - round with little height
Abstract entities		- always assigned feminine gender

Nouns referring to human beings and spirits are assigned gender based on their natural gender. This is true also for domestic animals, but other animals are assigned gender based on their shape and size (Lock 2011:47). In contrast, non-material and abstract nouns (e.g., *ok* ‘talk’, *mey* ‘work’, *nonkway* ‘knowledge’) are all feminine, with the sole exception of *prawk* ‘smell, scent’, which is masculine (Lock 2011:48).

Objects which are regarded as two-dimensional (i.e., very thin and flat, such as leaves and tables) are feminine (Lock 2011:48). This is also true for insects, which may be regarded as two-dimensional. In comparison, three-dimensional objects are masculine. However, not too voluminous objects with flat surfaces are feminine, which includes all fish, flat body parts (e.g., *iha* ‘hand’, *nweyk* ‘ear’, but not *makwey* ‘head’ or *kasaw* ‘nose’), and nouns like *howk* ‘lake’ and *iwa* ‘swamp’, but not *yawp* ‘river’ (Lock 2011:48). Furthermore, objects which are round from a two-dimensional perspective are feminine; this is seen in e.g., *sok* ‘snake’, which is masculine, but the green garden snake *wondeis* is feminine since it often coils up in the form of a circle (Lock 2011:49). Further examples are *hne* ‘bird’s nest’ and *kan* ‘a vine woven into a circle’, but not *bal* ‘ball’, which has a three-dimensional roundness (Lock 2011:49).

In contrast to the two-dimensional feminine objects, objects in which three-dimensionality is in focus are masculine. According to Lock, this depends on the perception of individual speakers: e.g., *mein* ‘stone’ can be feminine if features such as flatness and roundness are emphasized, whereas it can be masculine if the features of three-dimensionality or large size are in focus (Lock 2011:49). However, there are objects which are truly three-dimensional (and thus masculine), such as *su* ‘coconut’ and *am* ‘breadfruit’, and all liquids, e.g., *hu* ‘water’, *nioh* ‘blood’, and *sueyr* ‘rain’ (Lock 2011:49).

As stated above, animals are assigned gender based on size. Thus, larger animals tend to be masculine, such as *mu* ‘crocodile’ and *pareis* ‘wallaby’ (Lock 2011:49). This also included animals for which extendedness is a salient feature, such as *sok* ‘snake’ and *kie* ‘lizard’. Larger birds are also masculine, but birds like *sokua* ‘cokatoo’ and smaller are feminine (Lock 2011:49). One exception to the size-based gender assignment is *pruam* ‘cassowary’, which is feminine²¹; this may be due to the Abau world-view, since in a well-known creation story a cassowary gives birth to a human being (Lock 2011:50).

Finally, long and extended objects are masculine. Thus, *now* ‘tree’ and *iroum* ‘stick’ are masculine; however, if the referent has other salient features such as roundness and smoothness, it can be feminine, such as *now-ku* ‘tree stump’ and *youk* ‘paddle’ (Lock 2011:50). This is very much based on the speaker’s perception, as can be seen in Examples (14a) and (14b); in (14a), *youk* is masculine, since the speaker refers to a tree from which he makes a paddle (which is thus long and without any of

²¹ Except when referring to a specific male cassowary, in which case it becomes masculine (Lock 2011:50).

the roundness and smoothness of a paddle), whereas in (14b) *youk* is feminine since it refers to an actual paddle (Lock 2011:50). Such switches can occur in many other cases as well.

- (14) Abau (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.’

As stated above, Abau also has an extensive system of numeral classifiers. These can be divided into twelve groups on the basis of what numeral they occur with (Lock 2011:56). Thus, whereas gender is indicated through NP markers, the numeral classes of a noun is indicated by the set of numerals used for 1, 2, and 3 (Lock 2011:56). Nouns are assigned to the numeral classes based on humanness, surface, grouping, dimensionality, and extensionness; countable nouns which cannot be classified by their physical appearance are not restricted to one class but belong to classes 2, 3, or 4 (Lock 2011:57)²². An overview of the numeral classifiers can be seen in Table 24.

Table 24 (repeated from 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-eyn</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-eyn</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>

The choice of class and numeral can have a specific meaning; e.g., *su piron* ‘one coconut’ must refer to the whole coconut palm and not just the fruit, since class 5 signals long objects, while *su kamon* ‘one coconut’ is likely to be used when referring to just the fruit, since class 2 does not carry the semantic feature of length (Lock 2011:56). Furthermore, nouns can be used with different classifiers based on appearance, which is seen in Example (15).

²² Further specific assignment rules for numeral classifiers are discussed in Lock (2011:59–63).

(15) Abau (adapted from Lock 2011:57)

<i>pey piron</i>	‘one sugar cane (not cut)’	Class 5
<i>pey houkmon</i>	‘one piece of sugar cane’	Class 12
<i>pey eindmon</i>	‘one bundle of sugar cane, stored’	Class 9
<i>pey hnawmon</i>	‘one bundle of sugar cane, ready for transport’	Class 11

Thus, the main special characteristics of Abau is the gender assignment rules based on size and shape as well as the presence of a separate system of numeral classifiers.

C.2 Teop (Austronesian, Oceanic)

Teop has two genders, which are established on the basis of the form of indexed numerals and adjectives (Mosel & Spriggs 2000:322). Note that there are no gender differences for pronouns. Furthermore, the form of the article of the head noun leads to a further division of Gender I into two subgenders, viz., I-E and I-A. As seen in the Table 25 and Example (16), Teop exhibits an exceptional form of syncretism where the singular and plural articles show complete polarity (Mosel & Spriggs 2000:321–322).

Table 25: Gender marking 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>

(16) Teop (adapted from Mosel & Spriggs 2000:323):

a.	<i>e</i>	<i>maagee</i>	<i>my</i>	d.	<i>o</i>	<i>maagee</i>	<i>my</i>
	ART.I-E.SG	friend	my		ART.I.PL	friend	my
		‘my friend’				‘my friends’	
b.	<i>a</i>	<i>moon</i>		e.	<i>o</i>	<i>moon</i>	
	ART.I-A.SG	woman			ART.I.PL	woman	
		‘the woman’				‘the women’	
c.	<i>o</i>	<i>hoi</i>		f.	<i>o</i>	<i>hoi</i>	
	ART.II.SG	basket			ART.II.PL	basket	
		‘the basket’				‘the baskets’	

There is also a special plural article *ere*, which can be used with personal names, kinship terms, and the interrogative pronoun *teiee* (Mosel & Spriggs 2000:323). It can also be used with singular proper names, where it carries the meaning that the person is accompanied by one or more other people (17a). It is also used to coordinate two kinship terms or proper names (17b).

(17) Teop (Mosel & Spriggs 2000:323)

- a. *ere Kakato*
 PL.ART K.
 ‘Kakato and his company’
- b. *ere Kakato bo Sovavi*
 PL.ART K. and S.
 ‘Kakato and Sovavi’

The articles *e*, *a*, *o*, and *ere* precede nominal predicates, specific subject noun phrases, and certain specific object noun phrases; they contrast with four other kinds of articles, two of which are marked for gender (Table 26).

Table 26: The paradigm of Teop articles (Mosel & Spriggs 2000:324).

Gender	Singular			Plural		
	I-E	I-A	II	I-E	I-A	II
Basic article	<i>e</i>	<i>a</i>	<i>o</i>	<i>ere</i>	<i>o</i>	<i>a</i>
Object articles	<i>bene</i>	<i>bona</i>	<i>bono</i>	<i>bere</i>	<i>bono</i>	<i>bona</i>
Non-specific article		<i>ta</i>	<i>to</i>			
Specific partitive article		<i>eta</i>				
Non-specific partitive article		<i>sa</i>				

In general, plurality is shown by the plural articles. However, plurality can also be shown via reduplication of certain adjectives (with plural articles) (18) and the plural marker *maa* (which requires the article *a*) (19) (adapted from Mosel & Spriggs 2000:330–1). In the latter case, attributive or predicative adjectives take the simple plural article when the head noun is human (19a), but otherwise the *maa*-plural (19b) (except nouns referring to domestic animals, which allow either construction).

(18) Teop (Mosel & Spriggs 2000:330)

- a. *a inu a beera*
 ART.I-A.SG house ART.I-A.SG big
 ‘the big house’
- b. *o inu o be-beera*
 ART.I.PL house ART.I.PL RED.PL-big
 ‘the big houses’

(19) Teop (Mosel & Spriggs 2000:331)

- a. *a maa moon o be-beera*
 ART.I-A.SG PL woman ART.I.PL RED.PL-big
 ‘the big (adult) women’
- b. *a maa inu a maa be-beera*
 ART.I-A.SG PL house ART.I-A.SG PL RED.PL-big
 ‘the big houses’

According to Mosel & Spriggs (2000:334), gender assignment in Teop is highly predictable and based on semantics. Gender I-E contains all personal names, all kinship terms, nouns denoting pets, and a number of nouns referring to people with a particular communal or important social status (Mosel & Spriggs 2000:334–5). In contrast, gender I-A contains most words and can be regarded as the unmarked gender (Mosel & Spriggs 2000:336). It contains nouns denoting male and female humans (other than those of I-E), vertebrates, insects, and crabs (but not molluscs, shellfish, etc., > II), fruits and nuts (but not plants > II), utensils (but not wooden ones, mats, and clothes > II), prepared food, landmarks, wind and rain, the interrogative pronoun, and some abstract nouns derived from verbs (Mosel & Spriggs 2000:336–8). Finally, Gender II contains names of plants and their parts (but not fruits > I-A), objects made of plant material, invertebrates without legs, and many mass and abstract nouns (Mosel & Spriggs 2000:338).

References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. (Oxford Studies in Typology and Linguistic Theory). 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. D-level.
- 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-b-9780199281251> (15 May, 2015).
- Brown, Lea & Matthew S. Dryer. 2008. Diminutive as an inflectional category in Walman.
<http://linguistics.buffalo.edu/people/faculty/dryer/dryer/BrownDryerWalmanDimin.pdf> (4 February, 2015).
- Bruce, Leslie P. 1984. *The Alamlak language of Papua New Guinea (East Sepik)*. (Pacific Linguistics 113). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Bybee, Joan. 1985. Diagrammatic iconicity in stem-inflection relations. *Iconicity in syntax*. 11–48. (18 May, 2015).
- Conrad, Robert J. & Kepas Wogiga. 1991. *An outline of Bukiyip grammar*. (Pacific Linguistics 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 Textbooks in Linguistics). 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> (15 October, 2014).
- 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> (26 March, 2015).
- 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> (27 March, 2015).
- Corris, Miriam. 2008. *A grammar of Barupu: A language of Papua New Guinea*. Sydney: University of Sydney.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William. 2003. *Typology and universals*. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Croft, William. 2013. Agreement as anaphora, anaphora as coreference. *Languages Across Boundaries: Studies in Memory of Anna Siewierska*. 95.
- Di Garbo, Francesca. 2014. Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa. Stockholm: Stockholm University. Ph.D. thesis.
- 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. *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton*. 97–115.

- Donohue, Mark. 2004. A grammar of the Skou language of New Guinea. http://pubman.mpg.de/pubman/item/escidoc:402710/component/escidoc:402709/skou_donohue2004_s.pdf (6 February, 2015).
- Fedden, Sebastian. 2007. Women, houses, and plural objects?—Homophony in the Mian gender system. <https://minerva-access.unimelb.edu.au/handle/11343/34686> (3 May, 2015).
- Fedden, Sebastian. 2011. *A grammar of Mian*. Berlin, Boston: De Gruyter Mouton.
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Foley, William A. 1991. *The Yimas language of New Guinea*. Stanford University Press.
- Foley, William A. 2000. The languages of New Guinea. *Annual review of anthropology*. 357–404.
- Frowein, Friedel Martin. 2011. A grammar of Siar: An Oceanic language of New Ireland province, Papua New Guinea. Victoria: La Trobe University.
- Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg, ed., *Universals of Language*. 73–113. Cambridge, MA. <http://dSPACE.unive.cilea.it/handle/10278/3011> (12 February, 2014).
- Haspelmath, Martin. 2013. Argument indexing: a conceptual framework for the syntactic status of bound person forms. *Languages across boundaries: Studies in memory of Anna Siewierska*. 209–238.
- Heine, Bernd. 1982. African noun class systems. *Apprehension: Das Sprachliche Erfassen von Gegenständen* 1. 189–216.
- Hockett, Charles Francis. 1958. A course in modern linguistics. *Language Learning* 8(3–4). 73–75.
- Hoel, Hanna Marie, Tarja Ikäheimonen & Michiyo Nozawa. 1994. Mende grammar essentials. Seim, Sandaun Province, ms.
- 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 (ed.), *The Circum-Baltic languages: Typology and contact*, vol. 2 Grammar and typology, 615–750. Amsterdam: John Benjamins Publishing Company.
- Kulick, Don & Christopher Stroud. 1992. The structure of the Taiap (Gapun) Language. *The language game: Papers in memory of Donald C. Laycock*, vol. 110, 203–226. (Pacific Linguistics, Series C). 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. Ph.D. thesis.
- Lock, Arjen. 2011. *Abau grammar*. SIL-PNG Academic Publications. [http://www-01.sil.org/pacific/png/pubs/928474542179/DPv57_Abau%20Grammar_final%20\(pdfA\).pdf](http://www-01.sil.org/pacific/png/pubs/928474542179/DPv57_Abau%20Grammar_final%20(pdfA).pdf) (11 February, 2015).
- Loughnane, Robyn. 2009. A grammar of Oksapmin. Melbourne: University of Melbourne.
- Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45(2). 435–463.
- Mosel, Ulrike & Ruth Spriggs. 2000. Gender in Teop (Bougainville, Papua New Guinea). *Trends in Linguistics Studies and Monographs* 124. 321–350.
- Onishi, Masayuki. 1994. A grammar of Motuna (Bougainville, Papua New Guinea). Canberra: Australian National University. Ph.D. thesis.
- Phillips, Colin. 1993. Conditions on agreement in Yimas. *Papers on Case & Agreement I, MITWPL* 18. 173–213.
- Robinson, Stuart Payton. 2011. Split intransitivity in Rotokas, a Papuan language of Bougainville. Nijmegen: Radboud Universiteit. Ph.D. thesis.
- Rumsey, Alan. 2002. On the grammaticalization of posture verbs in Papuan languages, its bodily basis and cultural correlates. *The Linguistics of Sitting, Standing and Lying* 51. 179.
- Scorza, David. 1985. A sketch of Au morphology and syntax. *Pacific Linguistics. Series A: Occasional Papers*(63). 215–273.
- 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> (3 May, 2015).

- Stanhope, John M. 1972. The Language of the Kire People, Bogia, Madang District, New Guinea. *Anthropos*. 49–71.
- Steele, Susan. 1978. Word order variation: A typological study. *Universals of Human Language* 4. 585–623.
- Svärd, Erik. 2013. Selected topics in the grammar of Nalca. Stockholm: Stockholm University. <http://www.diva-portal.org/smash/record.jsf?searchId=1&pid=diva2:631006> (9 November, 2013). B.A. thesis.
- Tharp, Doug. 1996. Sulka grammar essentials. (Ed.) John M. Clifton. *Two Non-Austronesian Grammars from the Islands*. 77–179.
- Wälchli, Bernhard & Erik Svärd. 2014. The emergent rise of a gender system in Nalca (Mek, Tanah Papua): Inqorate noun classes, adjacency agreement, and switching gender on and off in the syntax. (Under Review).
- Zubin, David & Klaus-Michael Köpcke. 1986. Gender and folk taxonomy: The indexical relation between grammatical and lexical categorization. *Noun classes and categorization*. 139–180. (23 July, 2015).

Stockholm University
SE-106 91 Stockholm
Phone: 08 – 16 20 00
www.su.se



Stockholms
universitet