Plural Words in Austronesian Languages:

Typology and History

A thesis submitted in partial fulfilment of the
requirements for the degree of
Research Master of Arts
in
Linguistics
by

Jiang Wu

Student ID: s1609785
Supervisor: Prof. dr. M.A.F. Klamer
Second reader: Dr. E.I. Crevels
Date: 10th January, 2017
Faculty of Humanities, Leiden University

Universiteit
Leiden

Table of contents

Abstract

This thesis presents a systematic study on plural words, a particular type of nominal plurality marking, in Austronesian languages. More specifically, it investigates the synchronic distribution and diachronic developments of plural words in Austronesian languages from a typological perspective.

Plural words are defined as "separate words which modify nouns but which serve the same grammatical function as plural affixes in other languages" (Dryer 1989a: 865). Since Dryer's pioneering treatment, plural words have received very little attention, and no follow-up study has been carried out to characterise plural words in any particular language family. Some observations about plural words in Austronesian languages also remain preliminary due to the problems with language samples and the way in which plural words are identified. Building on previous studies, this thesis explores how plural words in Austronesian languages are distributed, and discusses the diachronic developments of these plural words.

An extensive new language sample is collected, which consists of 128 Austronesian languages across different genealogical subgroups and geographical areas. The languages are all selected as proportionally representing the most optimal genealogical subgroupings of Austronesian languages to our best knowledge so far (cf. Adelaar 2005a; Hammarström et al. 2016). In defining plural words, I apply a narrow definition and only consider pure plural words.

It is found that 54 Austronesian languages in my sample employ plural words, and their synchronic distribution is skewed. Plural words are mostly found in Philippine languages and Oceanic languages, and they are also frequently used in Central Malayo-Polynesian languages. As for diachronic developments, plural words in Austronesian languages have a number of independent origins, but some shared histories can also be identified. A great number of plural words (20/54) originate from a third person plural pronoun, and a few of them reflect a Proto Malayo-Polynesian (PMP) and Proto Oceanic (POc) reconstruction **maŋa*.

The results of this thesis can serve as a foundation upon which further investigation into plural words in individual languages can be conducted. The sparse presence of plural words reflecting **maŋa* also calls for a reconsideration of the PMP and POc reconstruction **maŋa*.

Acknowledgements

This thesis is the most valuable piece of academic work I have ever accomplished so far. During the process of writing, I have received abundant help from many people, without whom this work would not have been possible.

First and foremost, I would particularly thank my supervisor, Prof. Marian Klamer. In the months I was writing my thesis, she has been willing to spare her time to have a meeting with me and offer me plentiful remarks and inspiring suggestions. The outline and many parts of this thesis have been adjusted several times, and I am grateful that I am able to finish it within a tight schedule with her guidance. I would also like to seize this opportunity to thank her for her encouragement and help during my study in Leiden. She was one of the main reasons why I chose to do my master in Leiden University; and during the two-year Research Master Programme in Leiden, I have learned a lot from her (and of course other teachers!), and profoundly enhanced my understanding of Austronesian and Papuan linguistics.

I would also like to thank Dr. Mily Crevels (who eventually became the second reader of this thesis) and Kate Bellamy, who opened the door of linguistic typology for me when I first started my study in Leiden.

My gratitude also goes to anyone who have read (part of) my thesis and given me insightful comments, and those who have helped me clear my mind with refreshing discussions, especially to Wei-wei and Sara. I owe a special thank-you to Stefano Mattia, who has painstakingly read one chapter after another, and one draft after another of my thesis.

The teachers and friends I met in Porquerolle during the European Summer School of Linguistic Typology are also worth mentioning here (it was also Prof. Marian Klamer who suggested me to attend this summer school). The two-week summer school has been an invaluable experience for me, where I largely broadened my horizons and got to know many people. In the master class, Prof. Greville Corbett and Prof. Cristofaro Sonia shared their immense knowledge on this topic with me, and the comments I received from my fellow classmates are also much appreciated. Prof. Elizabeth Zeitoun also gave me some helpful advice on the structure of the thesis, and provided me with some suggestions for references. In particular, I would like to express my sincere appreciation to Lilian Li for her relentless help and encouragement (even if we have a seven-hour time difference!).

Finally, my heartfelt thanks go to my parents, for their unwavering support. Even if they probably do not understand what I have been studying, but thanks for letting me choose my own road!

Jiang Wu

List of tables

List of figures

List of maps

List of abbreviations

1	first person	IND	independent (pronoun)
2	second person	INTR	intransitive
3	third person	IPFV	imperfective
ACTV	actor verb	IRR	irrealis
AD	adnominal	LK	linker
ADV	adverb	LOC	locative
ANAPH	anaphoric	M	masculine
APRX	approximative	NEG	negation
AR	aorist	NM	noun marker
ART	article	NOM	nominative
AV	actor voice	OBJ	object
CAUS	causative	OBL	oblique
CLF	classifier	PAUC	paucal
COMPL	completive	PERS	personal
CONT	continuous	PFV	perfective
DEF	definite	PL	plural
DEM	demonstrative	POSS	possessive
DET	determiner	POTLOCV	potentive locative(-oriented) verb
DEX	indexer	PREP	preposition
DIR	directional	PROX	proximal
DIST	distal	PSR	possessor
DISTR	distributive	PST	past
DU	dual	RDP	reduplication
ERG	ergative	REAL	realis
F	feminine	RECP	reciprocal
FOC	focus	RES	resultative
FREQ	frequentative	SBJ	subject
GEN	genitive	SG	singular
GRTR.PL	greater plural	THMV	theme(-oriented) verb
IMM	immediate	TOP	topic marker
INCH	inchoative	TR	transitive

N.B. In this thesis, some adjustments have been made to the glossing in the examples I take from existing grammars, for the sake of consistency.

Chapter 1. Introduction

This thesis addresses a particular type of nominal plural marking called plural words. Unlike many languages which mark nominal plurality by morphological means (for instance, English uses suffix *-s* to denote plurality, thus *dog* > *dog-s*), a minority of languages do not mark nominal plurality on the noun itself at all, but somewhere else in the noun phrase. Such markers, defined by Dryer (1989a: 865) as "separate words which modify nouns but which serve the same grammatical function as plural affixes in other languages", are called plural words. An example can be seen in Tagalog as illustrated by example (1), where the plural marker *mga* is not an affix, but a separate word. A plural word is also present in other languages, such as Abui in (2).

(1) Tagalog (the Philippines, Austronesian)

Mga	*abogado*	*ang*	***mga***	*lalaki.*
PL	lawyer	TOP	PL	man

'The men are lawyers.' (Schachter & Otanes 1972: 111)

(2) Abui (Timor-Alor-Pantar)

neng	***loku***
man	PL

'the men' (Kratochvíl 2007: 165)

Some interesting observations about plural words in Austronesian languages can be found in previous studies. Synchronically, it is shown that plural words in Austronesian languages are mostly found in Philippine and Oceanic languages with a few instances in other areas (Dryer 1989a; 2013a), therefore exhibiting a skewed distribution. Also, within the Oceanic group alone, a great majority of languages employ plural words (Dryer 2013a). Diachronically, a plural word is reconstructed in Proto Oceanic (POc) as **maŋa*, as a descendant from Proto Malayo-Polynesian (PMP) **maŋa* (Lynch et al 2002: 90–91). The plural word *mga* in Tagalog is a reflex of such a reconstructed form.

However, the Austronesian sample languages examined in these studies are not balanced, since none of them aims at presenting a dedicated study on plural words in this language family. As a result, the above-mentioned observations have to be taken as preliminary rather than conclusive. Further study is needed before we can draw any conclusions about plural words in Austronesian languages.

In this thesis, I revisit the typology and history of plural words in Austronesian languages, examining their synchronic distribution and diachronic developments. A more extensive and balanced language sample is collected, and the coding of nominal

plurality of these sample languages is examined and analysed.

The following chapters of this thesis are organised as follows. Chapter 2 discusses previous literature, based on which I put forward my research questions. In Chapter 3, I reconsider the definition of plural words, and discuss the need to revise the criteria in identifying plural words in previous studies. Chapter 4 describes the methodology that I use to answer the research questions, and presents my language sample. Chapter 5 presents the distribution of plural words in Austronesian languages based on my sample languages, and Chapter 6 discusses the diachronic developments of the plural words found in the sample. Chapter 7 concludes the whole thesis.

Chapter 2. Background literature

The most important concept throughout this thesis is *plural words*. In this chapter, I review previous literature and discuss how plural words were introduced and defined, and then elaborate on the observations about plural words in Austronesian languages in more details.

2.1. Plural words as nominal plurality marking

Broadly speaking, plural words are a particular kind of nominal plurality marking, which concerns the grammatical feature of number. Grammatical number "encodes quantification over entities or events denoted by nouns or nominal elements" (Kibort & Corbett 2008). A distinction can be made between nominal number and verbal number; in this thesis, I limit myself to the nominal domain and focus on the most common type of values for grammatical number – plurality.

Within the scope of nominal plurality alone, there are a number of different ways in which plurality can be marked across world's languages. As for the languages in which a number distinction in the nominal domain can be made, Dryer (2013a) distinguishes two major types of nominal plurality marking:

- marking that involves changing the morphological form of the noun;
- marking that involves indicating plurality by means of a morpheme that occurs somewhere else in the noun phrase.

The difference between these two patterns lies in the level on which plural markers occur – on the noun itself, or on the noun phrase level when the noun is not marked for number by any morphological means. The most common plural marker in English, namely the suffix *-s*, as in *dog* > *dog-s*, falls into the first category because the noun itself undergoes the process of affixation. Other morphological marking can be found in other languages, for instance, plural prefix in Palauan illustrated by example (3), and stem change in Jamul Tiipay, as in example (4).

(3) Palauan (Palauan, Austronesian)
chad 'person' > *rę-chad* 'people'
kangkodang 'tourist' > *rę-kangkodang* 'tourists' (Josephs 1975: 43)

(4) Jamul Tiipay (Yuman, Cochimi-Yuman)
nyech'ak 'woman' > *nyech'aak* 'women'
xechany 'girl' > *xaachaaney* 'girls' (Miller 2001: 115)

Notwithstanding plural affixes being by far the most common type of nominal plurality markers, a minority of languages do not mark nominal plurality on the noun itself at all, but somewhere else in the noun phrase. Those markers on the phrase level are called plural words (1989a; 2007; 2013a). Dryer (1989a: 865) gives a description of this term: they are "separate words which modify nouns but which serve the same grammatical function as plural affixes in other languages". In addition to the samples from Tagalog and Abui in (1) and (2), the usage of plural words can be illustrated in more languages, as in (5) in Dogon and (6) in Raga.

(5) Dogon (Niger-Congo)
a. *ɛnɛ* ***mbe***
goat PL
'goats'
b. *ɛnɛ* *gɛ* ***mbe***
goat DEF PL
'the goats' (Plungian 1995: 9–10)

(6) Raga (Oceanic, Austronesian)
Ira *naturigi ra-m* *gan damu.*
PL child 3PL-CONT eat yam
'The children eat yam.' (Vari-Bogiri 2011: 97)

In those languages, the pluralisation of nouns is not expressed by an affix or other morphological changes on the stem, but by a separate word. The plural word serves similar functions as the plural suffix in English, but operating on a phrase level. It should be noted, however, that such words sometimes do not only encode the number value of plurality, but also singular, dual, trial or paucal number. Yapese is one of the languages which make a distinction between singular, dual and plural number words, as in example (7). I acknowledge that some dual or trial words are indeed present in my study, but I will still use the term *plural words* given that the majority of such number words are plural words, also as a continuity of previous terminology.

(7) Yapese (Oceanic, Austronesian)
a. *ea* ***rea*** *kaarroo neey*[1]
ART SG car this
'this car'

[1] *Ea* is analysed as a noun phrase connector, which might be considered to be a subtype of articles (Jensen 1977:157).

b. *ea* ***gäl*** *kaarroo neey*
ART DU car this
'these two cars'

c. *ea* ***pi*** *kaarroo ney*
ART PL car this
'these cars' (Jensen 1977: 155)

Even though Dryer (1989a) considers plural words to be separate words which resemble plural affixes in other languages, it is not always clear how to identify a plural word. *Plural words* for Dryer is not a word class defined by a set of definite morphosyntactic criteria, but a group of words (or morphemes) that share the following features across languages (Dryer 1989a: 866–867):

- firstly, plural words differ from 'many' in that 'many' inherently implies an amount of more than two, while plural words do not;
- secondly, plural words also differ from 'many' and 'some' in that 'many' and 'some' also encode indefiniteness, while plural words do not necessarily do so; yet, intrinsically encoding definiteness or indefiniteness does not disqualify a plural word as such;
- thirdly and most importantly, plural words are the sole indicators of plurality in noun phrases.

As can be seen, the descriptive criteria applied by Dryer are quite broad. Considering the fact that typologists do not always have first-hand sources, it is very difficult to be fully sure whether a plural word actually differs from 'many' in that the plural word can also be used to refer to two entities[2]. Whether encoding (in)definiteness is not diagnostic either, as some plural words defined by Dryer do simultaneously encode (in)definiteness. Therefore, the most important criterion is that the potential plural word has to be the only indicator of plurality on the phrase level. Using these criteria, Dryer identifies a group of plural words which do not only include the plural markers being a part of speech in their own right, but also include plural articles, and plural demonstratives (also see Dryer 2007: 167), as in the following cases:

(8) Hoava (Oceanic, Austronesian)
a. *na* *koburu*
ART child
'the child/a child'

[2] Also, when there is a dual number word alongside the plural word, the plural word would mean 'more than two'.

b. *sa* *koburu*
ART.SG child
'the child'

c. ***ria*** *koburu*
ART.PL child
'the children' (Davis 2003: 36)

(9) Kokota (Oceanic, Austronesian)

a. *kame=ḡu=ine*
arm=1SG=this.PROX
'this hand of mine'

b. *kame=ḡu=**ide***
arm=1SG=these.PROX
'these hands of mine' (Palmer 2009: 84)

Unlike the plural words in Abui or Tagalog, which constitute a particular grammatical category on their own, the plural word *ria* in Hoava is also a plural article. The word *koburu* in Hoava can mean either 'child' or 'children'. In many cases, its number can be inferred from the context or expressed by quantifiers, but using articles is also one way to overtly express number value. In (8c), the plural marker *ria* is the sole indicator of nominal plurality, therefore fitting into Dryer's criteria. At the same time, it simultaneously encodes plurality and definiteness. Similarly, in Kokota, demonstratives make a distinction between singularity and plurality, and plurality of a noun phrase can also be expressed by a plural demonstrative[3]. Both languages are considered to have plural words by Dryer (2013a), and the results can be found on The World Atlas of Language Structures (WALS) online.

One theoretical question one might ask here is: how can we define a plural marker to be a word, and distinguish it from a clitic or an affix? This question essentially boils down to a discussion of how we can identify a *word*, and it is indeed true that defining the concept of *word* is not an easy task.

As regards distinguishing plural words and plural clitics from plural affixes, one method is to check if the morpheme in investigation can be separated from the head noun by other elements. If yes, it is clearly a word or clitic, not an affix. Example (10) from Unua illustrates such a case:

[3] Articles in Kokota also differentiate singularity from plurality.

(10) Unua (Oceanic, Austronesian)

Go	*i-suatoxn-i*	*batin*	*nixe*	*demen*	***rin***
and	3SG-pull.down-TR	tree	wood	huge	PL

'And it pushed down huge trees.' (Pearce 2015: 188)

The plural marker *rin* and the noun *batin* 'tree' are separated by two adjectives, thus ruling out the possibility of it being an affix. However, in many languages employing plural words, the plural words always occur adjacent to the nouns, thus this criterion is not always applicable.

Differentiating words from clitics is even harder, since they are semantically and syntactically similar in many ways. Unlike affixes, clitics can be attached to different categories of words rather than a particular part of speech. But the essential difference between a word and a clitic sometimes relies on phonological aspects, and it is not always useful and necessary to distinguish them for analytical purposes. As Grimes (1991: 159) notes for the plural clitic in Buru, it "functions grammatically at the level of the NP, but phonologically at the word level". Nouns in Buru can be marked for plurality by a clitic *=ro*, as in example (11a). When attached to other clitics, *=ro* undergoes morphophonemic alternation, as in (11b). It is therefore not easy to judge if *=ro* is a clitic or a word by using semantic and syntactic evidence alone, and we need to seek for phonological evidence.

(11) Buru (Central-Malayo-Polynesian, Austronesian)

a. *fatu* 'rock' > *fatu=**ro*** 'rocks'
   huma 'house' > *huma=**ro*** 'houses' (Grimes 1991: 147–148)

b. *toho=n=**o***
   descend=GEN=PL
   'paths, trails' (Grimes 1991: 148)

In Dryer's analysis, he generally accepts the claims made in the grammars about whether a morpheme is an affix, a clitic or a word when no other analytical methods can be applied. Since the most distinctive feature of a plural word is that it operates on a phrase level, plural clitics are considered to be parts of plural words.

2.2. Plural words in Austronesian languages

Based on the preceding background, Dryer (1989a) conducts a typological study on plural words in world's languages. Plural words are also discussed in his chapter on coding of nominal plurality on The World Atlas of Language Structures online (WALS) (Dryer 2013a). From these works, some interesting observations about plural words in Austronesian languages can be found.

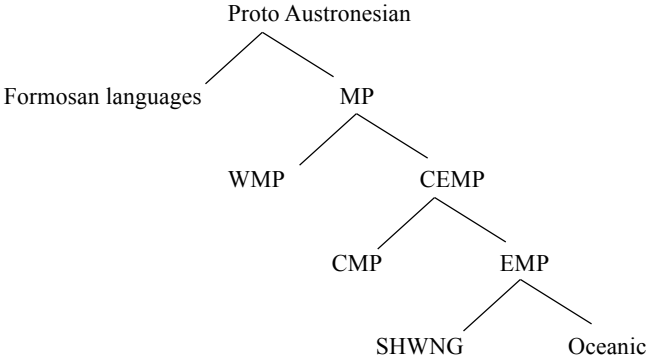
Firstly, it is observed in Dryer (1989a) that plural words are particularly frequent in Austronesian languages compared to other language families. He examines the coding of nominal plurality in a sample of 307 languages, among which 48 languages employ plural words. Out of these 48 languages with plural words, almost half of them are Austronesian (22/48). Secondly, these 22 Austronesian languages with plural words are skewedly-distributed genealogically and geographically. In terms of genealogical affiliations, these Austronesian languages with plural words are either Western Malayo-Polynesian (WMP) languages or Oceanic languages.[4] As for geographical distributions, Map 1 shows that most of these languages are either spoken in the Philippines (WMP languages) or on islands of the Pacific Ocean (Oceanic languages). Outside these two areas, only one language (Toba Batak, WMP) on the island of Sumatra in west Indonesia has a plural word.

A similar skewed distribution of plural words in Austronesian languages can also be found in the chapter by Dryer (2013a) on WALS. This chapter can be seen as an extended work of Dryer (1989a), and it uses a much more extensive language sample consisting of 1066 languages across the world, among with 115 are Austronesian. 76 out of these 115 Austronesian languages employ plural words.

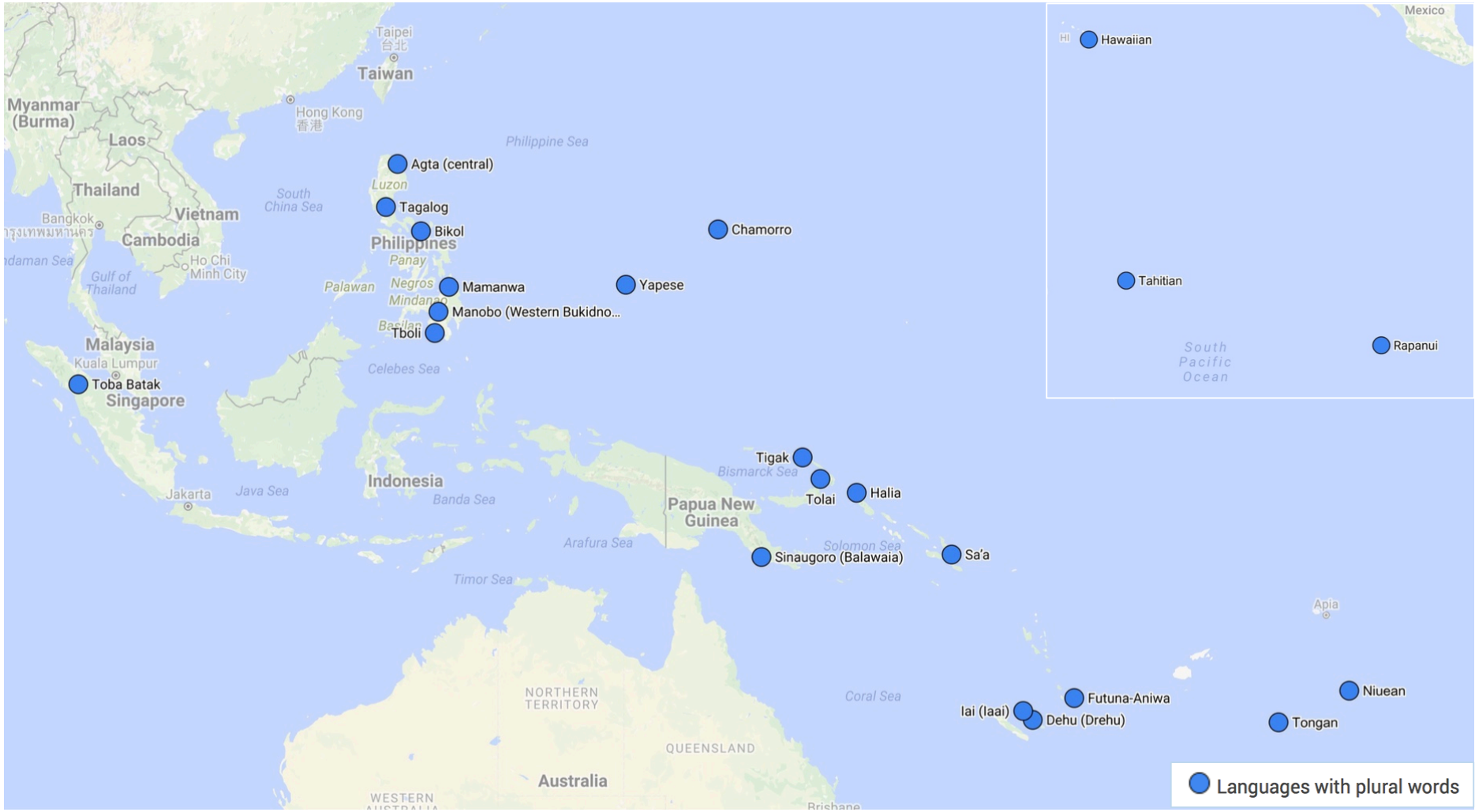
Similarly, as shown by Map 2, plural words in Austronesian languages are still mostly found in WMP languages in the Philippines, and Oceanic languages on Pacific islands and east Papua New Guinea. In west Indonesia, Toba Batak is still the only instance. But in Dryer (2013a), we also find some other areas where Austronesian languages with plural words are present: north Borneo, east Indonesia. The genealogical distributions of these languages are also more diverse: languages in east Indonesia are classified into the WMP group, and some SHWNG languages in the bird's head of Papua New Guinea also have plural words, as represented by Biak and Ambai.

---

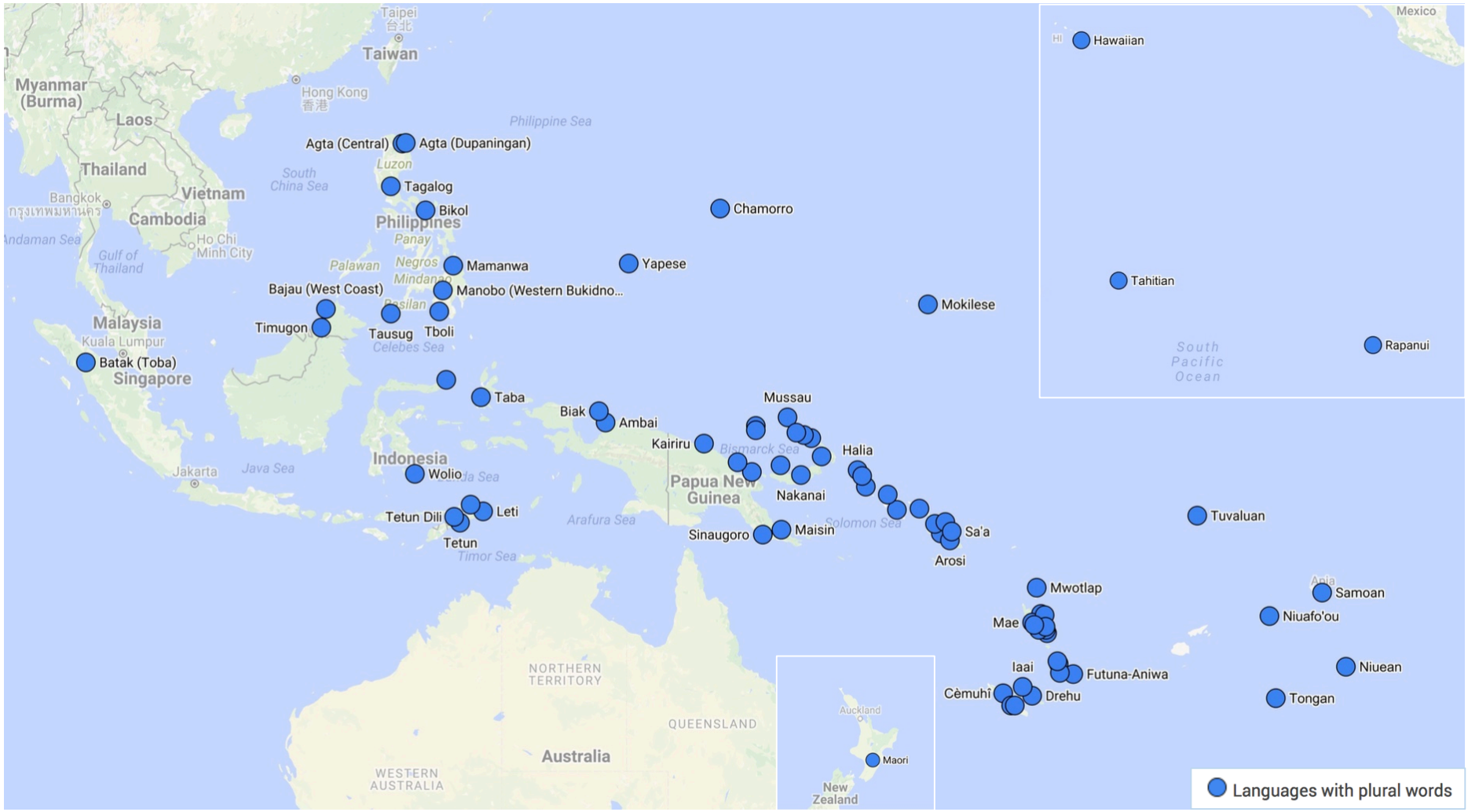
[4] Both Western Malayo-Polynesian (WMP) and Oceanic are major subgroups of the Austronesian language family. Other nodes in the Austronesian family tree include Malayo-Polynesian (MP), Central-Eastern Malayo-Polynesian (CEMP), Central Malayo-Polynesian (CMP), Eastern Malayo-Polynesian (EMP), South Halmahera-West New Guinea (SHWNG). At this stage, the Austronesian family tree can be roughly represented as follows (Blust 1999):



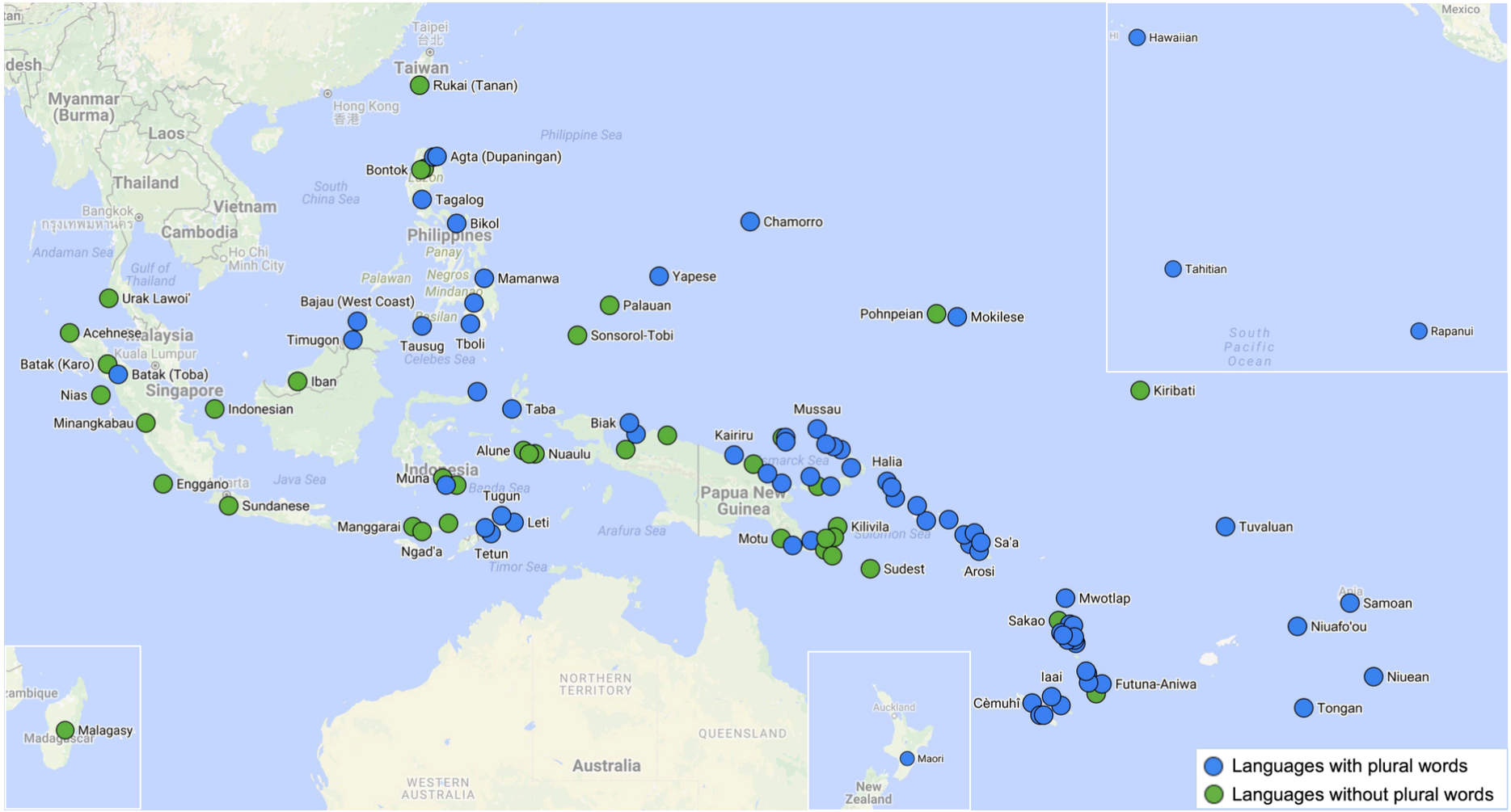
More discussion will be provided in Chapter 4.



Map 1: Distribution of Austronesian languages with plural words in Dryer (1989a)



Map 2: Distribution of Austronesian languages with plural words in Dryer (2013a)



Map 3: Distribution of all Austronesian sample languages in Dryer (2013a)

More patterns concerning the distribution of plural words can be observed if we compare the languages with plural words and those without plural words. In Map 3, blue dots refer to the Austronesian languages employing plural words, and green dots refer to the languages lacking plural words. From this map, we can see that there are some areas where the majority of languages have plural words, for instance, the Philippines and many Pacific islands; but at the same time, in other areas, languages with plural words are the minority group, for example in most parts of west Indonesia. Also, within the Oceanic group, except for those languages on the southeast tip of Papua New Guinea, almost all languages have plural words; and a calculation reveals that about 80% (55/71) Oceanic languages in Dryer's (2013a) sample have plural words.

Another noteworthy observation about plural words in Austronesian languages is that a plural word has been reconstructed in Proto Malayo-Polynesian (PMP) and its daughter subgroups. In discussing number marking in Proto Oceanic (POc), Lynch et al (2002: 74) argue that a plural word **maŋa*, which is used for marking plurality of common nouns (in contrast to human nouns), is reconstructable in POc, as a descendant from PMP **maŋa*. The reflexes of **maŋa* in Oceanic languages include Tigak *mamana*, Kara *mana*, Tolai *umana*, Halia *maman* and Nguna *maaŋa*, and they all behave like a plural word; other descendants of **maŋa* can also be found in various languages in the Philippines, as well as Wolio in Sulawesi (Lynch at al 2002: 90–91). Another source, the Austronesian Comparative Dictionary (ACD) online (Blust & Trussel 2010), confirms this reconstruction. On ACD, **maŋa* can be found as a PMP reconstruction, which is passed on to POc, as well as Proto Western Malayo-Polynesian (PWMP), Proto Central-Eastern Malayo-Polynesian (PCEMP), Proto Central Malayo-Polynesian (PCMP) and Proto Eastern Malayo-Polynesian (PEMP). The reflexes of this PMP reconstruction are illustrated by some WMP languages, such as Yami *maŋa* and Wolio *maŋa*, and two Oceanic languages, Weden *maga* and Nakanamanga *maaŋa.*

While these remarks on plural words in Austronesian languages are appealing, some cautions should also be taken. In order to draw a convincing conclusion about how plural words in Austronesian languages are distributed, one crucial basis is that the language sample needs to be balanced. However, this is not achieved in either Dryer (1989a) or Dryer (2013a). The total number of Austronesian languages examined in Dryer (1989a) is not provided; in Dryer (2013a), out of the 115 Austronesian sample languages, 71 are Oceanic languages – while in fact Oceanic languages consist of less than half of the total Austronesian languages. In comparison, whereas the total number of languages in WMP is similar to that in Oceanic, very few WMP languages outside the Philippines are selected. This does not necessarily mean that there are inherent problems with Dryer's studies, since both works do not aim at characterising plural words in Austronesian languages. The aim of Dryer (1989a) is to present a typology of plural words in languages across the world, and to examine the possible grammatical

categories of these plural words; Dryer (2013a) presents the coding of nominal plurality in world's languages, and discusses a typology of nominal plurality marking. Therefore, the observations discussed above have to be taken with caution, and a more balanced language sample is needed to depict a more objective picture and answer the question of how plural words are distributed in Austronesian languages.

With a new language sample, we can also examine the historical development of plural words in Austronesian languages. On the surface, the reconstruction of a plural word in POc seems to be in accordance with the massive presence of plural words in Oceanic languages: since a number of daughter languages as well as the Proto language share one similar grammatical feature, one might expect that the plural words in many Oceanic languages are inherited from POc. Some descendants of POc **maŋa* are indeed identified in previous studies, but do all of the Oceanic languages with plural words reflect **maŋa*? Also, since a plural word is claimed to be reconstructable in PMP, is it also the case with other Austronesian languages with plural words?

Based on this background, I revisit the typology and history of plural words in Austronesian languages, and ask the following questions:

a. How are plural words distributed in Austronesian languages?
b. Do these plural words all reflect the reconstructed form, or do they have various origins?

These two questions are the main research questions in this thesis, and they lay the basis for the following chapters.

Chapter 3. Defining plural words: A reconsideration

Before taking up the research questions, this chapter discusses another theoretical issue and reconsiders the definition of plural words and its application in previous studies.

In the definition of plural words, Dryer (1989a) considers plural words to be comparable to plural affixes in other languages. But we have also seen that the criteria he applies when identifying a plural word are quite broad: a plural word is the sole indicator of nominal plurality in a noun phrase when the noun itself is not marked, regardless of its grammatical category. Such criteria thus include the possibilities of articles and demonstratives being plural words, as discussed above. However, are such broad criteria actually helpful in cross-linguistic comparisons? In this chapter, Section 3.1 points out some problems with the way in which a plural word is identified in previous studies, and Section 3.2 presents an alternative analysis. A revised definition of plural words is given at the end of this chapter.

3.1. Plural words as a semantic category

As shown from the review in Chapter 2, plural words for Dryer are a group of morphemes that share certain similarities, rather than a grammatical category defined by a set of morphosyntactic criteria. Following his definition of plural words, Dryer (1989a) provides us with a typology of plural words based on their grammatical categories, and the results are:

- Plural words as numerals
- Plural words as articles
- Plural words as grammatical number words
- Plural words as a one-word minor category of their own
- Plural words as a multiword minor category of their own
- Miscellaneous categories of plural words

Accordingly, the plural word *ria* in Hoava in example (8) is an article, and *mga* in Tagalog is a one-word minor category of their own. Other categories will be no further elaborated here; as Dryer (1989a: 879) remarks, "we may speak of plural words as a semantic category, there is little basis for using the term as a syntactic category… at best, the term would be appropriate as a universal label for the one-word universal category of plural words". The same definition is also applied in Dryer (2013a).

Nevertheless, such a definition and such broad criteria in identifying plural words have caused some troubles and problems. First and foremost, plural words defined by Dryer's criteria are actually present in a great number of languages, including English. Consider an English example (12) offered by Corbett (2000: 136):

(12) *Those sheep are doing nothing about it.*

In this sentence, the noun *sheep* is not marked for number, but its plurality can be traced from the demonstrative *those*. The demonstrative is also the sole indicator of nominal plurality in the noun phrase *those sheep*. If we apply Dryer's criteria, *those* in this sentence can be thought of being a plural word.[5] It is comparable to the plural demonstrative *=ide* in Kokota (as illustrated by example 9, repeated here as example 13), since we can consider that all nouns in Kokota are not marked for number (just like *sheep* in English), and *=ide* is the only indicator of nominal plurality. If we follow this analysis, we would need to take English as a language with a plural word; and as a result, the value for nominal plurality marking in many other languages also need to be re-examined.

(13) Kokota (Oceanic, Austronesian)

a. *kame=ḡu=ine*
arm=1SG=this.PROX
'this hand of mine'

b. *kame=ḡu=**ide***
arm=1SG=these.PROX
'these hands of mine' (Palmer 2009: 84)

Another problem ensuing from this analysis is that the plural words identified by Dryer (1989a; 2013a) cannot be easily taken in cross-linguistic comparisons. The 22 Austronesian languages with plural words in Dryer (1989a), and the 71 Austronesian languages with plural words in Dryer (2013a) include all the possibilities presented above: a plural word might be a special word which consists of a grammatical category on its own in one language[6], while a plural article in another language. When we are presented with these languages with plural words, it would be very difficult to conduct

[5] In a similar way, Dryer (1989a: 873–874) himself also gives an example of French. As Dryer argues, spoken French has lost the plural suffixes on nouns, thus the article *les* is the only indicator of plurality in example (ii) and fits into his criteria of plural words.

i. *la* *pomme*
ART.F apple
'the apple'

ii. *les* *pommes*
ART.PL apple
'the apples'

[6] I call this type of plural words *pure plural words*.

any further comparisons, because a comparison between two plural words from the same grammatical category can be very different from a comparison between a pure plural word and a plural article. Comparing a plural word like *=ide* in Kokota and a plural word like *mga* in Tagalog is similar to comparing *those* in English and *mga* in Tagalog. Most likely they will have different origins, and they have different syntactic properties, etc. But since we know that *those* and *mga* are essentially different from each other, one being a demonstrative and the other being a pure plural word, such comparisons do not yield many useful results.

Lastly, some borderline cases cannot be easily diagnosed by using Dryer's definition. The example from Manam illustrates such a case.

(14) Manam (Oceanic, Austronesian)

a. *áine ŋára*
   woman that
   'that woman'

b. *áine ŋára-di*
   woman that-3PL.AD
   'those women' (Lichtenberk 1983: 267)

The typology of coding of nominal plurality in Dryer (2013a) starts with the position of the plural marker, i.e. on the nouns or on the phrase level. However, such a dichotomy made alongside the position of the plural marker can be problematic. In Manam, nominal number is also not marked on the noun itself; but at the same time, it is not marked on the noun phrase by a separate word either. The nominal plurality of this phrase is marked by a suffix *-di* on the demonstrative *ŋára*. In this case, shall we take *ŋára-di* as a plural word, or *-di* as a plural suffix? Either choice does not seem to offer a good explanation. On one hand, *ŋára-di* seems to fit into Dryer's criteria of plural words because it is a demonstrative which serves the function of the sole plural marker in this phrase; but in Manam, the suffix *-di* is not only used with demonstratives but also adjectives, as shown in example (15).[7] Then should the adjective *másare-di* 'broken' be considered a plural word? Such analysis is certainly questionable. On the other hand, *-di* itself cannot be called a plural suffix either, because a plural suffix has to be attached to the noun.

[7] The same affix is also used on the verbs in Manam to mark the plurality of the agent or patient.

(15) Manam (Oceanic, Austronesian)

a. *bóadi* *másare*
pot broken
'broken pot'

b. *bóadi* *másare-di*
pot broken-3PL.AD
'broken pots' (Lichtenberk 1983: 318)

Cases like this are thus hard to be dealt with in line with Dryer's typology of coding of nominal plurality. The value of coding of nominal plurality for Manam in Dryer (2013a) is *no plural*,[8] but nominal plurality in Manam can actually be marked in a noun phrase by using other words, thus this value also seems problematic.

3.2. Plural words as a syntactic category

Given the problems discussed above, it is necessary to re-define the concept of *plural words* and seek another basis on which plural words can be distinguished from other ways of nominal plurality marking.

An alternative way of analysing nominal plurality can be found in Corbett (2000) and Kibort & Corbett (2008). They also classify expressions of nominal number into three major groups based on the position where the number marking occurs: number expressed on the noun/nominal element, on or in the noun phrase, or on the verb. An obvious difference of such a classification from Dryer's is that it takes into account the fact that nominal number is not necessarily expressed in or on the noun phrase level, but can also be expressed outside the noun phrase, i.e. on the verbs, through agreement.[9] More importantly, the crucial difference between various ways of nominal number marking identified by Corbett (2000) does not hinge on their positions, but on the grammatical methods through which number is marked. Hence, regardless of the loci of number markers, coding of nominal plurality can be grouped into four types:

- Special number words
- A variety of morphological means
- Lexical means

[8] In Dryer (2013a), the value *no plural* is given to the languages that do not have morphological plurality marking or plural words (and clitics).

[9] Dryer (2013a) surely considers the possibilities of nominal plurality being marked on verbs as well. In discussing the value of *no plural* in his sample languages, he notes that "although such languages may simply not indicate plurality at all, the plurality of nominal referents is coded on the verb if the nominal is an argument of the verb and if the language is one that codes the number of that particular argument on the verb". But plurality marked on verbs is not included in his typology.

- Syntactic means

Some similarities to Dryer's classification can be found: Kibort & Corbett (2008) also identify a particular type of nominal number marking realised by using special number words, and *number words* in this sense are similar to Dryer's *plural words*. Morphological means are also clearly identified; but Corbett also includes lexical means and syntactic means. The meaning of *lexical means* is self-evident: in some languages, or in certain lexemes in a language, a noun itself encodes nominal number by a purely lexical manner. For instance, the plurality of *teeth* in English is lexically marked, and the pluralisation of the singular form *tooth* does not comply with the general rule by adding suffix *-s* or *-es* in English.

A fourth way of coding nominal number is through syntactic means, or in other words, agreement. As mentioned above, verbs are another source from which nominal number can be inferred. In many languages, verbs can or obligatorily need to be in agreement with the nominal elements. For example, in English, the copula *be* has to agree with the subject in number and person, thus from the conjugated form *is*, we can know that the subject is a third person singular noun or pronoun. For other verbs, third person singular subject can also be inferred from the suffix *-s* on the verb. Note that nominal number marking on verbs has to be distinguished from verbal number; it does not encode multiple events, but still refer to the quantity of the nominal elements. Examples of verbs in agreement with nouns in number are found in other languages as well, as in (16). In Amele, verbs agree with the number of the subject, marked by *-i-*, *-si-*, *-ig-* for singular, dual and plural number respectively. It is noteworthy that Amele also features number words, which are optional.

(16) Amele (Madang, Trans New Guinea)

a. *Dana* *(uqa)* *ho-**i**-a*
man 3SG come-3SG-TODAY'S.PAST
'The man came.'

b. *Dana* *(ale)* *ho-**si**-a*
man 3DU come-3DU-TODAY'S.PAST
'The two men came.'

c. *Dana* *(age)* *ho-**ig**-a*
man 3PL come-3PL-TODAY'S.PAST
'The men came.' (J. Roberts 1987, cited from Corbett 2000: 137)

Syntactic means are not only confined to marking nominal plurality on verbs, but also on demonstratives, articles, adjectives, pronouns and many other elements (Kibort & Corbett 2008). This is the aspect where Kibort & Corbett (2008) are fundamentally

different from Dryer: Dryer considers nominal plurality marked on demonstratives or articles to be similar to plural words, since they are number markers on the phrase level; but for Corbett, nominal plurality marked on demonstratives or articles is similar to that marked on the verbs, because both cases involve agreement. Adjectives can be another locus where nominal number is marked. It is well-known that in many Indo-European languages, in addition to demonstratives and articles, adjectives are also in agreement with nouns, as the Spanish example in (17). Similarly, nominal number in Kove, an Oceanic language, can also be expressed on adjectives by adding a third person plural object suffix, as illustrated in (18).

(17) Spanish (Romance, Indo-European)

a.	*el*	*amable*	*profesor*
	ART.SG.M	kind	teacher
	'the kind	teacher'	
b.	*los*	*amable-**s***	*profesor-es*
	ART.PL.M	kind-PL	teacher-PL

'the kind teachers' (my own knowledge)

(18) Kove (Oceanic, Austronesian)

niu	*moho-**ri***
coconut	old-3PL.OBJ

'old coconuts' (Sato 2013: 135)

Using Corbett's typology of nominal plurality, the number in the English example (12) above (repeated here as example 19) can be analysed as follows.

(19) *Those sheep are doing nothing about it.* (Corbett 2000: 136)

While the noun *sheep* is not marked for number, its plurality can be inferred from two elements – the verb *are* and the demonstrative *those*. Both elements are in their plural forms because they need to agree with the number of the noun, which is a syntactic rule in English. This analysis can also solve the dilemma of nominal plurality marking in Manam, as discussed above in example (14) and (15). While it is inaccurate to claim that the plural marker *-di* in Manam is a plural suffix, or that adjectives or demonstratives marked by *-di* are plural words, we can say that *-di* in Manam is a plural marker realised through syntactic means, and it can be attached to demonstratives, adjectives or verbs.

Let us return to the definition of *plural words*. Both Dryer and Corbett mention *plural words* or *number words* in their classification of nominal plurality marking, but

the actual definition of such words is different. While Dryer sees plural articles or demonstratives as subtypes of plural words, Corbett considers them to be similar to nominal plurality marked on verbs, both of which are operated by syntactic agreement. In this sense, *plural words* (or broadly speaking *number words*) in Corbett's definition are similar to the one-word minor category of plural words identified by Dryer. In Corbett's analysis, the crucial difference between nominal number expressed by syntactic means and other methods is that number expressed through agreement is not inherent to the noun, but contextual, depending on the phrasal or clausal structures.[10] Therefore, *number words* are also inherent, since they do not show agreement with other elements in the phrase and are not governed by syntactic rules.

This narrower definition of plural words can help us avoid the problems that we might encounter in Dryer's analysis. The merit of Dryer (1989a) is that it offers us an overview of how nominal number can be marked by words from different grammatical categories in a noun phrase, but as I have shown, a dichotomy between inside or outside noun phrases is not very favourable, since many similarities can be found between plurality marked on articles, demonstratives, adjectives (inside noun phrases) on one hand, and plurality marked on the verbs (outside noun phrases) on the other hand.

Lastly, if we apply a narrower definition for plural words, I do not see the reason why plural words have to be the sole indicators of nominal plurality in noun phrases. Different means of nominal plurality marking might co-occur, and plural words can also operate alongside other means. Example (20) from Unua illustrates such an example, where the plural word *rin* co-occurs with other plural markers (*ra-* and *re-*) cross-referenced on the verb. Even within a noun phrase, as (21) shows, a pure plural word *ira* can accompany other morphological plurality markers (in this case, reduplication).

(20) Unua (Oceanic, Austronesian)

Dabos ***rin*** ***ra-****vra* ***re-****b-ke-i* *xai.*

stranger PL 3PL-want 3PL-IRR-see-TR 2SG

'Strangers want to see you.' (Pearce 2015: 190)

(21) Raga (Oceanic, Austronesian)

Ira ***da-****daulato* *mai* ***ira*** ***mwal-****mwalagelo* *mai* ***ira*** *natu-ri-rigi.*

PL RDP-girl and PL RDP-boy and PL child-RDP-small

'the young girls, the young boys and the children.' (Vari-Bogiri 2011: 82)

All being said, for the purposes of the current study, Corbett's classification of nominal

[10] For the discussion about inherent and contextual, see Corbett (2006: 123–124).

plurality and his way of distinguishing number words will be applied. The term *plural words* I am using in the thesis thus resembles the special number word identified by Corbett, or the one-word minor category of plural words in Dryer's sense. It can be defined as follow. Plural words are inherent plural markers on the noun phrase level which have the shape of separate words. They are comparable to plural affixes in other languages in the sense that their main function is to express nominal plurality, but plural words do not need to be the sole indicators of nominal plurality in noun phrases.

Chapter 4. Methodology and samples

This chapter discusses the methodology and presents the language sample that I will investigate. Without knowing which languages feature plural words beforehand, a prior task will be to give a typological account of coding of nominal plurality in Austronesian languages as a whole, and then sift out the languages employing plural words. In this way, the geographical distribution of plural words in Austronesian languages can also be characterised.

This chapter is organised as follows. Section 4.1 describes the sampling method, and the sample languages are presented in Section 4.2. Values for the sample languages as well as some issues in defining the values are discussed in Section 4.3, followed by a list of languages employing plural words in my sample in Section 4.4.

4.1. Sampling method

In this study, I aim at exploring in which Austronesian languages plural words can be found and observing the distribution of plural words, thus the language sample should be a variety sample, which emphasises on a maximum of linguistic diversity.

Differing from a large-scale typological study which extracts samples from world's languages, the current study only addresses one language family, which means that all languages in the potential sample will be genetically related to some extent. Nevertheless, even within one language family, the genealogical distance between languages still varies considerably. If we compare all Austronesian languages to world's languages, then the subgroups in the Austronesian family would resemble different language families. As we want to find the maximal variations of coding of nominal plurality in this particular language family, and minimise the influence of close genealogical relatedness, the genetic distance between the languages selected in the sample should be maximised. A variety sample is thus still in request, which should be ideally selected based on the Diversity Value (DV) sampling method developed by Rijkhoff et al. (1993) and Rijkhoff & Bakker (1998). However, below I would argue that the difficulties and potential problems with the DV sampling method suggest that it is not very practical in this case. The sample selected for the present study is a proportionally representative sample. I will explain the exact methodology and some issues in selecting samples in the following sections.

4.1.1. Diversity Value (DV) sampling method and its problems

Here I will just briefly highlight some main arguments of the DV sampling method; for a detailed description, see Rijkhoff et al. (1993) and Rijkhoff & Bakker (1998).

The crucial point of the DV sampling method is that the number of sample languages we choose from each language family should be based on linguistic diversity,

rather than the total number of languages, because these two concepts do not always correlate. It is assumed that languages on a higher level of the family tree are structurally more different from each other than those on a lower level. Rijkhoff et al. (1993: 176) give an example, showing that there is more diversity in Afro-Asiatic with 258 languages than in Bantu with 500 languages. Thus a diversity value should be calculated based on the depth and width of each language family, and the number of languages chosen from each family should be determined propotionally by the diversity value. Within each family, the same method should be recursively applied in order to decide on the number of languages selected from each subgroup.

Theoretically, the DV sampling method is indeed the ideal way to select sample languages for the current research. However, in applying the DV sampling method, various problems arise. Firstly, one fundamental basis in using the DV sampling method is that there should be at least a widely-accepted language classification, because the diversity value of a language family or a subgroup is determined by the structure of the family tree. However, this is certainly not the case for Austronesian languages. As I will explain below, even high-level language classifications are still under much debate in this particular language family.

Figure 1 below represents the major branches of a tentative family tree of Austronesian languages.

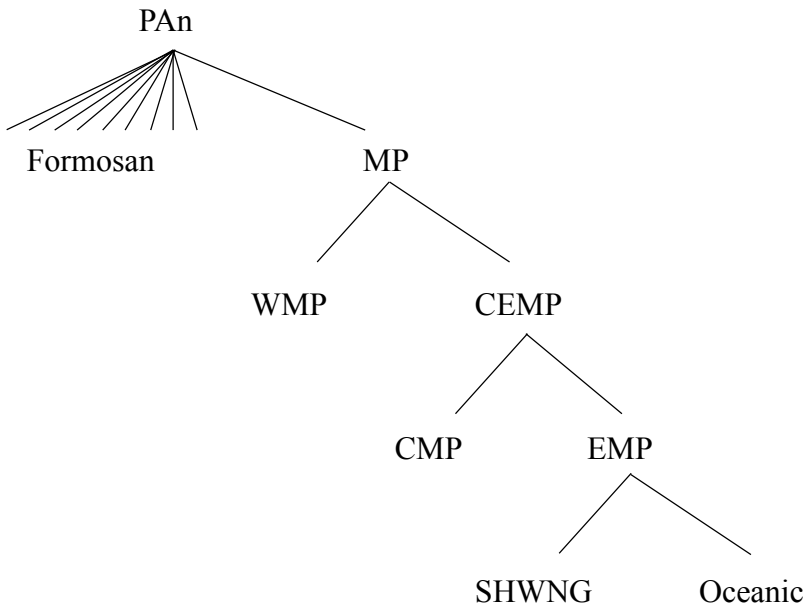


Figure 1: A tentative family tree of Austronesian languages (Blust 1999)[11]

[11] Abbreviations in the figure: PAn – Proto Austronesian; MP – Malayo-Polynesian; WMP – Western Malayo-Polynesian; CEMP – Central-Eastern Malayo-Polynesian; CMP – Central Malayo-Polynesian; EMP – Eastern Malayo-Polynesian; SHWNG – South Halmahera-West New Guinea

Even though the formulation of the Austronesian language family as a whole is generally uncontroversial, the genetic relationships among the daughter languages are far from certain. In the family tree above, the only well-founded and extensively accepted sub-families are Malayo-Polynesian (MP) languages, South Halmahera-West New Guinea (SHWNG) languages and the Oceanic languages (see e.g. Adelaar 2005a). The establishment of Central Malayo-Polynesian (CMP), Eastern Malayo-Polynesian (EMP) and Central-Eastern Malayo-Polynesian (CEMP) is argued by Blust (1978; 1993). However, the problem is that the diagnostic innovations for the subgrouping of CEMP or CMP are either not present in all languages, or not exclusive to the languages in the argued groups (see Ross 1995; Adelaar 2005a for a review, and Donohue & Grimes 2008 for more counterarguments). There is no clear phonological evidence for the grouping of EMP either (Ross 1995: 84–85). WMP, on the other hand, is only negatively defined: it refers to the languages that are not in CEMP, and there is no proper historical reconstruction as a foundation for its existence as a separate subgroup.

As for the subgrouping of Formosan languages, many questions still remain. Blust (1999) makes a classification of nine primary branches, which is a modification of Blust (1977) following Ferrell (1969), who originally makes a three-way classification: Atayalic, Tsouic and Paiwanic. Recently, new arguments for the subgrouping of Formosan languages and their relationship with MP are proposed by Ross (2009; 2012), who considers four primary branches on the first order of Austronesian family tree: Puyuma, Tsou, Rukai and Nuclear Austronesian. The detailed arguments for Austronesian subgrouping will not be elaborated here, but it should be emphasised that the classification of the Austronesian language family is still controversial and open for discussion, and such an uncertain subgrouping unavoidably undermines the applicability of the DV sampling method.

Secondly, the assumption that languages related to each other on a higher level have more structural diversity also requires reconsiderations. A parent language does not always split into two or more daughter languages due to seperation as depicted by the tree model in language classification (and this has often been acknowledged as a problem in the tree model, see François 2015 for a recent review); daughter languages might also arise via dialect differentiation. Ross (1995: 45–47) thus makes a distinction between a subgroup and a linkage, the second kind referring to languages arisen from a chain of diverse dialects. Pawley (1999: 130) notes that “a linkage is formed when a chain of diverse dialects persists for long enough for innovations to diffuse across parts of the chain, in overlapping or linking patterns, without spreading across the entire dialect chain”. If a group of languages arises from dialect differentation instead of decending from a parent language, grammatical features would diffuse across the chain and result in more similarites among the languages. Thus it would be misleading to assert that languages on higher levels of the family tree always have more structural

diversity, because it is possible that these languages could have converged through later diffusion. This is precisely the case in the Austronesian family. The Formosan languages, which occupy nine first-order nodes in the Austronesian family tree according to Blust (1999), are argued to have emerged as a linkage (Ross 1995). Thus even though they are on the first level in the family tree, the diversity they represent would be less than we expect from the DV sampling method. The same accounts for WMP; the subgroups in WMP are also very likely to have arisen as a linkage (Adelaar 2005a). Hence, if the DV sampling method were applied, I would have to select one language from each of the nine Formosan branches (because they are on the first level and resemble different language families), and this would violate the principle of maximising linguistic diversity because these languages are in fact argued to be a linkage.

Thirdly, bibliographic bias is a problem that typologists cannot easily overcome. The Austronesian family has over 1200 languages, but only a fraction of them are well documented, and these are usually centred in certain areas. For instance, many recent and well-written grammars for languages in Vanuatu are accessible, but grammars for languages in Borneo or Sulawesi are strikingly sparse. Under this circumstance, even if I were to apply DV sampling method, it is likely that data needed for the sample is not available.

4.1.2. Proportionally representative sample

Given all the limitations of the DV sampling method I listed above, the language sample I collected is a proportionally representative sample.

Such a method is discussed by Bell (1978) and applied by Tomlin (1986) in his typological study on basic word orders. As the name suggests, the number of sample languages taken from each language family is proportional to the total number of languages in that family. The same method applies to each subgroup, so that each family and each subgroup is represented. Even though this method has been criticised by Rijkhoff et al. (1993) and Dryer (1989b), it is the most appropriate and practical sampling method for the current study.

The state-of-the-art Austronesian classification is reviewed by Kikusawa (2015). Except for the subgrouping of Formosan languages, Blust's (1999) major branches are still now widely accepted and frequently cited, thus here I follow his classification.[12] The total number of Austronesian languages and the number of languages in each subgroup are drawn from *Glottolog 2.7* (Hammarström et al. 2016). In total, 1274 Austronesian languages have been identified so far, and the number of languages in

[12] And due to the reason that Formosan languages are considered to be a linkage (discussed above) and the total number of Formosan languages is relatively small, they are not very relevant here.

each branch is shown in Figure 2.

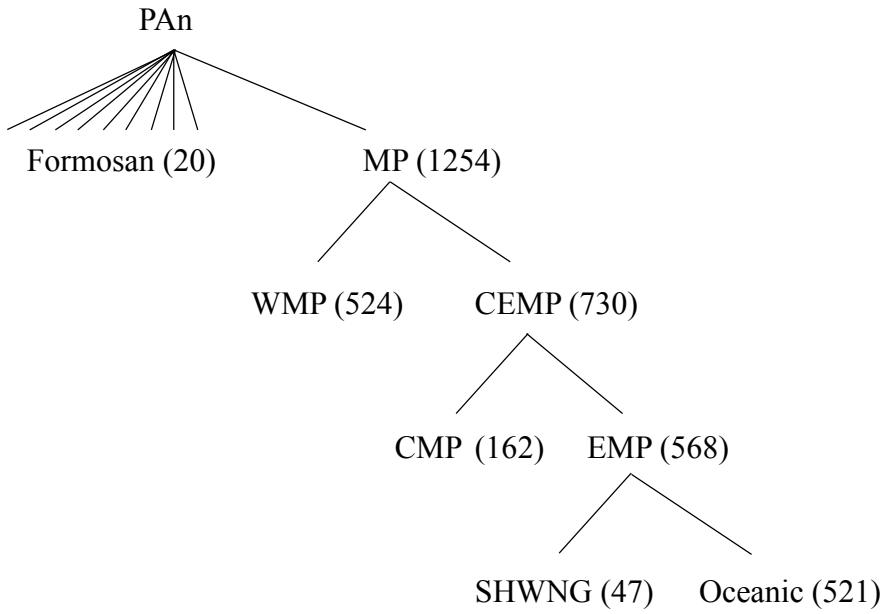


Figure 2: A tentative family tree of Austronesian languages (Blust 1999) with the number of languages in each branch

For lower-level subgrouping in the Austronesian family and the number of languages in each subgroup, I generally follow the classification and data on *Glottolog*. An exception is made for the classification of WMP, for which I take Adelaar's (2005a) subgrouping proposal, since it has a wider recognition. The subgrouping of WMP is presented in Table 1 below.

Table 1: Subgrouping of WMP languages (Adelaar 2005a)

1	Languages in Philippines	13	West Barito
2	Chamorro	14	Lampung
3	Palauan	15	Rejang
4	Sama-Bajau	16	Northwest Sumatra/Barrier Islands
5	Malayo-Sumbawan	17	Tomini-Tolitoli
6	Javanese	18	Kaili-Pamona
7	Moken-Moklen	19	Saluan
8	North Bornean	20	Bungku-Tokali
9	Kayanic	21	Muna-Buton
10	Land Dayak	22	Wolio-Wotu
11	East Barito	23	South Sulawesi
12	Barito-Mahakam		

When the classification of WMP in Adelaar (2005a) does not match that on *Glottolog*, I take the related subgroups on *Glottolog* into consideration and re-calculate the number of languages. For instance, Adelaar (2005a) considers languages in Philippines as one subgroup (as shown in Table 1 above), but *Glottolog* splits them into Batanic (2), Bilic (5), Central Luzon (10), Greater Central Philippines (95), Minasahan (5), Northern Luzon (52) and Sangiric (5). Thus the number of languages in Adelaar's (2005a) Philippine group would be roughly 174.

I choose 10 per cent of all Austronesian languages, therefore arriving at a 128-language sample. The number of languages from each branch is proportionally calculated; accordingly, there are two Formosan languages, 126 MP languages, among which 53 languages belong to WMP and 73 belong to CEMP, etc. The same procedure applies to each branch. Whenever possible, I choose languages from different subgroups in order to maximise their genetic distance. Still, some manual adjustments are required. On some occasions, the number of subgroups is larger than the number of sample languages that I am supposed to select, and one group might contain a substantial amount of daughter languages while there are language isolates on the same level. Under these circumstances, I opt for language isolates (also depending on the availability of the source), therefore the number of languages chosen might not be exactly 10 per cent of the total number of languages in that subgroup.

Bibliographic bias still exists. There are cases when no data is available at all for certain subgroups (e.g. Barito-Mahakam, WMP), then I manually adjust the sample by adding languages from other groups.

In a nutshell, the sampling method used here takes linguistic diversity into account and tries to maximise the diversity in the sample. Due to various factors beyond my control it may not be the ideal sample, but it shall sufficiently serve the purpose of the current study.

4.2. Language sample

By applying the methodology described above, I select 128 languages as my final sample, and they are presented in Table 2 to Table 6 with their names and primary classifications. For the full reference please refer to Appendix.

Table 2: Formosan languages in the sample

Number	*Name*	*Primary classification*
1	Rukai	Formosan
2	Puyuma	Formosan

Table 3: Western Malayo-Polynesian languages in the sample

Number	*Name*	*Primary classification*
3	Tagalog	Philippines – Greater Central Philippines (GCP) – Central Philippine
4	Bikol	Philippines – GCP – Central Philippine
5	Mansakan	Philippines – GCP – Central Philippine
6	Mamanwa	Philippines – GCP – Central Philippine
7	Cebuano	Philippines – GCP – Central Philippine
8	Central Tagbanwa	Philippines – GCP – Palawanic
9	Manobo	Philippines – GCP – Manobo
10	Subanen	Philippines – GCP – Subanen
11	Bontok	Philippines – North Luzon
12	Kankanaey	Philippines – North Luzon
13	Ibaloy	Philippines – North Luzon
14	Ilocano	Philippines – North Luzon
15	Dupangingan Agta	Philippines – North Luzon
16	Ayta Abenlen	Philippines – Central Luzon
17	Tboli	Philippines – Bilic
18	Tondano	Philippines – Minahasan
19	Chamorro	Chamorro
20	Palauan	Palauan
21	West Coast Bajau	Sama-Bajau
22	Madurese	Malayo-Sumbawan (MS) – Madurese
23	Acehnese	MS – North and East Malayo-Sumbawan (NEMS) – Aceh-Cham
24	Balinese	MS – NEMS – Bali-Sasak-Sumbawa
25	Indonesian	MS – NEMS – Malayic
26	Mualang	MS – NEMS – Malayic
27	Papuan Malay	MS – NEMS – Malayic
28	Salako	MS – NEMS – Malayic
29	Javanese	Javanese
30	Moklen	Moken-Moklen
31	Bulungan	North Borneo (NB) – Bulongan
32	Ida’an	NB – Northeast Sabahan
33	Belait	NB – North Sarawakan
34	Melanau	NB – Sarawak-Melanau-Kajang
35	Murut	NB – Southwest Sabahan

36	Tatana	NB – Southwest Sabahan
37	Kimaragang	NB – Southwest Sabahan
38	Bundu Dusun	NB – Southwest Sabahan
39	Kayan	Kayanic
40	Matéq	Land Dayak
41	Maanyan	East Barito
42	Malagasy	East Barito
43	Seruyan	West-Barito
44	Lampung	Lampung
45	Rejang	Rejang
46	Toba Batak	Northwest Sumatra/Barrier Islands
47	Pendau	Tomini-Tolitoli
48	Kaili	Kaili-Pamona
49	Balantak	Saluan
50	Mori	Bungku-Tokali
51	Tukang Besi	Muna-Buton
52	Wolio	Wolio-Wotu
53	Buginese	South Sulawesi
54	Makassarese	South Sulawesi
55	Pitu Ulunna Salu	South Sulawesi

Table 4: Central Malayo-Polynesian languages in the sample

Number	*Name*	*Primary classification*
56	Batuley	Aru
57	Donggo	Bima
58	Mono	Central Maluku (CM) – East Central Maluku (ECM) – Banda-Geser
59	Nuaulu	CM – ECM – Nunusaku
60	Alune	CM – ECM – Nunusaku
61	Larike	CM – ECM – Nunusaku
62	Buru	CM – West Central Maluku
63	Lamaholot Lewotobi	Flores-Lembata
64	Kambera	Flores-Sumba-Hawu
65	Kéo	Flores-Sumba-Hawu
66	Kei	Kei-Tanimbar
67	Selaru	Southern Southeast Maluku
68	Leti	Timoric A – Eastern Timoric A

69	Tetun Dili	Timoric A – Central Extra-Ramelaic
70	Tugun	Timoric A – Northern Timoric A
71	Southern Mambai	Timoric B

Table 5: South Halmahera-West New Guinea languages in the sample

Number	*Name*	*Primary classification*
72	Biak	Cenderawasih Bay
73	Ambai	Cenderawasih Bay
74	Taba	Raja Ampat-South Halmahera
75	Warembori	Lower Mamberamo
76	Irarutu	Nabi-Irarutu

Table 6: Oceanic languages in the sample

Number	*Name*	*Primary classification*
77	Wuvulu	Admiralty Islands
78	Paluai	Admiralty Islands
79	Loniu	Admiralty Islands
80	Vaeakau-Taumako	Central Pacific (CP) – East Fijian-Polynesian (EFP) – Polynesian
81	Samoan	CP – EFP – Polynesian
82	Hawaiian	CP – EFP – Polynesian
83	Nadrogâ	CP – West Fijian
84	Dehu	Loyalty Islands
85	Ponapean	Micronesian
86	Satawalese	Micronesian
87	Unua	North and Central Vanuatu (NCV) – Central Vanuatu (CV) - Malakula
88	Neve’ei	NCV – CV – Malakula
89	Tape	NCV – CV – Malakula
90	Abma	NCV – CV – South Pentecost
91	South Efate	NCV – CV – Epi-Efate
92	Mavea	NCV – Northern Vanuatu (NV) – Espiritu Santo
93	Tamambo	NCV – NV – Espiritu Santo
94	Araki	NCV – NV – Espiritu Santo
95	Mwotlap	NCV – NV – Torres-Banks linkage
96	Raga	NCV – NV – Hano
97	Wala	Southeast Solomonic
98	Longgu	Southeast Solomonic

99	Belep	Southern Melanesian (SM) – New Caledonian (NC) – Extreme Northern
100	Tinrin	SM – NC – Southern New Caledonian
101	Cèmuhî	SM – NC – Cemuhî
102	Anejom̃	SM – South Vanuatu
103	Mussau	St. Matthias
104	Engdewu	Temotu
105	Vitu	Western Oceanic linkage (WOL) – Meso Melanesian linkage (MML) – Bali-Vitu
106	Kara-Lemakot	WOL – MML – New Ireland-Northwest Solomonic linkage (NINSL)
107	Siar	WOL – MML – NINSL
108	Ughele	WOL – MML – NINSL
109	Kokota	WOL – MML – NINSL
110	Teop	WOL – MML – NINSL
111	Nakanai	WOL – MML – Willaumez
112	Bukawa	WOL – North New Guinea linkage (NNGL) – Huon Gulf
113	Jebem	WOL – NNGL – Huon Gulf
114	Adzera	WOL – NNGL – Huon Gulf
115	Mato	WOL – NNGL – Ngero-Vitiaz linkage (NVL)
116	Kove	WOL – NNGL – NVL
117	Mangap-Mbula	WOL – NNGL – NVL
118	Lote	WOL – NNGL – NVL
119	Manam	WOL – NNGL – Schouten
120	Kairiru	WOL – NNGL – Schouten
121	Tobati	WOL – NNGL – Sarmi-Jayapura Bay
122	Maisin	WOL – Papuan Tip linkage (PTL) – Nuclear Papuan Tip linkage (NPTL)
123	Tawala	WOL – PTL – NPTL
124	Koluwawa	WOL – PTL – NPTL
125	Gapapaiwa	WOL – PTL – NPTL
126	Sinaugoro	WOL – PTL – Peripheral Papuan Tip
127	Motu	WOL – PTL – Peripheral Papuan Tip
128	Yapese	Yapesic

4.3. Data coding and values

Marking of nominal plurality of all the languages in the sample is coded. In data coding, I only consider productive ways of nominal plurality marking. Thus any special marking for certain lexical items are ruled out, and plurality marking for a closed word class (e.g. kinship terms) is also excluded. Syntactic means to express nominal plurality are excluded as well, since agreement occurs very frequently across languages, and it often co-occurs with other means of nominal plurality marking. It is found that Austronesian languages show much variation in ways to mark nominal plurality, and the possible values of nominal plurality marking in my sample are as follows.

- Plural prefix
- Plural suffix
- Plural infix
- Reduplication (full or partial)
- Plural clitic
- Plural word
- No plural
- Not clear

Some other theoretical issues in data coding should also be explained here.

Firstly, not all lexemes in one language can take nominal plurality marking. For instance, in Palauan (as in example 3, repeated here as 22), plurality is marked by the prefix *rę-*, but only for human nouns. Non-human nouns are not marked for number, thus *bilis* 'dog' cannot take the plural prefix (**rę-bilis* does not exist). In some other languages, such as Anejom̃, only animate nouns can be marked. (23) shows that non-singular animate nouns in Anejom̃ are marked by a prefix *eplu-* (often *iplu-*). Therefore, in these cases, the value I give is only for a certain word class in that particular language.

(22) Palauan (Palauan, Austronesian)
chad 'person' > ***rę-****chad* 'people'
kangkodang 'tourist' > ***rę-****kangkodang* 'tourists' (Josephs 1975: 43)

(23) Anejom̃ (Southern Melanesian, Oceanic)
a. *Et awod etwa-m̃ a Nalmunai.*
3SG.AR hit brother-your.SG SBJ Nalmunai
'Nalmunai hit your brother.'
b. *Et awod* ***elpu-****etwa-m̃ a Nalmunai.*
3SG.AR hit PL-brother-your.SG SBJ Nalmunai
'Nalmunai hit your brothers.' (Lynch 2000: 50)

There are also some languages in which different word classes are marked by different strategies. In Leti, plurality of human nouns can be marked by encliticising the third person plural clitic *=ra* on NPs featuring a deictic modifier or a possessive suffix (24a, 24b), but plurality of non-human nouns is marked by reduplication (24c).

(24) Leti (Timoric A, Central Malayo-Polynesian)

a.	p*üata*		*püat=e=**ra***
	woman	>	woman=DEX=PL
	'woman'		'the women'
b.	*isüòn-ne*		*isüòn-ne=**ra***
	witch-3SG.POSS	>	witch-3SG.POSS=PL
	'his witch'		'his witches'
c.	*kuda* 'horse'	>	***kuda-**kuda* 'horses'
	vatu '(the) stone'	>	***vatu-**vatu* '(the) stones'

(Engelenhoven 2004: 116)

Dryer (2013a) considers that nominal plurality in Leti is marked by plural clitics, but I do not see why he discards reduplication. My solution in these cases is to give two values for the coding of nominal plurality.

There are also a few languages where different ways of nominal plurality can be used regardless of word classes. For instance, nominal plurality in Puyuma (Formosan) can be marked by reduplication (25a), or a suffix *-an* (25b) or both (25c).[13] I also keep two values, reduplication and plural suffix, as the values for nominal plurality marking for Puyuma.

(25) a. *tu=lasaD-aw=dar* *i* ***TaLu-**TaLun*
3.GEN=hide-TR=FREQ LOC RDP-grass
'She hid it in the field.'

b. *ma-la-lemes* *naDu* *na* *lalak-**an***
INTR-RDP-disappear those.NOM DEF.NOM child-PL
'Those children were going to disappear.'

c. *saDu* *ku=**'ali-**'ali-**an***
many 1S.POSS=RDP-male.friend-PL
'I have a lot of friends.' (Lit. My friends are many.) (Teng 2007: 142)

Apart from having a clear value for coding of nominal plurality, few languages receive

[13] Nominal plurality in Puyuma is also often not marked (Teng 2007: 142).

the values *no plural* or *not clear*. *No plural* in my data coding means that there is no productive inherent nominal plurality marking. It is still possible that nominal plurality in that language can be inferred from other elements contextually, or certain lexemes can also be marked for plurality in certain ways. There are four languages in which nominal plurality marking cannot be determined, therefore receiving the value *not clear*, and they will be omitted in the following analysis.[14] However, both *no plural* and *not clear* remain tentative. For the coding of nominal plurality for each language, also see Appendix.

Finally, in dealing with plural clitics, I follow Dryer's suggestion and consider them to be a particular kind of plural words.

4.4. Sample languages with plural words

By applying the foregoing methodology and the criteria in identifying plural words, it is found that 54 languages in the 128-language sample employ plural words (or plural clitics), and they are listed in Table 7 below. The following investigations will be centred around these 54 sample languages and their plural words.

Table 7: Sample languages with plural words

Number	*Name*	*Primary classification*
1	Tagalog	WMP – Philippines – Greater Central Philippines (GCP) – Central Philippine (CP)
2	Bikol	WMP – Philippines – GCP – CP
3	Mansakan	WMP – Philippines – GCP – CP
4	Mamanwa	WMP – Philippines – GCP – CP
5	Cebuano	WMP – Philippines – GCP – CP
6	Central Tagbanwa	WMP – Philippines – GCP – Palawanic
7	Manobo	WMP – Philippines – GCP – Manobo
8	Subanen	WMP – Philippines – GCP – Subanen
9	Kankanaey	WMP – Philippines – North Luzon
10	Ibaloy	WMP – Philippines – North Luzon
11	Dupangingan Agta	WMP – Philippines – North Luzon
12	Tboli	WMP – Philippines – Bilic
13	Chamorro	WMP – Chamorro
14	West Coast Bajau	WMP – Sama-Bajau
15	Tatana'	WMP – North Borneo – Southwest Sabahan

[14] It is because the source does not discuss nominal number marking, and there is also no enough information for me to determine if there is any inherent nominal number marker.

16	Kimaragang	WMP – North Borneo – Southwest Sabahan
17	Toba Batak	Northwest Sumatra/Barrier Islands
18	Pendau	WMP – Tomini-Tolitoli
19	Wolio	WMP – Wolio-Wotu
20	Batuley	CMP – Aru
21	Alune	CMP – Central Maluku – East Central Maluku – Nunusaku
22	Buru	CMP – Central Maluku – West Central Maluku
23	Selaru	CMP – Southern Southeast Maluku
24	Leti	CMP – Timoric A – Eastern Timoric A
25	Tetun Dili	CMP – Timoric A – Central Extra-Ramelaic
26	Tugun	CMP – Timoric A – Northern Timoric A
27	Southern Mambai	CMP – Timoric B
28	Taba	SHWNG – Raja Ampat-South Halmahera
29	Paluai	Oceanic – Admiralty Islands
30	Loniu	Oceanic – Admiralty Islands
31	Hawaiian	Oceanic – Central Pacific – EFP – Polynesian
32	Nadrogâ	Oceanic – Central Pacific – West Fijian
33	Dehu	Oceanic – Loyalty Islands
34	Unua	Oceanic – North and Central Vanuatu (NCV) – Central Vanuatu (CV) - Malakula
35	Neve’ei	Oceanic – NCV – CV – Malakula
36	Tape	Oceanic – NCV – CV – Malakula
37	Abma	Oceanic – NCV – CV – South Pentecost
38	Mavea	Oceanic – NCV – Northern Vanuatu (NV) – Espiritu Santo
39	Araki	Oceanic – NCV – NV – Espiritu Santo
40	Mwotlap	Oceanic – NCV – NV – Torres-Banks linkage
41	Raga	Oceanic – NCV – NV – Hano
42	Wala	Oceanic – Southeast Solomonic
43	Longgu	Oceanic – Southeast Solomonic
44	Mussau	Oceanic – St. Matthias
45	Engdewu	Oceanic – Temotu
46	Kara-Lemakot	Oceanic – Western Oceanic linkage (WOL) – Meso Melanesian linkage (MML) – New Ireland-Northwest Solomonic linkage (NINSL)
47	Teop	Oceanic – WOL – MML – NINSL

48	Nakanai	Oceanic – WOL – MML – Willaumez
49	Mato	Oceanic – WOL – North New Guinea linkage (NNGL) – Ngero-Vitiaz linkage (NVL)
50	Mangap-Mbula	Oceanic – WOL – NNGL – NVL
51	Lote	Oceanic – WOL – NNGL – NVL
52	Kairiru	Oceanic – WOL – NNGL – Schouten
53	Sinaugoro	Oceanic – WOL – Papuan Tip linkage – Peripheral Papuan Tip
54	Yapese	Oceanic – Yapesic

Chapter 5. Distributions of plural words in Austronesian languages

This chapter describes the distribution of plural words in Austronesian languages based on the language sample represented in Chapter 4. The investigation will be conducted from two perspectives: I will not only investigate the genealogical and geographical distribution of languages with plural words, but also attend to how languages without plural words are distributed. In this way, we can observe if using plural words is a common way to mark nominal plurality in languages in certain subgroups or areas, and a more complete picture of plural words in Austronesian language can be presented.

Section 5.1 discusses the genealogical and geographical distribution of the 54 Austronesian languages presented in Table 7. Section 5.2 contrasts the distribution of languages with plural words and those languages without, investigating the frequency of using plural words. An interim summary is presented in Section 5.3, where I compare my results with the observations found in previous studies.

5.1. Genealogical and geographical distribution of Austronesian languages with plural words

Out of 128 languages in my sample, expect for four languages whose values of nominal plurality marking are not clear, 54 languages employ plural words. These languages are distributed across various genetic subgroups and geographical areas.

Chart 1 represents the genealogical distribution of the 54 languages in Table 7.

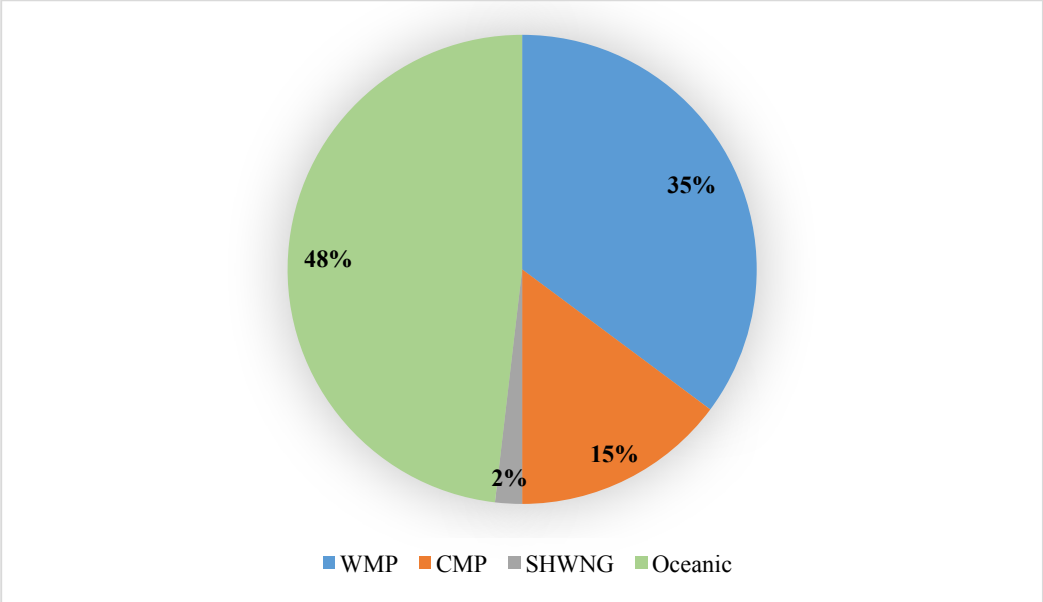


Chart 1: Genealogical distribution of Austronesian languages with plural words

As can be seen, languages with plural words are found in WMP, CMP, SHWNG and Oceanic group, and none of them is Formosan. They are not evenly distributed across these subgroups: almost half of them (26/54) are Oceanic languages; and another big portion (35%) is found in WMP, where 19 languages use plural words to denote nominal plurality. There are also eight languages from CMP employing plural words, and one from SHWNG.

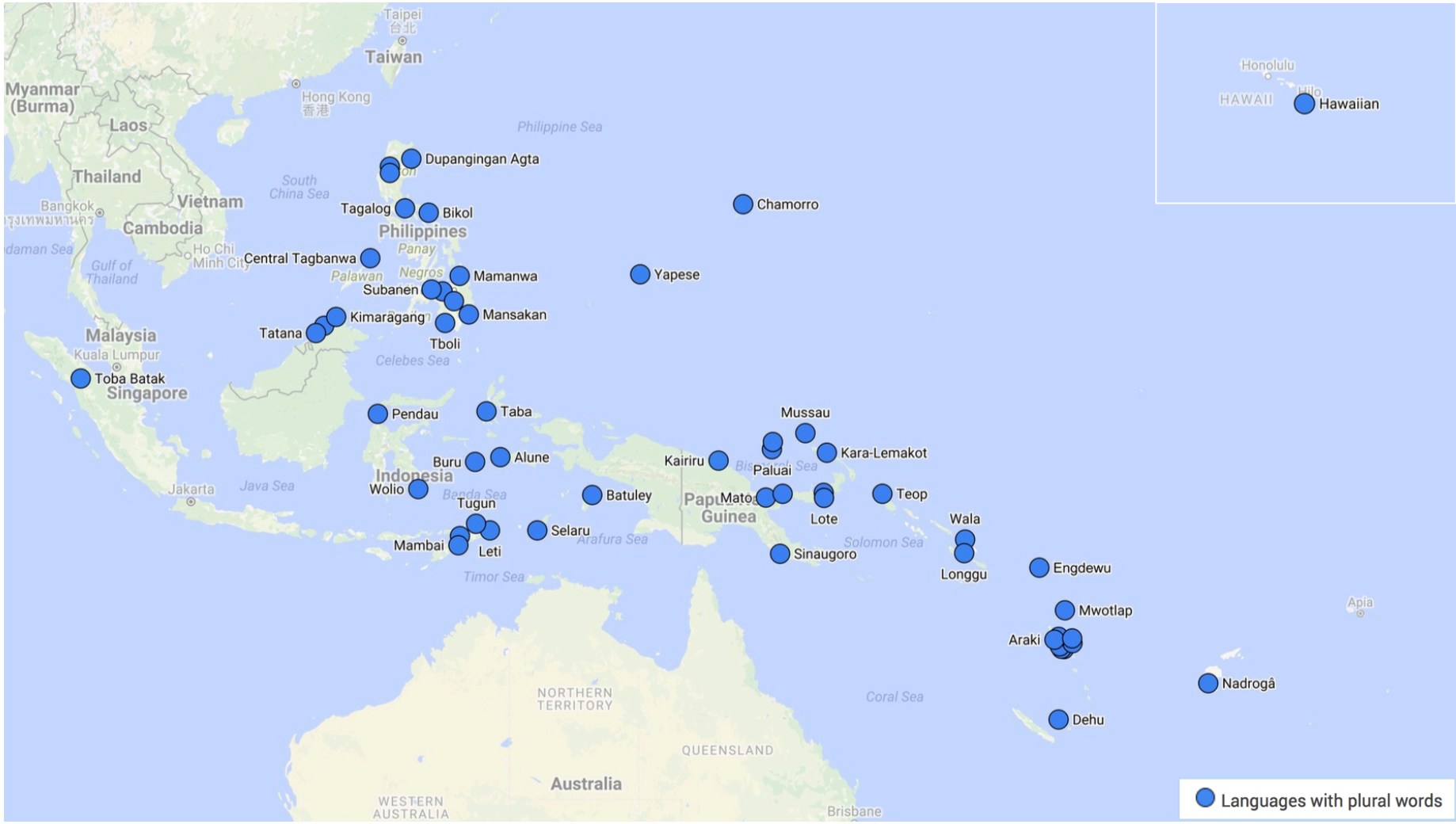
Languages with plural words are also not evenly distributed within each of these subgroups. In WMP, for instance, about two thirds (12/19) of the languages with plural words are Philippine languages, while the remaining one third consists of WMP languages from all other subgroups. In Oceanic as well, among the 26 languages with plural words, eight are in the subgroup of North and Central Vanuatu, and another eight are from Western Oceanic Linkage. Some of these languages also show close genealogical relationships on a lower level: the eight Philippine languages with plural words are either Central Philippine languages (Tagalog, Bikol, Mansaka, Mamanwa and Cebuano), or North Luzon languages (Kankanaey, Ibaloy and Dupangingan Agta). Also, four out of the eight CMP languages (Leti, Tetun Dili, Tugun and Southern Mambai) are Timoric.[15]

In addition to the dispersed yet unbalanced genealogical distribution, the geographical distribution of Austronesian languages with plural words also shows much diversity. Map 6 on the next page illustrates the geographical distribution of languages with plural words.

Languages with plural words are mostly found in several geographical areas: The Philippines, east costal line of Papua New Guinea and islands of Pacific Ocean. Some other areas where plural words are also commonly found include north Borneo (Sabah), represented by Tatana', Kimaragang, West Coast Bajau, as well as east Indonesia, mostly on Timor and its neighbouring islands, represented by the Timoric languages mentioned above. A few languages with plural words are found in Sulawesi, i.e. Pendau and Wolio. There are also several isolated dots on the Pacific Ocean, e.g. Yapese, Chamorro, and Hawaiian.

By and large, the geographical distribution of these languages corresponds with their genealogical distribution pattern: languages in the Philippines, north Borneo as well as Sulawesi belong to the WMP group, and languages on the Pacific Ocean and east Papua New Guinea are Oceanic languages. In the area in between, languages in east Indonesia belong to the CMP group; and the only SHWNG sample, Taba, is spoken in North Maluku, close to the tip of the bird's head of Papua New Guinea.

[15] As discussed in Chapter 4, many of these subgrouping proposals are putative.



Map 4: Distribution of Austronesian languages with plural words

5.2. Frequency of using plural words

The previous section offers a description of the genealogical and geographical distribution of Austronesian languages with plural words. However, such descriptions remain one-sided: since the number of sample languages from each subgroup is different, the absolute number of languages with plural words does not necessarily imply the frequency of using plural words in languages of a certain subgroup. It has been shown that most Austronesian languages with plural words are either Oceanic or WMP, but since the total number of Oceanic and WMP languages is considerably big compared to CMP or SHWNG, it does not necessarily mean that plural words are most frequently found in Oceanic or WMP languages.

To address this issue, I contrast the distribution of Austronesian languages with plural words and those without in this section. Chart 2 below contrasts their genealogical distributions. Four languages whose coding of nominal plurality is not clear are excluded here, and three of them are from WMP and one from SHWNG. Blue bars represent Austronesian languages with plural words, while green bars represent languages without plural words.

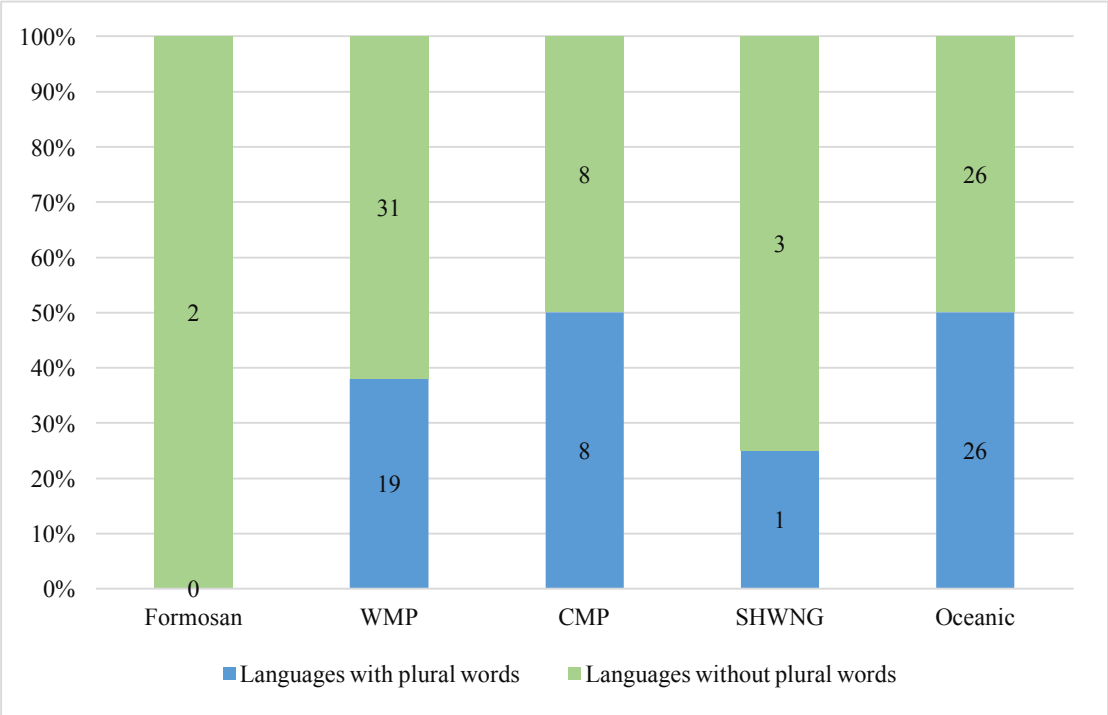


Chart 2: Percentage of languages with plural words in different subgroups

Chart 2 suggests that overall, languages without plural words still outnumber languages with plural words in most subgroups. In terms of frequency, languages with plural words have the highest percentage in CMP and Oceanic; in both groups half of the languages use plural words to mark nominal plurality. It is also shown that even though

26 out of 54 languages Austronesian languages with plural words are Oceanic, there are another 26 Oceanic languages without plural words. In WMP as well, although the total number of languages with plural words is quite large, they only account for about one third of the total number (19/50). In contrast, in CMP, while the absolute number of languages with plural words is only eight, it makes up half of the total sample languages. The only SHWNG language with plural word represents 25%.

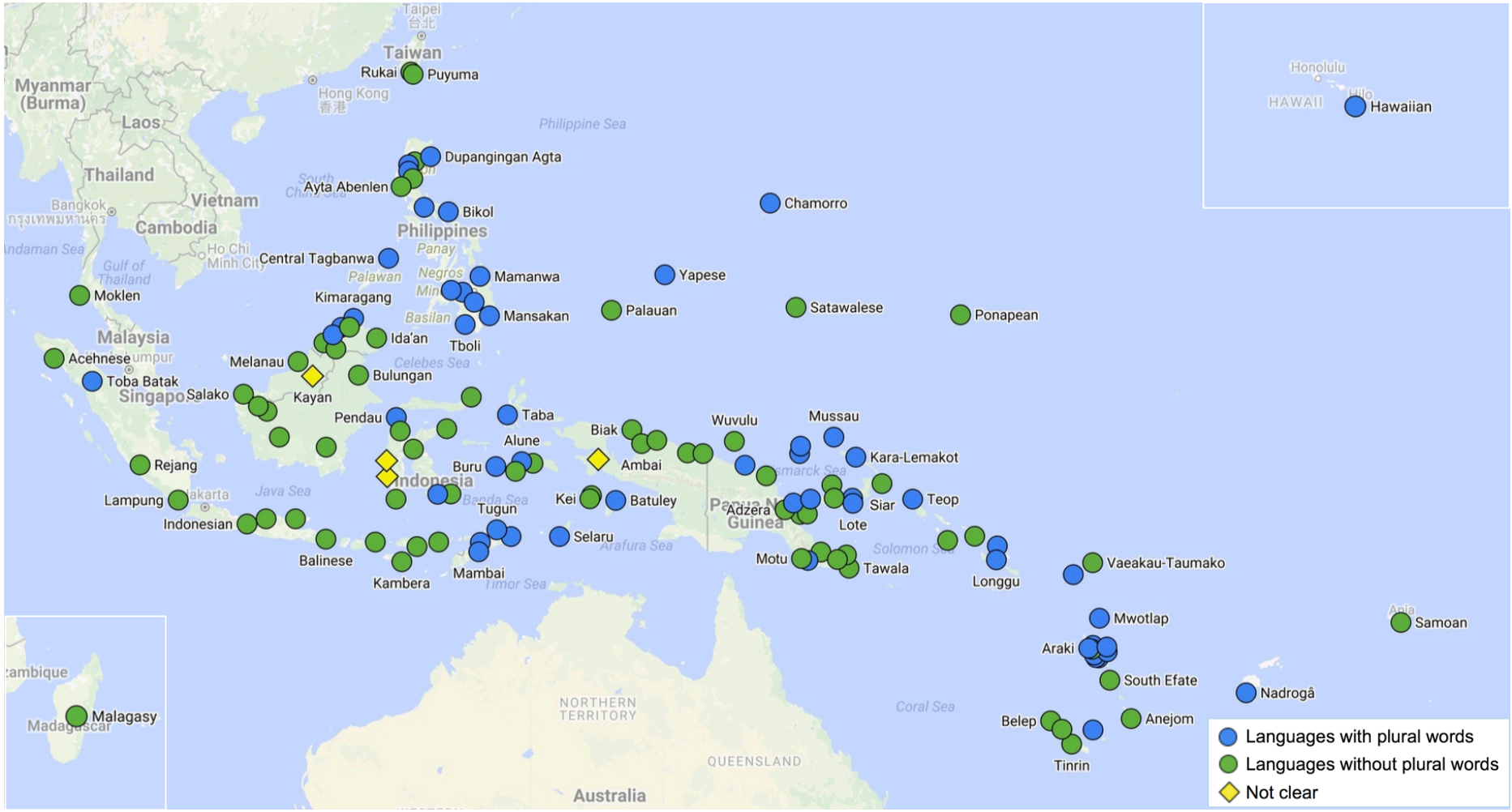
For lower-level branches, there are certain subgroups in which plural words are used by the majority of languages. In the Philippine languages, for example, 12 out of 16 languages employ plural words, and all of the eight Greater Central Philippines languages have plural words. A similar dominance of using plural words to mark nominal plurality is also found in other branches. Eight out of ten North and Central Vanuatu languages use plural words, and all Timoric languages have plural words.

We could now conclude that in terms of genealogical affiliations, languages with plural words are quite often found in CMP and Oceanic, but not ubiquitous – they are used by half of the languages in these two subgroups. Plural words show fairly little presence in SHWNG, and not at all in Formosan languages.

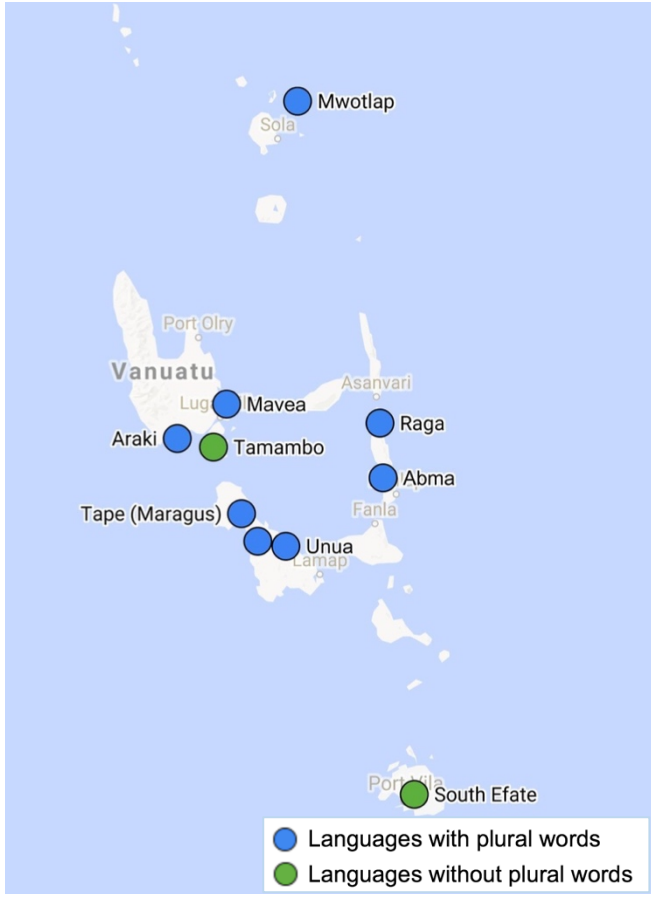
The geographical distribution of plural words can also be demonstrated in a map where both languages with and without plural words are represented. In Map 5, blue dots still represent languages with plural words, while green dots represent languages without plural words. Four languages whose values are not clear are marked by yellow diamonds.

It can be observed from Map 5 that languages with plural words are particularly common in contrast to languages without plural words in certain areas: The Philippines, east Indonesia and Vanuatu. In Vanuatu, for instance, eight out of ten languages use plural words, as illustrated by Map 6. Such a high presence of plural words can also be found in the Philippines, where 12 out of 15 languages employ plural words.

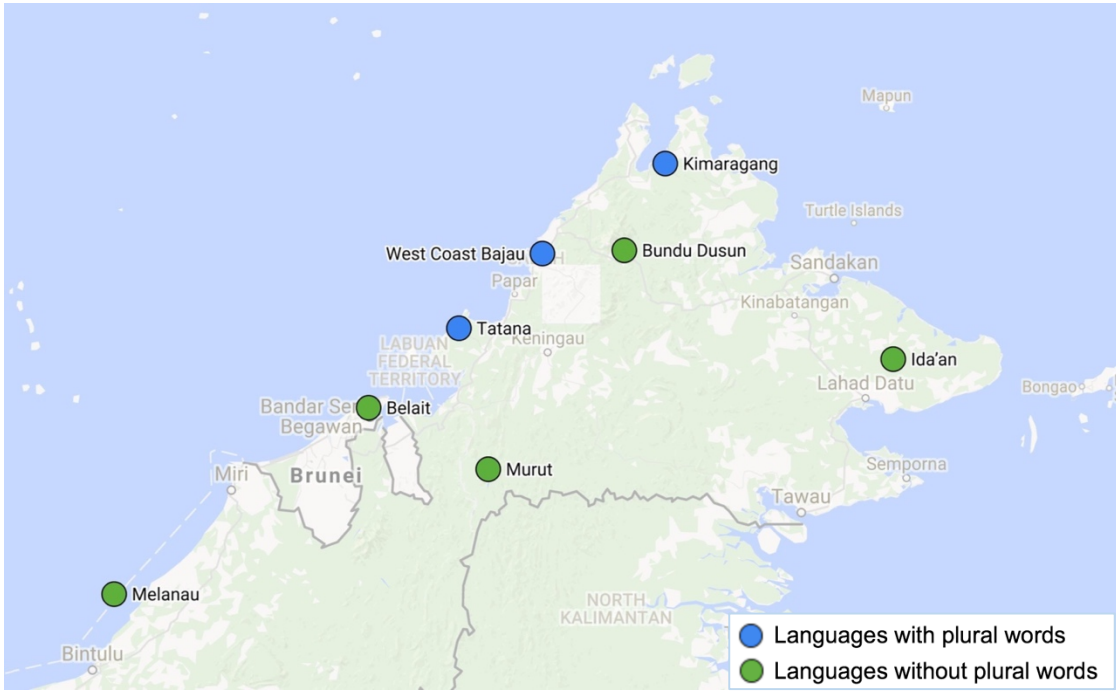
There are some other areas where languages without plural words largely outnumber languages with plural words: Borneo, Sulawesi, east costal line and south-eastern tip of New Guinea. In Borneo, languages with plural words are only found on the northern tip, i.e. Sabah, and altogether missing in the remaining areas (Map 7). In Sulawesi as well, although plural words do have their presence, the majority of languages do not use plural words to denote nominal plurality. Note that while Map 4 above suggests that many languages in east costal line of Papua New Guinea (and the adjacent islands) have plural words, a closer look shows that there are equally many languages that lack plural words (Map 8). In the south-eastern tip of New Guinea, only one out of six languages has plural words (Map 9). Some other languages in Micronesia, represented by Satawalese and Ponapean in my sample, also lack plural words.



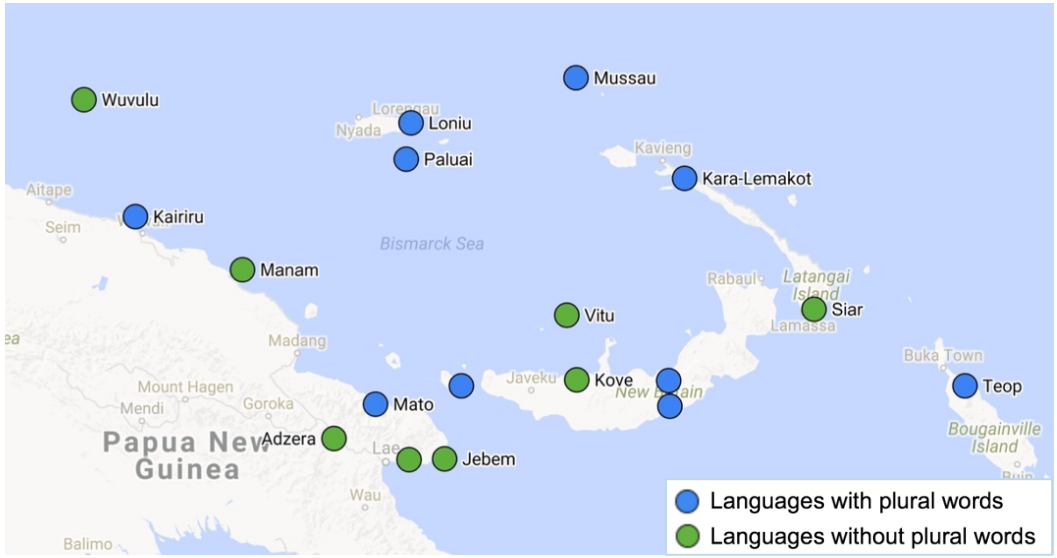
Map 5: Distribution of all sample languages



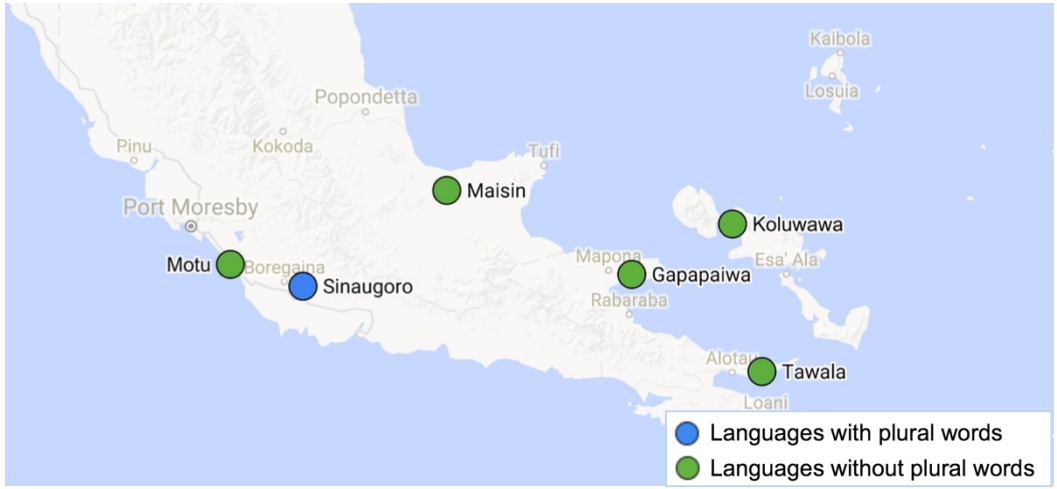
Map 6: Distribution of sample languages in Vanuatu



Map 7: Distribution of sample languages in Northern tip of Borneo (Sabah)



Map 8: Distribution of sample languages in New Guinea and adjacent islands



Map 9: Distribution of sample languages in south-eastern tip of New Guinea

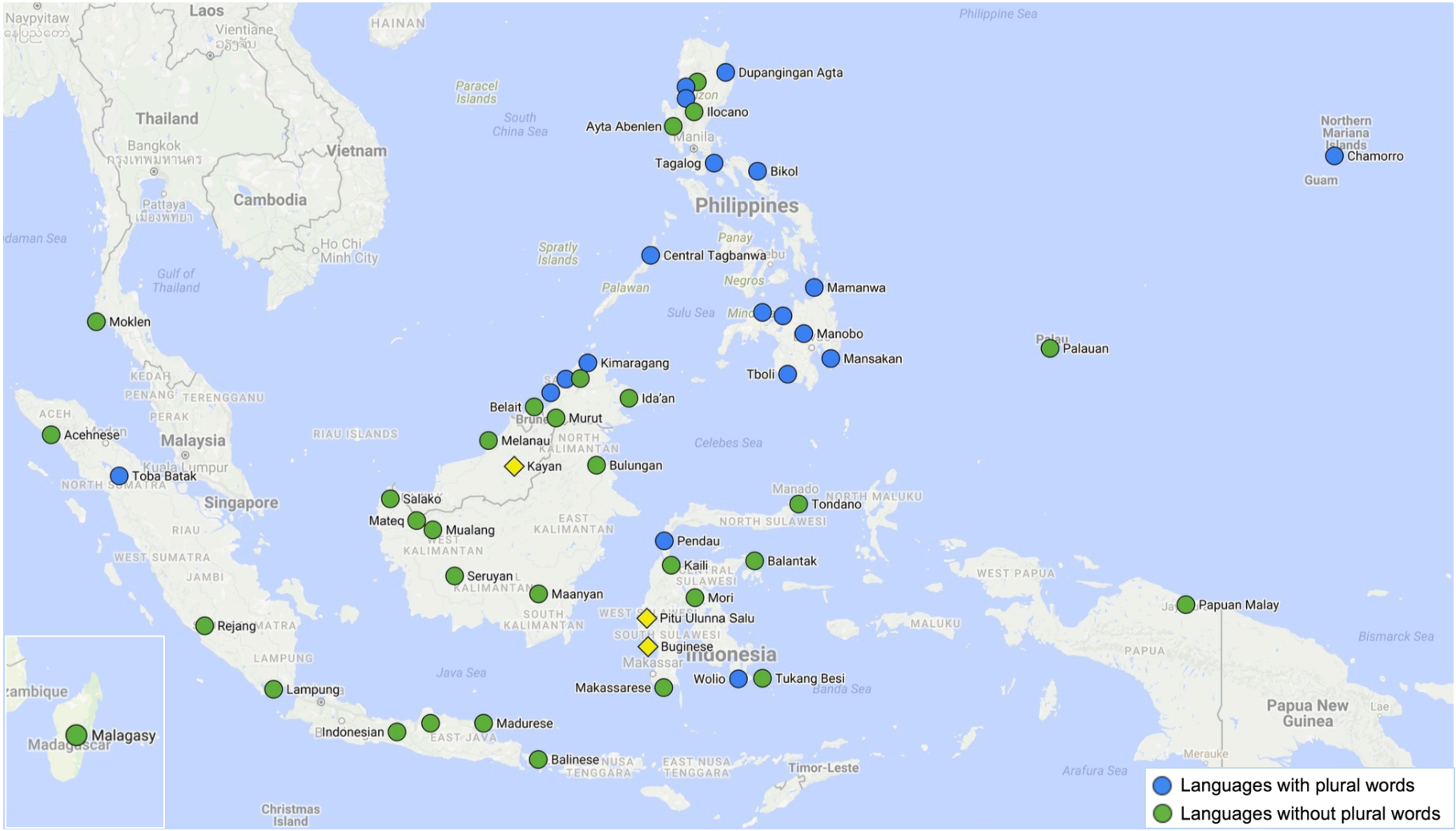
Plural words are strikingly absent in certain areas: the majority of Indonesian islands, including Java, Sumba and Flores; north coastal New Guinea; and New Caledonia. In the whole west Indonesia, Toba Batak on Sumatra is the only language marking nominal plurality with a plural word, while all other sample languages in its adjacent areas lack plural words. All the way to the west, plural words are not found in Malagasy on Madagascar either.

A skewed geographical distribution of plural words in Austronesian languages can now be inferred: plural words are not present in all the regions where Austronesian languages can be found, but only in some of them, especially in the Philippines, Vanuatu, north Borneo and east Indonesia. In some other areas, for instance west Indonesia, almost none the languages in investigation employs plural words.

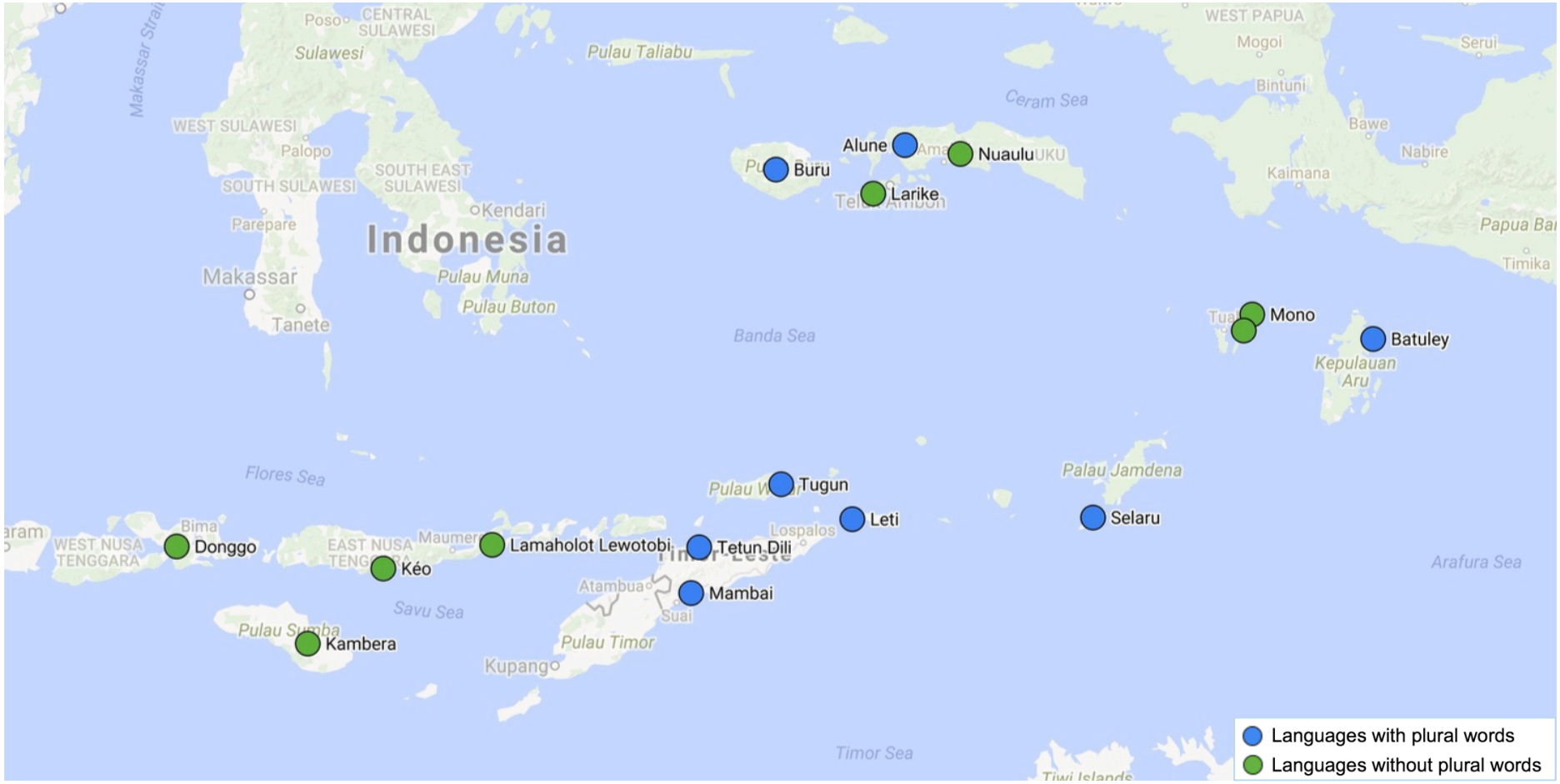
Similarly, for some of the major subgroups in the Austronesian family, languages with plural words also show skewed geographical distribution. The geographical distribution of WMP languages is demonstrated in Map 10. As can be seen, languages in The Philippines, Borneo, and most parts of Indonesia all belong to WMP, but within WMP, most languages with plural words are found in the Philippines, while in west Indonesia almost all languages lack plural words.

A similar uneven distribution of plural words is also seen in CMP languages, as illustrated by Map 11. Within the area where CMP languages are spoken, plural words are only present in the east part of this region. All languages on Timor and surrounding islands, as represented by Tugun, Leti, Tetun Dili and Southern Mambai, have plural words. To the west, all languages (Donggo, Kambera, Kéo and Lamaholot Lewotobi) lack plural words.

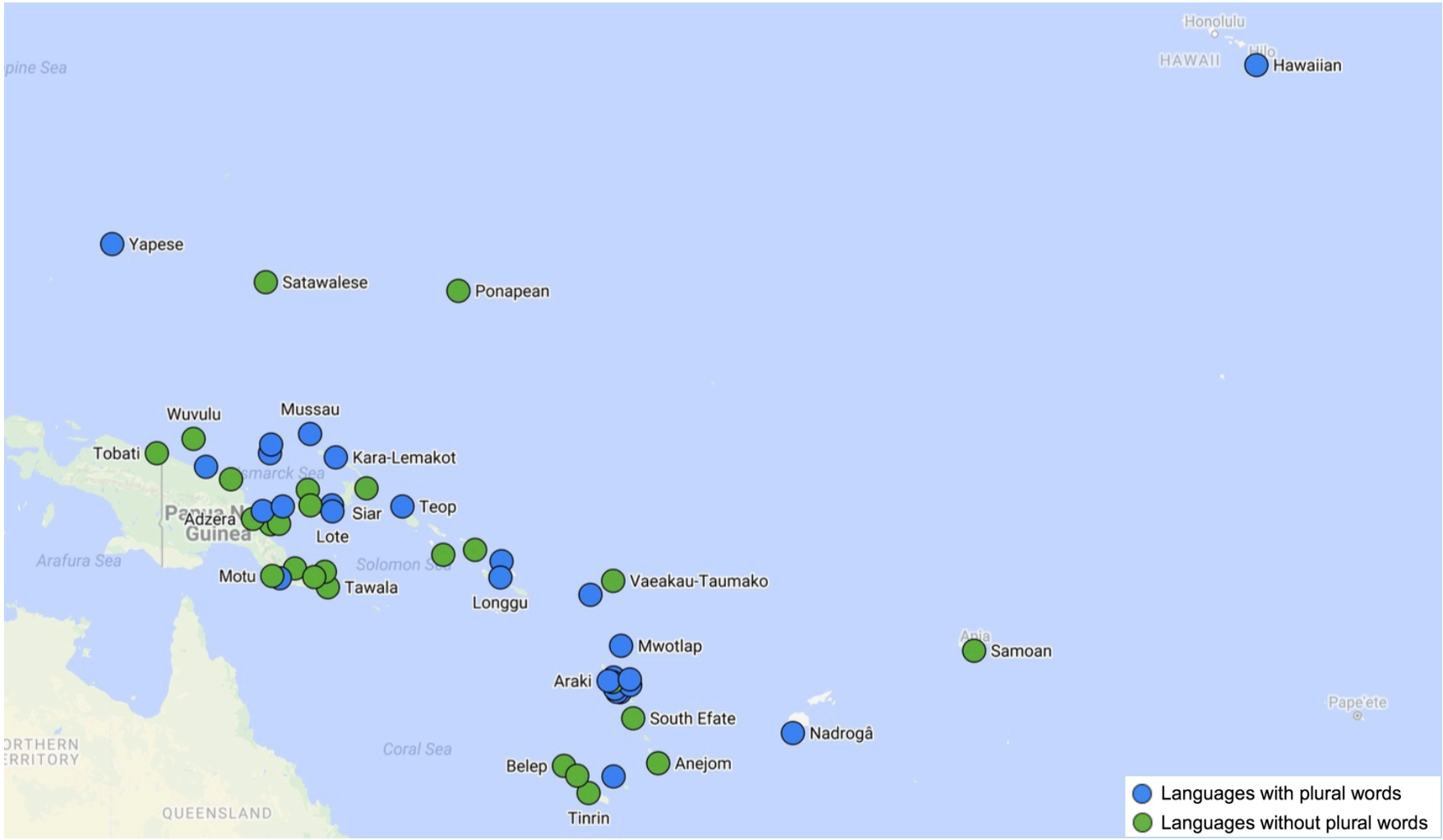
In contrast, Map 12 suggests that there is virtually no areal dichotomy as for Oceanic languages with plural words and those without. There are indeed certain areas where most languages have plural words, e.g. Vanuatu, and some other areas where all sample languages (Cèmuhî, Belep and Tinrin in New Caledonia) do not have plural words. But on the whole, both of languages with and without plural words can be found across different areas.



Map 10: Distribution of WMP sample languages



Map 11: Distribution of CMP sample languages



Map 12: Distribution of Oceanic sample languages

5.3. Interim summary and discussions

Taken together, the findings in this section can be summarised as follows:

- In terms of genealogical distribution, most Austronesian languages with plural words are WMP or Oceanic languages. But as for frequency, plural words are more commonly used in CMP and Oceanic languages, where half of the sample languages employ plural words. There are also some lower-level subgroups presenting a substantial dominance of using plural words to denote nominal plurality, such as Philippine languages and Vanuatu languages.
- As for geographical distribution, plural words are not evenly distributed over all Austronesian languages, but show a skewed distribution. They are mostly found in The Philippines, east Indonesia and Vanuatu, while remarkably absent in west Indonesia, north coastal New Guinea and New Caledonia. A skewed distribution is also found for plural words in some subgroups, i.e. WMP and CMP, but not in Oceanic.

These results can be compared with the observations found in previous studies (discussed in Chapter 2). If we start by comparing the distribution of Austronesian languages with plural words alone (e.g. Map 2 and Map 4), my result is similar to what Dryer has observed: genealogically, most plural words are found in Oceanic languages, followed by WMP, especially in Philippine languages. Geographically, plural words are mostly found in languages in The Philippines and islands of the Pacific Ocean, with few instances in Sabah and east Indonesia. Meanwhile, there are still some differences. For instance, the great majority of Austronesian languages with plural words in Dryer's findings are Oceanic (around 70% in Dryer 2013a). But as can be seen from my results (Chart 1), Oceanic languages with plural words make up less than half of the total number.

As a further step, if we examine the distribution of Austronesian languages with plural words vis-à-vis those without plural words, a number of other differences can be identified. In previous studies, Toba Batak is the only language with a plural word in west Indonesia. However, since the total number of languages selected from west Indonesia is comparably small, a convincing conclusion can hardly be drawn. My results show that even if more sample languages are selected from west Indonesia, Toba Batak is still the only language with a plural word, which suggests that languages with plural words have very little presence in this area. It is also observed in Dryer (2013a) that plural words have particular high presence in Oceanic language (around 80%). In contrast, the results based on my language sample show that half of the Oceanic languages do not have plural words.

These differences do not only result from the fact that my sample languages are

different from Dryer's, but also from the different criteria I use in identifying plural words (see Chapter 3). Some languages are selected in both Dryer (2013a) and my sample, while the value that I gave is different from Dryer's. For instance, for Tinrin, an Oceanic language spoken in New Caledonian, Dryer (2013a) considers it to have a plural word. From the reference given by Dryer (2013a), he takes the plural determiner *mê* in Tinrin as a plural word, as illustrated by the following sample.

(26) Tinrin (New Caledonian, Oceanic)
mê *drae* *rra*
DET.PL thing DIST
'those things' (Osumi 1995: 159)

In the first place, such markers do not fit into my criteria of plural words. *Mê* is not an inherent plural marker, and its plurality is realised in agreement with the nominal element. In addition, there are several prefixes that can be used to mark number in Tinrin: *ke-*, *a-*, *au-*, *truu-* and *mê-*, among which the first three mark singularity, *truu-* mark duality, and *mê-* marks plurality (Osumi 1995: 101).[16] Because of the presence of the plural prefixes *mê-*, I gave the value of *plural prefix* for Tinrin. There is one more reason why considering Tinrin having plural words is imprecise: when the plural determiner is used with a human noun, it is not the sole indicator of nominal plurality, because the noun itself is also marked for nominal plurality. In (27a) and (27b), both nouns, *vîê* 'women' and *voo* 'men', are marked for number by lexical means, thus the plural determiner *mê* does not even fit into Dryer's criteria of plural words.

(27) Tinrin (New Caledonian, Oceanic)
a. *mê* *vîê*
DET.PL women.PL
'the women'
b. *mê* *asirri* *voo*
DET.PL three men.PL
'the three men' (Osumi 1995: 159)

To conclude, this chapter presents the genealogical and geographical distribution of Austronesian languages with plural words. The results partially overlap with what previous studies have presented, but there are also some significant differences which result from the differences in methodology.

[16] They can only be used with human nouns, thus *a-via* 'soldier' > *mê-via* 'soldiers', and *a-droa* 'youth' > *mê-droa* 'youths' (Osumi 1995: 102).

Chapter 6. History of plural words in Austronesian languages

Following the description of how Austronesian languages with plural words are distributed, this chapter deals with the diachronic developments and historical origins of plural words in Austronesian languages.

As discussed in Chapter 2, the reconstruction of a plural word **maŋa* has been proposed in PMP (but not in PAn) and daughter subgroups of PMP, such as PCEMP and POc. At first sight, the results from the previous section confirms the proposed reconstruction of **maŋa*: plural words are indeed not found in Formosan languages, but only in MP languages and its sub-branches, and half of the Oceanic samples have plural words. The distribution pattern also shows that in some subgroups, e.g. Philippine languages, plural words can be found in almost all member languages, which seems to suggest a particularly close historical relatedness. Thus some relevant questions to be asked here are: do plural words in Austronesian languages all reflect the reconstructed form **maŋa*? If not, to what extent do they reflect **maŋa*?

An investigation into the plural words in the 54 languages in my sample offers a negative answer to the first question. Plural words in these 54 Austronesian languages appear in various forms, and their historical origins also vary. Some of them do reflect the reconstructed form, but most of them do not. Third person plural pronoun is a very common source from which plural words have developed, and some other origins can also be identified. In the following discussions, the origins of plural words are grouped into three broad categories: plural words reflecting the reconstructed form **maŋa*, plural words originating from a corresponding third person plural pronoun, and plural words with other miscellaneous origins.

6.1. Plural words reflecting **maŋa*

As Lynch et al (2002: 90–91) note, **maŋa* in POc is taken to be descended from PMP **maŋa*, and this form is widely reflected in Philippine languages, Wolio (Sulawesi), and some Oceanic languages. In my sample languages, plural words reflecting **maŋa* are also found in Philippine languages, Oceanic languages, as well as Wolio. However, overall, they only make up a small portion of all plural words. I will start with a comparison of the plural words that transparently reflect **maŋa* or that have been argued to be descendants of **maŋa*, then attend to other plural words which might also be grouped into this category. Their distribution will then be presented and discussed.

*6.1.1. A comparison of plural words reflecting *maŋa*

Out of 54 languages with plural words, only nine have a relatively clear reflex of **maŋa*, and seven of them are Philippine languages. Outside the Philippines, Wolio, a WMP languages in Sulawesi, also has a plural word descended from **maŋa*; and only one

Oceanic language in my sample, Kara-Lemakot, has a plural word *maana*, which has been taken as a reflex of **maŋa*.

In the Philippine languages, reflexes of **maŋa* are found as Tagalog *mga*, Bikol *mga*, Mansaka *manga*, Mamanwa *manga*, Cebuano *mga*, Central Tagbanwa *manga*, and Western Bukidnon Manobo *menge*. All these languages are close to each other in terms of their geographical distribution. Genealogically, they all belong to the Greater Central Philippine group, and the first five belong to the same lower-level group, Central Philippine. Most of them have a clear reflex of the reconstructed form **maŋa* without many sound changes. *Mga* in Tagalog, Bikol and Cebuano is the conventional orthography for */maŋa/*, thus except for Western Bukidnon Manobo *menge*, in which language the vowel */a/* has changed to */e/*, the plural words in these Philippine languages are identical to PMP **maŋa*.

Apart from the similarities in forms, some common grammatical properties can also be identified in plural words across these languages. The usage of each of these plural words is exemplified by (28).

(28) a. Tagalog (Central Philippine)

Mga	*abogado*	*ang*	***mga***	*lalaki*
PL	lawyer	TOP	PL	man

'The men are lawyers.' (Schachter & Otanes 1972: 111)

b. Bikol (Central Philippine)

laláki 'boy' > ***mga*** *laláki* 'boys'

áyam 'dog' > ***mga*** *áyam* 'dogs' (Mintz 1971: 99)

c. Mansaka (Central Philippine)

yang	***manga***	*baboy*	*na*	*maitum*
TOP	PL	pig	LK	black

'the black pigs' (Svelmoe & Svelmoe 1974: 52)

d. Mamanwa (Central Philippine)

Manga	*lodzoq*	*na*	***manga***	*tao*	*ini.*
PL	bolo.knife	of	PL	person	this

'These are the people's bolo-knives.' (Miller & Miller 1976: 27)

e. Cebuano (Central Philippine)

Mga	*ka-uban*	*nako?,*	***mga***	*lalaki.*
PL	RECP-company	1SG.GEN	PL	male

'My companions, (they're) all male.' (Tanangkingsing 2009: 157)

f. Central Tagbanwa (Palawanic)

layan	***mga***	*punti*
DEM.NOM	PL	banana.plant

'those banana plants' (Scebold 2003: 61)

g. Western Bukidnon Manobo (Manobo)
menge *valey* *ni* *Huwan*
PL house LK Huwan
'Huwan (John)'s houses' (Elkins 1970: 6)

Plural words in these languages are mostly prenominal. It has also been observed that the plural words almost always occur immediately before the noun (Schachter & Otanes 1972: 111; Tanangkingsing 2009: 55; Svelmoe & Svelmoe 1974: 51–52; Scebold 2003: 60), only with few exceptions found in Mansaka and Cebuano, where the plural marker may follow the linker *na* or *nga*, as illustrated in example (29).[17]

(29)a. Mansaka (Central Philippine)
yang *baboy* *na* ***manga*** *maitum*
TOP PIG LK PL black
'the black pigs' (Svelmoe & Svelmoe 1974: 52)

b. Cebuano (Central Philippine)
syempre *kami-nga* ***mga*** *g<in>a<g>may…*
of.course 1PL.NOM-LK PL small<RES><PL>
'Of course, we tiny (employees)…' (Tanangkingsing 2009: 55)

Another similarity in grammatical properties can be found for *mga* in both Tagalog and Cebuano. It these two languages, the plural marker *mga* does not co-occur with numerals in a noun phrase; and when it does, it marks approximation of number. (30) illustrates such an example, but whether this polysemous usage of *mga* is also found in other languages remains unclear.

(30)Cebuano (Central Philippine)
'*unsa=ka* *oras-a* *ni-abot* *sa* *balay'*
what=2SG.NOM hour-DEF AV-arrive LOC house
'***mga*** *seven-thirty'*
APRX seven-thirty
'What time did you get home?'
'Around seven-thirty.' (Tanangkingsing 2009: 56)

Outside the Philippines, a reflex of **maŋa* is found in Wolio. Wolio is a WMP language in South Sulawesi, and it belongs to a putative subgroup Wolio-Wotu. It has a plural

[17] *Nga*, or *na*, or other related forms is a linker for head and attributives, and it is commonly present in Philippine languages.

word *maqa* in Anceaux (1988), and in Anceaux's orthography, *q* represents the voiced velar nasal, thus again the plural word is identical to the reconstructed PMP form **maŋa*.[18] The same word order property similar to reflexes of **maŋa* in the Philippine languages can also be seen in Wolio, as example (31) illustrates that the plural marker *maqa* in Wolio is also placed immediately before the noun. Also, as Anceaux (1988: 36) remarks, *maqa* is not present when plurality is marked by other elements in the noun phrase, such as numerals; but it can be used to mark plurality of the third person pronoun *incia* which does not distinguish number, thus *maqa incia* 'they'.

(31) Wolio (Wolio-Wotu)

a. ***maqa*** *wutitinai*
PL relative
'the relatives'

b. *o* ***maqa*** *mia*
ART PL people
'the people' (Anceaux 1988: 36)

In addition to Philippine languages and Wolio, one Oceanic language in my sample has been argued to have a plural word descended from **maŋa*. Lynch et al (2002: 90–91) list *mana* in Kara, a Tungak-Nalik language, as a reflex of **maŋa*. It is found in my sample that Kara-Lemakot, the Lemakot dialect of Kara, has a plural word *maana*.

Different from other reflexes of **maŋa* discussed above, *maana* in Kara-Lemakot is analysed as a greater plural word, based on the observation that *maana* is only used when the number of referents is greater than just a few (Dryer 2013b: 71), as illustrated by example (32).

(32) Kara-Lemakot (Tungak-Nalik)

...e	***maana***	*rabuna*	*xaves*	*ri*	*sangas*	*usi*	*a*	*salan*	*aave.*
and	GRTR.PL	people	many	3PL	walk	follow	ART	path	that

'... and many people follow that route.' (Dryer 2013b: 71)

In this example, *rabuna* 'people' is not only modified by *maana*, but also by a quantifier *xaves* 'many'; such a co-occurrence of *maana* with *xaves* 'many' or *xapiak* 'all' is not uncommon in Kara-Lemakot.

The analysis of *maana* being a greater plural word is also based on the fact that other number words, such as *mu* 'plural', *ro* 'dual' and *nu* 'paucal', are used in other

[18] Anceaux (1988) is a reprinted version of Anceaux (1952). In *Wolio Dictionary* (Anceaux 1987), the velar nasal sound has been transcribed as *ng*, and an entry for *manga* can be found.

cases in Kara-Lemakot. (33) demonstrates the usage of the plural word *mu*, which is the most common number word in this language.

(33) *A* ***mu*** *fefeng* *si-na* *ri* *fe* *fa-para-ye* *e ...*
ART PL disciple POSS-3SG 3PL go.and CAUS-wake.up-3SG and
'His disciples went and woke him up and ...' (Dryer 2013b: 70)

It should be noted, however, that the greater plural word *maana* can also co-occur with the plural word *mu*, as in example (34). Even more unexpectedly, *maana* sometimes appears with the paucal number word *nu*, as in (35).

(34)... *re* *fe-xuus* *a* *maana* *rabuna* *xaves* *la*
… 3DU INCH-tell ART GRTR.PL people many LOC
maana ***mu*** *bina* *xapiak* *pa-na.*
GRTR.PL PL place all PREP-3SG
'... they told people throughout that entire region about him.'
(Dryer 2013b: 71)

(35) *Na* *Yesus* *e* *mu* *fefeng* *si-na* *ri* *falet* *xe* *la*
ART Jesus and PL disciple POSS-3SG 3PL go to LOC
maana ***nu*** *bina* *faasilak se* *Sesaria* *Filipai.*
GRTR.PL PAUC place near PREP Caesarea Philippi
'Then Jesus and his disciples went to the villages of Caesarea Philippi.'
(Dryer 2013b: 73)

The exact reason for the co-occurrence of two (seemingly contradictory) number words in Kara-Lemakot is not clear; but one possible reason could be that in many cases *maana* corresponds to the meaning of 'all'. *Maana nu bina* in example (35) can thus be interpreted as 'all of the few villages'.[19] It is also not clear whether *maana* can appear with numerals, but no example can be found in Dryer (2013b).

These fundamental differences between *maana* in Kara-Lemakot and its putative cognates in Philippine languages, plus the fact that it is so far the only reflex of **maŋa* that has been attested in an Oceanic language, and the fact that it has cross-linguistically very unmarked and common segments could also warrant the conclusion that this is actually an accidentally similar form and not a cognate of **maŋa*.

[19] It is my interpretation but not Dryer's.

*6.1.2. Other plural words potentially related to *maŋa*

Apart from the above-mentioned reflexes of PMP **maŋa*, there are some other plural words which might also be genetically related to this form.

A potential cognate of *maana* in Kara can be found in Teop, an Oceanic language spoken on the Bougainville island. Teop has a plural word *maa*, as illustrated by example (36).

(36) Teop (Nehan-North Bougainville)

Ahaik,	*a*	***maa***	*ruene*	*toro*	*isuvu*	*ma-ori*	*mohina…*
NEG,	ART PL		water	must	fetch	DIR-3PL	garden

'No, the water must be fetched from the garden....' (Mosel & Thiessen 2005)

Maa in Teop is also prenominal, and it also always precedes the head noun immediately in all the examples containing *maa* provided in Mosel & Thiessen (2005). But it is special in that it always requires an article *a*, which marks word classes. The relationship between *maa* in Teop and other reflexes of **maŋa* cannot be determined at this stage, since the historical phonology of these languages is not well-understood; but considering the radical sound change from **maŋa* to *maa*, the plural word in Teop might have also developed from a completely different source.

It is also noteworthy that some transparent reflexes of **maŋa* do not appear in the shape of a plural word, but a plural affix. A fossilised plural prefix *manga-* is identified in Mori, a Bungku-Tokali language in Central Sulawesi, as in *manga-litau* 'youth' (Esser & Mead 2011: 455). It has also been argued that a plural prefix *manga-* was once productive in Tontemboan, a Minasahan language in North Sulawesi (Adriani & Adriani-Gunning 1908). In discussing the plural prefix *manga-* in Tontemboan, Adriani & Adriani-Gunning (1908) propose that *mang-* is a fusion of *ma-* and *nga-*, and its cognates can be found in Philippine languages as *manga*, and in Tomini languages as *mongo*, and also in Toraja (South Sulawesi), Maori and Samoan.[20]

Two important points in this reference should be highlighted. Firstly, although I am not able to give a thorough analysis of the origin of *manga* here, judging from the existing evidence, the analysis of *manga* as a bimorphemic item is not impossible. The Matigsalug dialect of Manobo has a plural word *me*, which has the same usage of *menge* in Western Bukidnon Manobo. *Me* and *menge* in these two closely-related varieties of Manobo presumably have the same origin, thus *me* could be a part of *menge*. In addition,

[20] The original cite is as follows: 'Met het voorvoegsel ma- wordt nga samengesteld tot manga, dat in het Tt [Tontemboan], evenals in de meeste andere Philippijnsche talen, verder in het Mongondousch en in de Tominische talen (in den vorm mongo), in de Toradja'sche talen en verder in het Maorisch en Samoaansch wordt gebruikt om een meervoud aan te duiden.' (Adriani & Adriani-Gunning 1908: 118)

a prefix *man-* marking plurality can be found in some other languages, for instance Chamorro, as illustrated by example (37). Topping (1973: 234–235) notes that the prefix *man-* is mostly used when the noun functions as a stative predicate, but there are also nouns which take the *man-* prefix when they are not predicates. However, even if *ma-* can be analysed as a single morphemic item, we still need to explain why *nga-*, no matter it being a prefix or infix, is attached to *ma-* to form *manga*.

(37) Chamorro (WMP)

a. ***Man-****estudiante* *siha.*
PL-student 3PL
'They are students.' (Topping 1973: 234)

b. *pale'* 'priest' > ***ma****male'* 'priests'
saina 'parent' > ***ma****ñaina* 'parents'[21] (Topping 1973: 235)

Secondly, Adriani & Adriani-Gunning (1908) mention that a cognate of *manga* can be found in Tomini languages in the form of *mongo*. If true, we would then expect *ongo* in Pendau, a Tomini language, is also possibly related to this form. The usage of the plural word *ongo* in Pendau can be illustrated in example (38). Quick (2007: 190) analyses it as a human group marker, but it does fit into my criteria of plural words.

(38) Pendau (Tomini-Tolitoli, WMP)

unga 'child' > ***ongo*** *unga* 'children'
lei 'young girl' > ***ongo*** *lei* 'young girls' (Quick 2007: 190)

This link of *manga* ~ *mongo* ~ *ongo* opens much wider possibilities of other plural markers being genetically related as well. Some forms which are clearly related to *ongo* can be found in some north Bornean languages in my sample: a plural word *tongo* in Kimaragang, and *banga(n)/bengen/bongon* in West Coast Bajau, and a plural infix *-ongo-* in Bundu Dusun. Furthermore, a plural proclitic *ngo-* in Tatana' and a plural prefix *ŋaN-* in Murut might also have the same origin. Note that all these five languages are spoken in Sabah, and the first four belong to the same subgroup of Southwest Sabahan in WMP. (39) illustrates the usage of plural markers in each of these languages.

(39) a. Kimaragang (Southwest Sabahan)

It	***tongo***	*torigi*	*dirih*	*nga*	*aso*	*noh.*
NOM	PL	house.post	ANAPH	but	NEG.exist	already

'Even the house posts were gone (burned up).' (Kroeger 2005: 410)

[21] Considering the morphophonological alternation of *man-*, it should be best analysed as *maN-*.

b. West Coast Bajau (Sama-Bajau)
*Abis mangan **bongon** pakir pan be-pule' no.*
finish AV.eat PL religious.man TOP DISTR-go.home FOC
'After eating, the men with religious knowledge went home.'
(Miller 2007: 113)

c. Bundu Dusun (Southwest Sabahan)
tulun 'person' > *t<**ongo**>ulun* > *tongoulun* 'a group of people'
tanak 'child' > *t<**ongo**>anak* > *tanganak* 'children'
(Price 2007: 34)

d. Tatana' (Southwest Sabahan)
*Aro duo **ngo-**bua baloi*
there.is two PL-CLF house
'There are two houses.'
(Chan & Pekkanen 1989: 12, cited from Dillon (1994: 25))

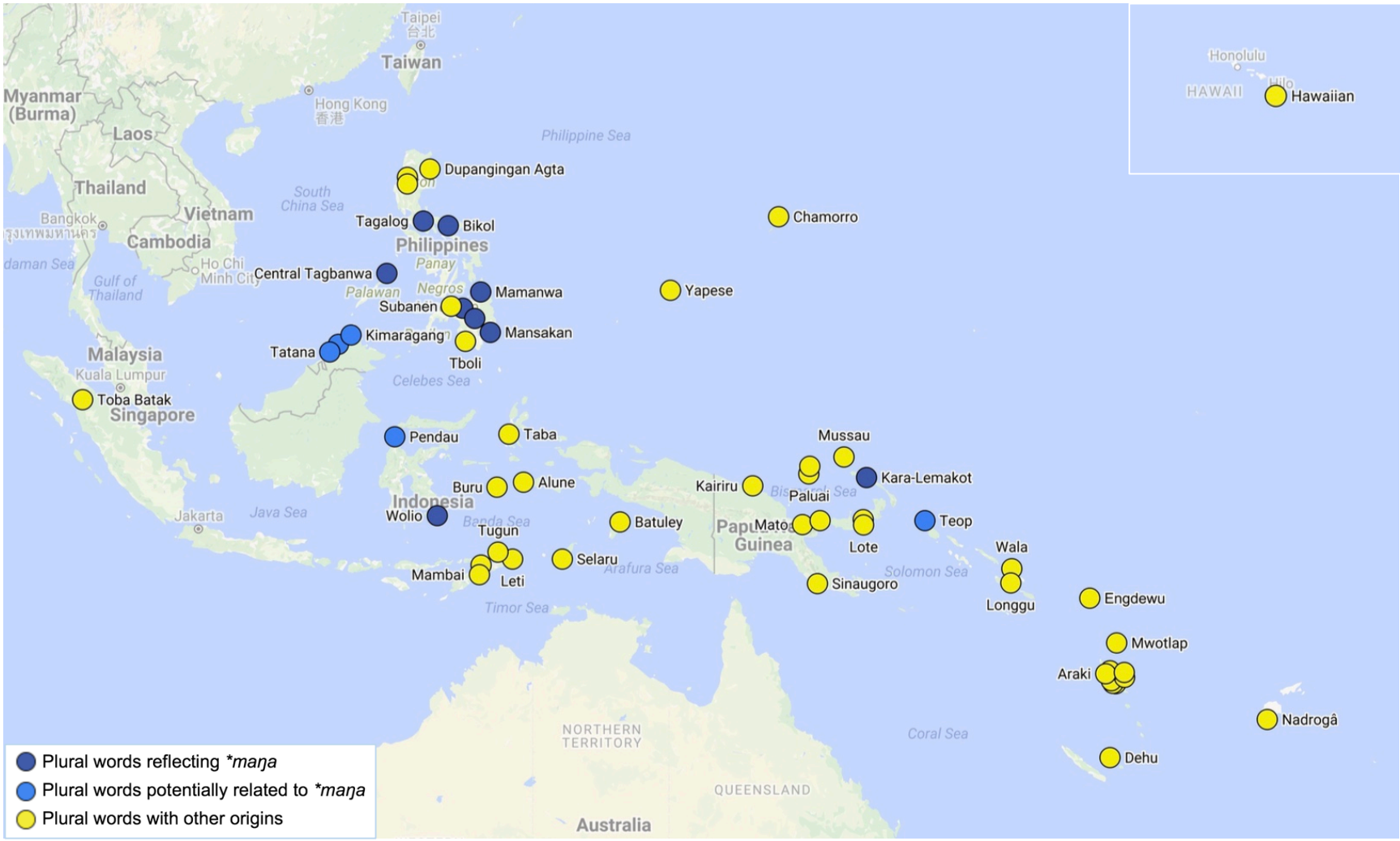
e. Murut (Southwest Sabahan)
taun 'year' > ***ŋa**taun* 'years'
bulan 'month' > ***ŋam**bulan* 'months'
baloy 'house' > ***ŋam**baloy* 'households' (Prentice 1971: 175)

Given the fact that languages in Sabah (as well as in Sulawesi) are indeed typologically very similar and historically related to Philippine languages, such a connection of plural markers among Philippine-Sabah-Sulawesi languages is not surprising. And if such correlation were true, it in turn backs up the proposal that *manga* might be a fusion of *ma-* and *nga-*, since it is the *nga-* part that is shared by all these languages.

The hypothesis of *manga* being fused from *ma-* and *nga-*, and the correlation among *manga ~ (mongo) ~ ongo ~ tongo ~ bongon ~ ngo-* is a hypothesis at this stage. But the foregoing discussions also suggest that some other forms might also be related to **manga*. Further studies may reveal more of the history of *manga* (and its related forms) and test if **manga* is indeed reconstructable as a morphologically simple lexeme.

*6.1.3. Distribution pattern of plural words reflecting *maŋa*

The previous sections show that only nine present-day Austronesian languages have plural words reflecting **maŋa*. Even if we include the possible items discussed in 6.1.2, the total number of plural words reflecting PMP **maŋa* is still relatively small. The distribution of these plural words reflecting **maŋa* can be illustrated in Map 13. In this map, dark blue dots represent languages with a plural word reflecting **maŋa*, light blue dots represent the possibly-related items, and yellow dots represent languages with a plural word from a different origin.



Map 13: Distribution of plural words reflecting **maŋa*

As can be seen, most of the plural words reflecting **maŋa* are found in Philippine languages, and all these Philippine languages are genealogically closely-related and geographically adjacent. However, not all Philippine languages have a plural word that is descended from PMP **maŋa*.[22] Kankanaey and Ibaloy, for example, have developed their plural words from third person plural pronouns, and this development will be further discussed in the next section. Outside the Philippines, Wolio is the only WMP language having a reflex of **maŋa*. In CMP and SHWNG, plural words reflecting **maŋa* are altogether missing; in Oceanic languages, only one language, Kara-Lemakot, has the reflex of **maŋa*.

Lynch et al (2002: 90–91) list the following Oceanic languages as having plural words as decedents of **maŋa*, Tigak *manama*, Kara *mana*, Tolai *umana*, Halia *maman* and Nguna *maaŋa*. This quote seems to suggest that Oceanic languages as a whole widely reflect the reconstructed form **maŋa*, but an inspection of the genealogical affiliation and geographical distribution of these five languages suggests that they have close genealogical relationships, and they are spoken in geographically adjacent areas. All five languages belong to Western Oceanic linkage, and New Ireland-Northwest Solomonic linkage (NINSL) on a lower level. Therefore, no instance of reflexes of **maŋa* has actually been proposed outside languages in Western Oceanic linkage, and in my sample they are not found in any other subgroups of Oceanic languages.[23]

In addition, even within Western Oceanic linkage, it appears that plural words reflecting **maŋa* are not frequently attested either. Among the five languages listed in Lynch et al (2002: 90–91), only Kara, as presented by the variety of Kara-Lemakot, is included in my sample. The exclusion of other languages is not intentional. According to *Glottolog* (Hammarström et al. 2016), the total number of languages spoken in NINSL is 50. Therefore, five languages in NINSL are selected in my sample with considerations of their lower-level genetic classification: Kara-Lemakot, Siar, Ughele, Kokota and Teop. In these five sample languages, only Kara-Lemakot has a reflex of **maŋa* (which might also be an accidentally similar form), while further study might also identify a cognate in Teop; but the other three languages, Siar, Ughele and Kokota, simply do not employ a plural word.[24]

In such a case, the synchronic reflexes of PMP **maŋa* are only widely attested in the Philippines (and possibly in Sabah), a distribution that casts serious doubts on its status as reconstructed form for the entire MP family. In other words, at this stage, even

[22] And as discussed in Chapter 5, not all Philippine languages have a plural word.

[23] It does not mean that they are absolutely absent, but still, it implies that even if reflexes of **maŋa* can be found in other subgroups, they are quite rare.

[24] Kokota is also examined in Dryer (2013a), who gave it the value of *plural words*. The plural words identified by Dryer are in fact a plural article and a plural demonstrative, therefore they are excluded in my analysis.

if the reconstruction of a PMP **maŋa* were true, there is very little supporting evidence. If **maŋa* is indeed reconstructable, some other issues need to be explained in further studies. For example, in Philippine languages, I have shown that some languages do not have a plural word, and some others have a plural word that is not related at all to the reconstructed from. It would then be necessary to explain how some daughter languages lost the plural words in their histories, and developed a new plural word from another origin.

6.2. Plural words originating from third person plural pronouns

Other than having a plural word relating to **maŋa*, a great number of Austronesian languages have a plural word developed from a third person plural pronoun. This section will be organised as the previous one, starting with a comparison of plural words originating from third person plural pronouns, followed by a discussion of other potentially-related forms and a presentation of the distribution pattern.

6.2.1. A comparison of plural words originating from third person plural pronouns

Example (40) below illustrates a correspondence of the plural word and the third person plural pronoun in Tugun, a CMP language spoken on the Wetar island in east Indonesia. In (40a), *hira* functions as a post-nominal plural word which modifies the head noun *manu* 'bird', while in (40b), *hira* functions as a third person plural pronoun, being the subject the sentence.

(40) Tugun (Timoric)

a. *Manu* ***hira*** *r-ihi-fere* *pari* *futun* *faitu.*
bird PL 3PL-RECP-divide become group seven
'The birds divided into seven groups.' (Hinton 1991: 50)

b. ***Hira*** *r-a* *e,* *am toa* *naha.*
3PL 3PL-eat COMPL 1PL NEG yet
'They have eaten, but we haven't yet.' [25] (Hinton 1991: 93)

Twenty languages in my sample shows such a correspondence between the plural word and a certain type of third person plural pronoun. One might wonder how I could claim that in these cases the plural word originated from the third person plural pronoun, since it can also be argued that the similarities in forms are due to mere chance. I do not deny this possibility in certain cases, and it is true that we do not have much evidence to make a strong assertion. But as a whole, plural words in many languages are identical

[25] Note that in (38b) the subject is also marked on the verb by a third person plural suffix *r-*. *Hira* as a free pronoun is actually seldom used in Tugun, here it creates a contrast between *hira* 'they' and *ami* 'we'.

to the third person plural pronoun, and they cannot be all due to chances.

There are another two reasons to confer this development. On one hand, the development from a third person plural pronoun to a plural marker has been suggested as a common grammaticalisation process cross-linguistically. Heine & Kuteva (2002: 237–238) have given examples from a number of languages across the world and they see such a grammaticalisation as a classical instance of desemanticisation, through which process the main semantic content of the plural pronoun is bleached out. On the other hand, it also is explicitly stated in many grammars that the third person plural pronoun can be used as a plural marker.

While a similar development is shared by many Austronesian languages, the exact developmental paths encode many variations. Synchronically, it can be observed that these plural words and their corresponding third person plural pronouns vary in forms, and the grammatical function served by the third person plural pronoun is not the same. The relative word order of the plural word and the head noun is also different: some plural words are prenominal while other are post-nominal. A comparison of these plural words is summarised in Table 8 on the next page, with the consideration of the basic word order of these languages as well.

Table 8: Correspondences of plural words with third person plural pronouns

Language name	*Plural word*	*Corresponding third person plural pronouns*	*Grammatical role or functions of the third person plural pronouns*	*Order of the plural word and head noun*	*Basic word order*
Kankanaey	*da*	*=da*	Subject	Prenominal	VSO
Ibaloy	*ira /ʔida/*	*ira /ʔida/*	Nominative	Post-nominal	VSO
Dupangingan Agta	*hidi*	*hidi*	Nominative, topic	Post-nominal	VSO
Chamorro	*siha*	*siha*	Subject, object	Post-nominal	VSO
Buru	*=ro*	*=ro*	Object (undergoer)	Post-nominal	SVO
Leti	*=ra*	*=ra*	*Not clear*	Post-nominal	SVO
Tetun Dili	*sira*	*sira*	Subject, object	Post-nominal	SVO
Tugun	*hira*	*hira*	Subject, object	Post-nominal	SVO
Taba	*=si*	*si*	Subject, object	Post-nominal	SVO
Paluai	*ip*	*ip*	*Not clear*	Prenominal	SVO
		ip=	Subject		
		=ip	Object		
Loniu	*sɛh*	*sɛh*	Subject, object	Prenominal	SVO
Nadrogā	*kura*	*kura*	Topic	Prenominal	SVO
Neve‘ei	*ar*	*ar*	Subject	Post-nominal	SVO
Tape	*er*	*er*	Subject	Post-nominal	SVO
		=ër	Object		
Abma	*nii*	*nii*	Topic, object	Post-nominal	SVO

Nakanai	*egite*	*egite*	Subject	prenominal	SVO
Mato	*=di*	*-di*	Object	Post-nominal	SVO
		di-	Subject		
		ding	Subject, object		
		-ding	Possessive		
Mangap-Mbula	*zin*	*zin*	Topic, nominative, accusative	Prenominal	SVO
Kairiru	*rri*	*rii*	Subject, object	Prenominal	SOV
		=rri	Object, possessive		
Sinaugoro	*-ri*	*-ri*	Subject, object, possessive	Post nominal	SOV

From this table we can see that the forms of the plural words and their corresponding third person plural pronouns exhibit many variations across these languages. Some cognates are identifiable, for instance, Kankanaey *da*, Ibaloy *ira*, Dupangingan Agta *hidi*, Chamorro *siha* all descend from PMP **ida* or **si ida* 'third person plural pronoun', where **si* is a nominative case marker. *Sira* in Tetun Dili and *hira* in Tugun, as well as the clitic forms *=ro* in Buru, *=ra* in Leti also presumably have the same origin. On the other hand, some other plural words and pronouns have distinctive forms, such as *ip* in Paluai, *sɛh* in Loniu and *kura* in Nadrogā. The origins of these pronouns are not clear,[26] but still, a striking correspondence between the plural word and a certain type of third person plural pronoun can be found in all these languages.

The grammatical functions of the pronouns also vary considerably. In the Tugun example above, the plural word *hira* corresponds to a free pronoun, which can function as either a subject or an object in Tugun. But some other languages make a distinction between the subject pronoun and the object pronoun, and the plural word is only comparable to one of them. In Dupangingan Agta, for example, the plural word *hidi* as in (41a) corresponds to the nominative third person plural pronoun in (41b), but not the genitive pronoun *=di* in (39c).

(41) Dupangingan Agta (North Luzon)

a. *<in><um>angay golang* ***hidi*** *ha Tuguegarao*
<COMPL><AV>go orphan PL OBL Tuguegarao
'Orphans went to Tuguegarao.' (Robinson 2011: 69)

b. *mag-pokpok* ***hidi***
AV-wash.clothes 3PL.NOM
'They are washing clothes.' (Robinson 2011: 83)

c. *i-labbang****=di*** *ni kakay*
TV-bury=3PL.GEN PERS grandfather
'They will bury grandfather.'[27] (Robinson 2011: 86)

It should be noted, again, that such correlations are only made based on the forms of the plural word and the third person plural pronoun, given that the exact diachronic development is under-researched. Thus in some other languages, a plural word has

[26] Hamel (1994: 52) suggest that plural pronoun in Loniu, *tanah* 1INC, *uwɛh* 1EXC, *hah* 2PL, *sɛh* 3PL appear to contain the root for 'four' and might have derived from a quadruple pronoun forms.

[27] Although from the English translation it seems that the third person pronoun is still in the nominative case, but in Dupangingan Agta (and many other Philippine languages), the case of the third person pronoun depends on the grammatical voice marker used on the verb. Here the sentence is in theme voice, so literally it means 'The grandfather will be buried by them.'

several corresponding forms of third person pronouns, and it cannot be determined at this stage which pronoun eventually developed into a plural word. In Paluai, for example, the form of the plural word is identical to the free pronoun *ip*, but at the same time the subject bound form *ip*=, and the object bound form =*ip*, might also be the source of the plural word. In the following example, (42a) demonstrates the usage of *ip* as a plural marker, (42b) and (42c) exemplify the usage of the subject clitic *ip*= and object clitic =*ip* respectively.

(42) Paluai (Admiralty Islands)

a. ***Ip*** *pein* *pari* *Nauna*
PL woman belonging.to Nauna
'(The) women from Nauna.' (Schokkin 2014:199)

b. *ipno* *rok* *wot* *onga* *ippe* *yong* *tuktuk tou*
ip=no rot wot onga **ip**=pe yong tuktuk.tou
3PL=IPFV sit go.level and.so 3PL=PFV hear drum.beat
'They were sitting around and they heard the drumbeat for calling people.'
(Schokkin 2014: 216)

c. *Ingan* *mat* *nêmip*
yi=ngan mat nêm=**ip**
3SG=eat die be.finished=3PL
'He killed and ate them all.' (Schokkin 2014: 211)

A possessive pronoun can also be the origin of a plural word, and Sinaugoro is such an example.

(43) Sinaugoro (Peripheral Papuan Tip)

a. *Numa* *vovoka* *namo**-ri*** *ḡe* *raga-ri-ni*
house many good-PL 3PL.REAL.IMM build-3PL-IPFV
'They are building many nice houses.' (Tauberschmidt 1999: 63)

b. *Sina**-ri*** *na* *e* *nari-ri-to*
mother-3PL.POSS ERG 3SG.REAL.IMM care.for-3PL-PERF
mo *ḡe* *barego-to…*
till 3PL.REAL.IMM big-PFV
'Their mother brought them up…' (Tauberschmidt 1999: 102)

Another special kind of pronoun from which the plural word can develop is called *independent pronoun*. Some languages have a special pronoun that can be differentiated from either subject pronoun or object pronoun. Thus *independent* here does not mean a free form (in contrast to a bound form), but refers to a particular kind of pronoun that

has special functions, usually topicalisation.[28] Nadrogā is a language that makes such a distinction. In Nadrogā, the independent third person plural pronoun *kura* functions as the nucleus of an NP (corresponding to the function of *topic* in Table 8); but a distinctive subject third person pronoun *ara* occurs in a verbal phrases, and the object pronoun is in the form of a suffix *-ra*. (Geraghty 2002: 836). It is the independent pronoun *kura* that can be used as a plural word; example (44) illustrates such a distinction.

(44) Nadrogā (West Fijian)

a. *o* ***kura*** *coko* *na* *lewa*
ART.PERS PL all ART woman
'all the women' (Geraghty 2002: 838)

b. ***Ara*** *nō* *koto* ***ara*** *vī-vohaki.*
3PL stay CONT 3PL RECP-discuss
'They stayed and they discussed.' (Geraghty 2002: 846)

Ibaloy is another language that has such a distinctive independent plural pronoun. Interestingly, unlike Nadrogā, the plural word *ira* in Ibaloy does not originate from the independent pronoun, but the nominative pronoun. Example (45a) illustrates the usage of plural word *ira* in Ibaloy, and a corresponding nominative plural pronoun can be found in (45b).[29] In contrast, in (45c), the independent third person plural pronoun *si'kato* is used for topicalisation.

(45) Ibaloy (North Luzon)

a. *inkowan* *nonta* *Maodi* *soni* *too* ***ira***
ʔin-kowan nonta maʔodi so=ni toʔo ʔida
THMV.PFV-say GEN.RECP Maodi OBL=GEN person PL
'Maodi said to the people.' (Ruffolo 2004: 342)

b. *Mimotok* ***ira*** *nonta* *abril.*
<im>motok ʔida nonta abril
<ACTV.PFV>arrive 3PL when.PAST April
'They arrived last April.' (Ruffolo 2004: 129)

[28] Dupaningan Agta also has a set of such independent pronouns, which are called 'long-form nominatives' (Robinson 2011: 83–84). But in Dupaningan Agta, the third person plural independent pronoun has the same form as the nominative pronoun. In some other languages, it is usually either the free subject pronoun or the object pronoun that is used for topicalisation.

[29] Genitive pronoun in Ibaloy is *=cha* or *=ra*, both pronounced as */da/*.

c. *tep* *ebatekan* ***si'kato***
tap ʔa-batak-an siʔgato
because PoTLOCV.PFV-tattoo-LOCV 3PL.IND
'because he was tattooed' (Ruffolo 2004: 174)[30]

Apart from the variations in the forms of the plural words, Table 8 also shows that the relative position of plural words and head nouns differs in these languages.

In earlier work of linguistic typology, it has been suggested that nominal modifiers are dominantly post-nominal in VO languages, and pre-nominal in OV languages (Greenberg 1963: 88–90; Lehmann 1973: 48). This statistical universal has been rejected in Dryer (1986; 1988), who argues that there is no evidence of any correlation between the verb-object order and modifier-noun order. As for the relationship between verb-object and plural word-noun order, Table 8 also shows that there is no transparent correlation. Even though VO is the dominant word order pattern in sample languages above, we can find many prenominal plural words in these VO languages as well. Also in the two OV languages, Kairiru and Sinaugoro, both prenominal and post-nominal plural words can be found.

I argue that the word-order pattern of these plural words corresponding to third person pronouns has to be taken as a result of diachronic developments. As mentioned above, while claiming that plural words in these languages have developed from a certain type of third person plural pronoun, little is known about from which pronoun the plural word has developed, and how this grammaticalisation has taken place. A plural word developed from a subject pronoun might have the same form as a plural word developed from an object pronoun, but their relative positions as regards to the head nouns might differ, depending on the basic word order of the language in investigation. We might expect that a plural word developed from a subject third person pronoun in a SVO language is likely to be prenominal, as in the Nakanai example below.

(46) Nakanai (New Ireland-Northwest Solominic linkage)
a. ***Egite*** *tovo* *lalai* *la* *viliti.*
3PL try ADV NM fishing
'They tried fishing' (Johnston 1980: 134)
b. ***egite*** *la* *malu*
PL NM bird
'the birds' (Johnston 1980: 175)

[30] A more literal translation would be 'it was he who got tattooed.'

The basic SVO order is illustrated in (46a), and (46b) shows that the plural word, *egite*, corresponds to the subject pronoun.[31] *Egite* in (46b) is analysed as a plural word, which marks the plurality of the nominal element *malu* 'bird'; but at the same time, it is also possible to take (46b) as a noun-noun juxtaposition. Nakanai generally lacks a copula, and in (46b), the plural word precedes the nominal marker, thus it can also be interpreted as 'they are birds'. This reinterpretation suggests a possible diachronic development: *egite* once only functioned as a subject pronoun, and in noun-noun juxtaposition structures it is gradually reanalysed as a plural marker. For this particular grammaticalisation, i.e. from a subject third person plural pronoun in a SVO language to a plural marker, it is then not surprising to find that the plural marker is prenominal. But again, other languages do not always share the same grammaticalisation path, thus the result of the word-order pattern can be very different.

All being said, some areal features can still be identified for languages in Table 8. All languages in east Indonesia (Buru, Leti, Tetun Dili and Tugun) employ a post-nominal plural word with a basic SVO word order, and the forms of the plural words are also very similar. In Tape and Neve'ei, two closely-related SVO languages in Vanuatu, the post-nominal plural word *ar* and *er* also show many similarities in forms and grammatical properties.

6.2.2. Other plural words potentially related to third person plural pronouns

The foregoing section discusses the plural words which synchronically have an identical form to a third person plural pronoun. However, if a plural word originated from a third person plural pronoun in a relatively early stage, say, in an ancestor language, then synchronically the plural word and the current third person plural pronoun in a daughter language might not always correspond in forms. Some languages in my samples do reflect such a situation.

Southern Mambai, a language spoken on the Timor island in east Indonesia, has a post-nominal plural word *sêr*, as in *mu* 'banana' > *mu sêr* 'bananas'. On the surface, this plural word has no connection to the third person plural pronoun *rom*, but a comparison with other genealogically-related and geographically-adjacent languages suggests a historical relatedness. Plural words in languages in Timor and their corresponding third person plural pronouns are summarised in Table 9.

[31] *Egite* is analysed as *e-gite*, where *-gite* is the pronoun root, *e-* being a personal number marker. When third person pronoun functions as an object, it is in the root form and attaches to the verb as a suffix *-gite*.

Table 9: Plural words and third person plural pronouns in languages in Timor

Language name	*Plural word*	*Third person plural pronoun*
Leti	*=ra*	*=ra*
Tetun Dili	*sira*	*sira*
Tugun	*hira*	*hira*
Southern Mambai	*sêr*	*rom*

The historical relatedness of *sira* with *hira* and the clitic form *=ra*, which involves the sound change from */s/* to */h/*, is transparent. If these words are indeed historically related, rather than purely synchronically similar, we could reconstruct a plural word **sira* based on these sound changes. But is *sêr* also a cognate? Two pieces of evidence suggests that it is indeed very possible. In discussing the plural word in Southern Mambai, Hull (2003: 11) considers *sêr* to be comparable to *sira* in Tetum (also known as Tetun),[32] and mentions another plural word *seir* in the Daissua-Betano sub-dialect of Tetum. The relatedness between *seir* and **sira* can be justified by the process of metathesis, and it has been found that metathesis is not uncommon in languages in Timor (Steinhauer 2008; Edwards 2016). A correlation thus can be made: *sêr ~ seir ~ (sire) ~ sira*.

The conclusion of this reasoning is that the ancestor of the current Timor languages had already developed a plural word from the third person plural pronoun **sira*, and this plural word was passed onto the daughter languages. In Southern Mambai, the plural word underwent regular sound changes and metathesis, thus having the current form *sêr*. In the same historical period, its original third person plural pronoun was lost and a new third person plural pronoun *rom* has developed, resulting in the fact that even though the plural word is related to third person plural pronoun, it is not reflected in the forms of these two elements at the current stage.

In addition to Southern Mambai, the plural word *ira* in Raga, and *iɣe* in Mwotlap might also be historically related to a third person plural pronoun. Example (47) and (48) illustrate the usage of the plural words and the free third person plural pronoun in Raga and Mwotlap.

(47) Raga (North Vanuatu)

a. ***Ira*** *vavine* *ra-m* *bavatu.*
PL woman 3PL.CONT weave
'The women are weaving.' (Vari-Bogiri 2011: 82)

[32] The plural word *sira* in Tetum can be found in Hull & Eccles (2001: 14).

b. ***Kera*** *ira nitu-na*.
3PL PL child-3SG.POSS
'They are his/her children.' (Vari-Bogiri 2011: 177)

(48) Mwotlap (North Vanuatu)
a. ***iɣe*** *lok^{w}oven*
PL woman
'women' (Crowley 2002: 592)
b. ***ikey*** *mɛ-mtiy-to*
3PL PST-sleep-PST
'they slept' (Crowley 2002: 595)

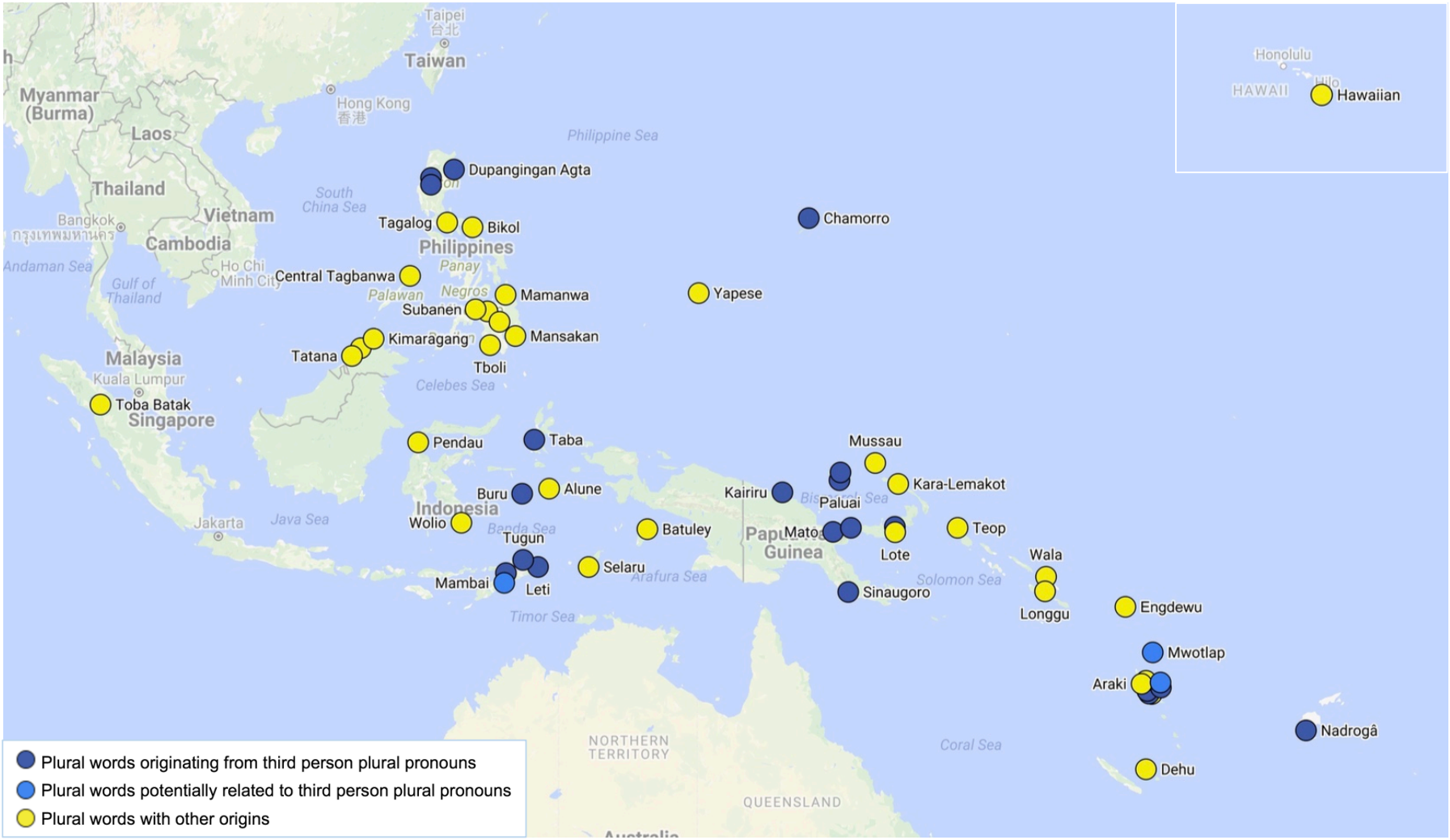
At first sight, *ira* and *kera*, *iɣe* and *ikey* do not indicate much correlation. However, all these forms are related to the third person plural (non-singular) pronoun reconstructed in POc, **[k]ira* (Lynch et al 2002: 67). Thus it is also very likely that the plural words *ira* and *iɣe* have developed from a third person plural pronoun at an earlier stage, after which the grammaticalised plural word underwent different developing paths from the plural pronouns.

6.2.3. Distribution pattern of plural words originating from third person plural pronouns

The distribution pattern of the plural words discussed above can be depicted in Map 14. Dark blue dots represent languages with a plural word corresponding to a third person plural pronoun, light blue dots represent languages that are potentially related, and yellow dots represent languages with plural words of other origins.

Overall, the development from a third person plural pronoun to a plural word can be found in many different areas: The Philippines, east Indonesia and various spots on the Pacific Ocean (particularly Vanuatu) and east Papua New Guinea. In certain areas, plural words in all languages (potentially) correspond to a third person plural pronoun: north Luzon, Timor island, and east coastal line of Papua New Guinea.

Considering the fact that forms and grammatical properties of these plural words also differ for languages in different areas, we might not expect that all of them have a common ancestor, but that such a development is parallel in many places. However, as discussed above, plural words in languages on the Timor island might have developed at an earlier stage, and it is very likely that the current plural words share a common ancestor. Such a scenario is also possible for the plural words in Vanuatu languages, such as Raga and Mwotlap.



Map 14: Distribution of plural words corresponding to third person plural pronouns

6.3. Plural words with other origins

The preceding two sections discussed two major historical sources of plural words in my sample languages. The diachronic developments of some other plural words remain unclear, and it is also possible that more plural words can be related to the reconstructed form **maŋa* or a third person plural pronoun. I am not able to determine each of these origins and provide a thorough explanation, but some distinctive plural words will be presented in this section, and I will discuss their origins when possible.

Toba Batak is the only language with a plural word in west Indonesia. Plural word in Toba Batak is *akka*, as in *jabu* 'house' > *akka jabu* 'houses' (Nababan 1981: 87). As far as its form is concerned, this plural word does not show any affiliation to plural words in other languages. Tuuk (1971: 116) proposes that this plural word was originally used as a kinship term referring to same-sex elder siblings, and a potential cognate *angka* is still used in this way in the Mandailing dialect of Batak. As Tuuk (1971: 116) suggests, since it is obligatory to use the term to show respect, it gradually lost its lexical meaning and became a plural marker.[33]

Another distinctive plural word, *etoa*, with a variant *atoa*, can be found in the Oceanic language Mussau. Example (49) illustrates its usage, and as can be seen, the plural word can co-occur with a quantifier. As for its history, the plural marker *etoa/atoa* seems to be related to a quantifier *katoa* 'some, a few'.

(49) Mussau (St. Matthias, Oceanic)

a. *nau* ***etoa***
day PL
'days'

b. *nau* *ekapa* ***atoa***
day all PL
'every day' (Brownie & Brownie 2007: 64)

Lote, another language in New Ireland-Northwest Solomonic linkage (closely related to Nakanai and Kara-Lemakot discussed above), has a post-nominal plural word *mur*, as in *ure* 'thing' > *ure mur* 'things', *ngaunga* 'food' > *ngaunga mur* 'foods' (Pearson & Berg 2008: 34).

And as mentioned in Chapter 2, some languages have other number words other than plural words, and in the Yapese examples (repeated here as example 50).

[33] This reasoning about the development from a kinship term for respect to a plural marker is not quite clear. Cross-linguistically, a plural pronoun can be commonly used to indicate respect, but I am not aware of any development from a plural kinship term indicating respect to a plural marker in other languages.

(50) Yapese (Oceanic, Austronesian)

a. *ea* ***rea*** *kaarroo neey*
ART SG car this
'this car'

b. *ea* ***gäl*** *kaarroo neey*
ART DU car this
'these two cars'

c. *ea* ***pi*** *kaarroo ney*
ART PL car this
'these cars' (Jensen 1977: 155)

In these languages, the origins of plural words might also be very different from other pure plural words. As Corbett (2000: 267) suggests, number system other than singular-plural opposition can rise from numerals, but if this is the case in Yapese also needs to be evaluated by further research.

Note that Yapese does not only have number words with different number values, but also have two plural words for different functions. *Pi* in (50c) is only used in nouns phrases without classifiers, and when classifiers are present, another plural word *yuu* is used instead, as in (51). This is thus another distinctive feature of plural words in this language.

(51) *ea* ***yuu*** *kea* *niiw* *neey*
ART PL CLF coconut.tree this
'these coconut trees' (Jensen 1977: 155)

6.4. Interim summary and discussions

In this chapter, I discussed the diachronic developments of plural words in Austronesian languages and presented a general typology of their origins.

Nine languages (around 17%) have a plural word reflecting the reconstructed PMP and POc form **maŋa*, and twenty languages (around 37%) have a plural word originating from a certain type of third person plural pronoun. The origins of most plural words in my sample languages fall into either of these two categories; and if we included the possibly-related items as well, the percentages stand at 26% and 43% respectively. Other idiosyncratic origins are also present, as shown by the case of Toba Batak and Yapese, but they only make up less than one third of the total number, and it is also possible that some of them can be categorised into the first two types.

It has therefore shown that plural words in Austronesian languages do have a number of independent origins, confirming Dryer's (1989a: 885) speculation. But in the meanwhile, some general patterns can also be identified, and a number of plural

words in certain geographical areas and certain genealogical subgroups, such as Philippine languages in The Philippines and Timoric languages in east Indonesia, have shown close historical relatedness.

Based on this general picture, I present a background on which further studies can be conducted to investigate the histories of plural words in individual languages. I have also pointed out that while a plural word in PMP has been reconstructed as **maŋa*, and its reflexes can be found in several present-day daughter languages, the sparse presence of reflexes of **maŋa* raises some doubts for the reconstructed form. It does not mean that the reconstruction is necessarily incorrect, but we would need more evidence to explain why the daughter languages have lost it, and how they developed a plural word from other origins.

Finally, although I have shown that plural words in many Austronesian languages correspond to a certain type of third person plural pronoun, the exact mechanisms of these developments remain to be demonstrated.

Chapter 7. Conclusions

Since Dryer's (1989a) introduction of plural words in his pioneering work, very little attention has been paid to this particular type of nominal plurality marking, and no systematic work has been done to investigate plural words in any particular language family. While some observations about plural words in Austronesian languages can be found in earlier large-scale typological work, I have pointed out that these observations have to be taken as preliminary, since the Austronesian language samples could be largely improved. I have also shown that although Dryer's typology provides us with an overview of how nominal plurality can be marked in noun phrases, treating too many different grammatical elements all as plural words is not an ideal way to conduct cross-linguistic comparisons in a particular language family. Based on this background, this study explored the synchronic distribution pattern and diachronic developments of plural words in Austronesian languages. The definition of plural words has been revised, and a more extensive and more balanced Austronesian language sample has been collected. The foregoing investigation has offered answers to the research questions, and I shall formulate the conclusions here.

Firstly, 54 out of 128 Austronesian languages in my sample employ a plural word, and I have demonstrated that these languages with plural words display a skewed distribution, both genealogically and geographically. Genealogically, as far as the absolute number is concerned, plural words are mostly found in WMP, especially Philippine languages, and in Oceanic, especially Vanuatu languages. In terms of frequency, plural words are commonly used in CMP, Philippine languages in WMP, and Oceanic languages (particularly Vanuatu languages). Geographically, plural words are not evenly distributed either. There are several regions where a great number of languages with plural words can be found: The Philippines, north Borneo, east Indonesia and Vanuatu. In contrast, Austronesian languages with plural words are almost absent in west Indonesia, where Toba Batak is the only instance. Also, half of the Oceanic languages in my sample lack a plural word, and such an observation shows a contradiction to the observations in previous studies.

Secondly, plural words in Austronesian languages have a number of independent origins; but at the same time, they can be categorised into three major groups: plural words reflecting PMP **maŋa*, plural words originating from third person plural pronouns, and plural words with other miscellaneous origins. While a plural word **maŋa* has been reconstructed in PMP and POc, the results from the current study shows that reflexes of **maŋa* are only found in few present-day languages, most of them being Philippine languages. Twenty plural words in my sample correspond to a certain type of third person plural pronoun in the same language, which suggests a grammaticalisation path from a third person plural pronoun to a plural word. Apart from

reflecting **maŋa* or originating from a third person plural pronoun, other plural words with their own idiosyncratic developments are also present in my sample.

While offering new insights into the typology and history of plural words in Austronesian languages, this study also raises many issues.

The scanty presence of the reflexes of **maŋa* in present-day Austronesian languages calls for a reconsideration of the reconstruction in PMP and its daughter subgroups. If this reconstruction is correct, we still need to explain the fact that in most languages with plural words, this form was replaced by a different item.

In discussing plural words with reflexes of **maŋa*, I suggest a possibility that **maŋa* might be an originally bimorphemic item. This analysis remains hypothetical, and more data on other plural words (and plural markers in general) have to be brought together to confirm or reject this claim.

The variations in the grammaticalisation paths from a third person plural pronoun to a plural word suggest that while the overall pattern is shared, the precise diachronic development of plural words has to be investigated for each language individually.

To sum up, this thesis presents a new analysis of plural words in Austronesian languages based on new datasets. It is shown that the synchronic distribution of plural words in Austronesian languages is skewed, and the diachronic developments of these plural words also vary to some extent.

Appendix: Sample languages and their coding of nominal plurality

Number	Name	Primary classification	Coding of nominal plurality	Source	Notes
1	**Rukai**	Formosan	plural prefix	Zeitoun (2007: 170)	
2	**Puyuma**	Formosan	reduplication, plural suffix	Teng (2007: 142)	
3	**Tagalog**	Philippines – Greater Central Philippines (GCP) – Central Philippine	plural word	Schachter & Otanes (1972: 111–113)	*manga*, for non-personal nouns
4	**Bikol**	Philippines – GCP – Central Philippine	plural word	Mintz (1971: 99)	*mga*
5	**Mansaka**	Philippines – GCP – Central Philippine	plural word	Svelmoe & Svelmoe (1974: 49)	*manga*
6	**Mamanwa**	Philippines – GCP – Central Philippine	plural word	Miller & Miller (1976: 27)	*manga*
7	**Cebuano**	Philippines – GCP – Central Philippine	plural word	Tanangkingsing (2009: 55)	*mga*
8	**Central Tagbanwa**	Philippines – GCP – Palawanic	plural word	Scebold (2003: 60)	*manga*, abbreviated as *mga*
9	**Manobo**	Philippines – GCP – Manobo	plural word	Elkins (1970: 6–7)	West Bukidnon Manobo, *menge*
			plural word	Wang et al. (2006: 49)	Matigsalug Manobo, *me*

10	**Subanen**	Philippines – GCP – Subanen	plural clitic	Daguman (2013: 59, 85)	*na*, subject to glottalisation by a case marker, thus becoming an enclitic *ŋna-* or *ŋa-*
11	**Bontok**	Philippines – North Luzon	reduplication, plural prefix	Fukuda (1997: 22–23)	
12	**Kankanaey**	Philippines – North Luzon	plural word, reduplication	Allen (2014: 116–118)	*da*, as third person nominative pronoun, used before reference phrase marker; reduplication is possible
13	**Ibaloy**	Philippines – North Luzon	plural word, reduplication	Ruffolo (2004: 341)	*ira*, as third person plural ronoun
14	**Ilocano**	Philippines – North Luzon	reduplication	Rubino (1997: 73)	articles distinguish plurality
15	**Dupangingan Agta**	Philippines – North Luzon	plural word, reduplication	Robinson (2011: 69–71)	*hidi*, as third person plural nominative pronoun; reduplication is possible.
16	**Ayta Abenlen**	Philippines – Central Luzon	reduplication	Nitsch (2009: 10)	
17	**Tboli**	Philippines – Bilic	plural word	Porter (1977: 52)	*kem*
			plural word	Forsberg (1992: 10)	*kem*
18	**Tondano**	Philippines – Minahasan	no plural	Sneddon (1975: 115)	there is a noun class marker dedicated to animate plural nouns, *se*, in contrast with the marker for animate singular nouns, *si*
			reduplication	Brickell (2014: 73)	
19	**Chamorro**	Chamorro	plural word	Topping (1973: 234–235)	*siha*, as third person plural pronoun
20	**Palauan**	Palauan	plural prefix	Josephs (1975: 42–43)	*rę-*, only for human nouns

21	**West Coast Bajau**	Sama-Bajau	plural word	Miller (2007: 112–113)	*banga(n)/bengen/bongon*
22	**Madurese**	Malayo-Sumbawan (MS) – Madurese	reduplication	Davies (2010: 129–131)	
23	**Acehnese**	MS – North and East Malayo-Sumbawan (NEMS) – Aceh-Cham	no plural	Durie (1985)	
24	**Balinese**	MS – NEMS – Bali-Sasak-Sumbawa	reduplication	Clynes (1995: 203)	
25	**Indonesian**	MS – NEMS – Malayic	reduplication	Sneddon et al. (2010: 21)	
26	**Mualang**	MS – NEMS – Malayic	reduplication	Tjia (2007: 48)	
27	**Papuan Malay**	MS – NEMS – Malayic	reduplication	Kluge (2014: 171)	
28	**Salako**	MS – NEMS – Malayic	reduplication	Adelaar (2005b: 36)	
29	**Javanese**	Javanese	reduplication	Suharno (1982: 25)	
30	**Moklen**	Moken-Moklen	no plural	Larish (2005)	
31	**Bulungan**	North Borneo (NB) – Bulongan	reduplication	Adul, Maswan & Yazidi (1990: 33–34)	
32	**Ida'an**	NB – Northeast Sabahan	reduplication	Goudswaard (2005: 56)	there is a collectivity marker *iro*, meaning 'N and company' (Goudswaard 2005: 274); it is similar to the third person plural pronoun *(m)iro*

33	**Belait**	NB – North Sarawakan	reduplication	Noor Alifah (2004: 188)	
34	**Melanau**	NB – Sarawak-Melanau-Kajang	no plural	Blust (1988)	
35	**Murut**	NB – Southwest Sabahan	plural prefix	Prentice (1971: 118, 175–176)	*ŋaN-*, for various classes of metrical nouns, e.g. day, year, occur only in numeral phrases
36	**Tatana**	NB – Southwest Sabahan	plural proclitic	Dillon (1994: 24)	*ngo-*. Dillon considers it to be prefix, but when the classifier is present, *ngo-* attaches to the classifier, thus a proclitic.
37	**Kimaragang**	NB – Southwest Sabahan	plural word	Kroeger (2005: 411)	*tongo*
38	**Bundu Dusun**	NB – Southwest Sabahan	plural infix	Price (2007: 34)	*-ongo-*
39	**Kayan**	Kayanic	not clear	Clayre & Cubit (1974)	
40	**Mateq**	Land Dayak	no plural	Timothy (2013: 73)	
41	**Maanyan**	East Barito	no plural	Gudai (1985: 31)	
42	**Malagasy**	East Barito	no plural	Rasoloson & Rubino (2005: 45)	
43	**Seruyan**	West Barito	reduplication	Poerwadi, Iper & Purwaka (2003: 30)	
44	**Lampung**	Lampung	reduplication	Satun et al. (1985: 75)	
45	**Rejang**	Rejang	reduplication	Napsin et al. (1980: 38)	
46	**Toba Batak**	Northwest Sumatra/Barrier Islands	plural word	Tuuk (1971: 115) [1864]	*akka*
			plural word	Nababan (1981: 87)	*akka*
47	**Pendau**	Tomini-Tolitoli	plural word	Quick (2007: 190)	*ongo*, only for human nouns
48	**Kaili**	Kaili-Pamona	reduplication	Sofyan et al. (1979: 50–51)	

49	**Balantak**	Saluan	no plural	Berg & Busenitz (2012: 46, 174)	there is an associative plural *ari*
50	**Mori**	Bungku-Tokali	no plural	Esser & Mead (2011: 308)	adjectives can be marked by a prefix *ma-*. A nominal prefix *manga-* is fossilised, as in *mangalitau* 'youth'
51	**Tukang Besi**	Muna-Buton	no plural	Donohue (1999a: 346–347)	there is an 'inalienable plural marker' *mai*. When it is used with alienable things, it emphasises the plurality of the object
52	**Wolio**	Wolio-Wotu	plural word	Anceaux (1988: 36) [1952]	*maqa /maŋa/*
53	**Buginese**	South Sulawesi	not clear	Samsuri (1965)	third person pronouns can be overtly marked as plural by *manəŋ* 'all'
54	**Makassarese**	South Sulawesi	no plural	Jukes (2006)	
55	**Pitu Ulunna Salu**	South Sulawesi	not clear	Campbell (1989)	
56	**Batuley**	Aru	plural word	Daigle (2015: 128)	*je*, reduplication is possible, but not common; demonstratives also differentiate singularity from plurality
57	**Donggo**	Bima	reduplication	Sunihati et al. (1997: 109)	
58	**Mono**	Central Maluku (CM) – East Central Maluku (ECM) – Banda-Geser	plural prefix	Olson (2001: 91–93)	*à-*, for animate nouns
59	**Nuaulu**	CM – ECM – Nunusaku	plural suffix	Bolton (1990: 51–52)	suffixes vary for different word classes, e.g. *one* 'star' > *one-u* 'stars'

60	**Alune**	CM – ECM – Nunusaku	plural word	Wattimury, Haulussy & Pentury (1996: 27, 34)	*boka*, as in *manus boka* ‘chicken’ (*ayam-ayam* in Indonesian)
61	**Larike**	CM – ECM – Nunusaku	plural suffix	Laidig & Laidig (1991: 14, 36)	*-i* or *-u*
62	**Buru**	CM – West Central Maluku	plural clitic	Grimes (1991: 147)	*=ro*, identical to the third person undergoer pronoun
63	**Lamaholot Lewotobi**	Flores-Lembata	no plural	Nagaya (2012)	
64	**Kambera**	Flores-Sumba-Hawu	no plural	Klamer (1998)	articles can distinguish number
65	**Kéo**	Flores-Sumba-Hawu	no plural	Baird (2002)	
66	**Kei**	Kei-Tanimbar	reduplication	Tetelepta et al. (1985: 32–33)	
67	**Selaru**	Southern Southeast Maluku	plural clitic	Coward (1990: 22–23)	*=re*, having different morphological variations
68	**Leti**	Timoric A – Eastern Timoric A	plural clitic, reduplication	Engelenhoven (2004: 115–116)	*=ra*, for human nouns; non-human nouns mark plurality by repetition
69	**Tetun Dili**	Timoric A – Central Extra-Ramelaic	plural word	Williams-van Klinken, Hajek & Nordlinger (2002: 30)	*sira*, thrid-person plural pronouns, definite, not restricted to humans, also used as an associative plural
70	**Tugun**	Timoric A – Northern Timoric A	plural word, reduplication	Hinton (1991: 50)	*hira*, as third-person plural pronouns
71	**Southern Mambai**	Timoric B	plural word	Hull (2003: 11)	*sêr*, as Tetum *sira*

72	**Biak**	SHWNG – Cenderawasih Bay	no plural	Heuvel (2006: 201–204)	non-specific articles differentiate singular and non-singular
73	**Ambai**	SHWNG – Cenderawasih Bay	no plural	Silzer (1983: 93)	there is a plural definite article
74	**Taba**	SHWNG – Raja Ampat-South Halmahera	plural clitic	Bowden (2001: 190)	*=si*, only for human referents, as third person plural free pronouns
75	**Warembori**	SHWNG – Lower Mamberamo	plural suffix	Donohue (1999b: 11–12)	*-na*
76	**Irarutu**	SHWNG – Nabi-Irarutu	not clear	Jackson (2014: 141)	there is a plural marker mentioned, but no example. Reduplication is possible. Determiners and demonstratives also differentiate number
77	**Wuvulu**	Admiralty Islands	no plural	Hafford (2015: 66–67)	demonstratives distinguish number
78	**Paluai**	Admiralty Islands	plural word, reduplication	Schokkin (2014: 190–191)	*ip*, as third person plural pronouns. Reduplication is possible
79	**Loniu**	Admiralty Islands	plural word	Hamel (1994: 89–90)	*sɛh*, as third person plural pronoun
80	**Vaeakau-Taumako**	Central Pacific (CP) – East Fijian-Polynesian (EFP) – Polynesian	plural prefix	Næss & Hovdhaugen (2011: 145–147)	for kinship terms as well as ‘thing’, ‘root’
81	**Samoan**	CP – EFP – Polynesian	no plural	Mosel & Hovdhaugen (1992)	Dryer consider articles to be plural markers, e.g. non-specific plural article *ni*
82	**Hawaiian**	CP – EFP – Polynesian	plural word	Elbert & Pukui (1979: 162)	a number of different forms: *mau, po‘e, kau, wahi nāhi* and *ona*.
83	**Nadrogā**	CP – West Fijian	plural word	Geraghty (2002: 838)	*kura*, independent third person plural pronoun

84	**Dehu**	Loyalty Islands	plural word	Tryon (1968: 59)	*ite*
85	**Ponapean**	Micronesian	no plural	Rehg (1981: 144)	demonstratives differentiate number
86	**Satawalese**	Micronesian	no plural	Roddy (2007: 46)	determiners and demonstratives differentiate singularity from plurality.
87	**Unua**	North and Central Vanuatu (NCV) – Central Vanuatu (CV) - Malakula	plural word	Pearce (2015: 188)	*rin*
88	**Neve'ei**	NCV – CV – Malakula	plural word	Musgrave (2007: 68)	*ar*, as third person non-singular independent pronoun
89	**Tape (Maragus)**	NCV – CV – Malakula	plural word	Crowley (2006: 130)	*er*, as third person non-singular pronominal pronoun
90	**Abma**	NCV – CV – South Pentecost	plural word, reduplication	Schneider (2010: 65, 139)	*nii*, as third person plural object/independent pronoun. Reduplication to mark plurality is possible
91	**South Efate**	NCV – CV – Epi-Efate	no plural	Thieberger (2006: 107)	number is expressed by the pronominal markers on the verb
92	**Mavea**	NCV – Northern Vanuatu (NV) – Espiritu Santo	plural word, reduplication	Guérin (2011: 157–161)	*re*, for specific or referential human nouns. Reduplication is possible. Kinship terms can be marked by a prefix *na-*
93	**Tamambo**	NCV – NV – Espiritu Santo	plural prefix	Jauncey (2011: 113–114, 152)	*va-/vai-/ra-* for human nouns, *na-* for kinship terms; terms for trees are marked by prefixing *lo-*. Reduplication is possible but not productive

94	**Araki**	NCV – NV – Espiritu Santo	plural word, reduplication	François (2002: 37, 54, 80)	*dai*, reduplication is possible
95	**Mwotlap**	NCV – NV – Torres-Banks linkage	plural word	Crowley (2002: 591–592)	*iye*, there are also dual and trial words, demonstratives also mark plurality
96	**Raga**	NCV – NV – Hano	plural word, reduplication	Vari-Bogiri (2011: 81–82)	*ira*, reduplication is possible
97	**Wala**	Southeast Solomonic	plural word	Lovegren, Mitchell & Nakagawa (2015: 115)	*gi*, used with non-singular noun phrases, except those containing indefinite determiners
98	**Longgu**	Southeast Solomonic	plural clitic	Hill (2011: 224)	*-gi*
99	**Belep**	Southern Melanesian (SM) – New Caledonian (NC) – Extreme Northern	no plural	McCracken (2001: 281–282)	demonstratives and determiners distinguish number
100	**Tinrin**	SM – NC – Southern New Caledonian	plural prefix	Osumi (1995: 101)	*mê-*, only for human nouns
101	**Cèmuhî**	SM – NC – Cemuhî	no plural	Lynch (2002: 756–767)	number can be marked by articles
102	**Anejom̃**	SM – South Vanuatu	plural prefix	Lynch (2000: 50)	*eplu-/ilpu*, for animate nouns
103	**Mussau**	St. Matthias	plural word	Brownie & Brownie (2007: 64)	*atoa/etoa*, possibly related to *katoa* ‘some’
104	**Engdewu**	Temotu	plural clitic	Vaa (2013: 172)	*=yo*
105	**Vitu**	Western Oceanic linkage (WOL) – Meso Melanesian linkage (MML) – Bali-Vitu	reduplication	Berg & Bachet (2006: 32)	

106	**Kara-Lemakot**	WOL – MML – New Ireland-Northwest Solomonic linkage (NINSL)	plural word	Dryer (2013b: 70)	*maana*, greater plural. There are other number words, *mu* 'plural', *ro* 'dual' and *nu* 'paucal'
107	**Siar**	WOL – MML – NINSL	no plural	Frowein (2011: 120)	articles distinguish number.
108	**Ughele**	WOL – MML – NINSL	no plural	Frostad (2013: 50)	demonstratives distinguish number
109	**Kokota**	WOL – MML – NINSL	no plural	Palmer (2009: 84)	number can be marked by articles and demonstratives
110	**Teop**	WOL – MML – NINSL	plural word, reduplication	Mosel & Thiessen (2005)	*maa,* on Chapter 6.2
111	**Nakanai**	WOL – MML – Willaumez	plural word, reduplication	Johnston (1980: 154, 175)	*egite*, as third person plural pronoun. Reduplication is possible
112	**Bukawa**	WOL – North New Guinea linkage (NNGL) – Huon Gulf	no plural	Eckermann (2007: 27–28)	a very limited group of nouns can take enclitics *-i*, and there are also few nouns denoting plurality are formed with 'people' as a compound
113	**Jebem**	WOL – NNGL – Huon Gulf	no plural	Ross (2002: 277)	Kinship terms take the clitic *-i*
114	**Adzera**	WOL – NNGL – Huon Gulf	no plural	Holzknecht (1986: 104–105)	reduplication is possible but rare
115	**Mato**	WOL – NNGL – Ngero-Vitiaz linkage (NVL)	plural clitic	Stober (2013: 42)	*=di*, as third person pronoun object suffix form.
116	**Kove**	WOL – NNGL – NVL	no plural	Sato (2013: 134–135)	number can be marked on ajectives.
117	**Mangap-Mbula**	WOL – NNGL – NVL	plural word	Bugenhagen (1995: 201–203)	*zin*, as third person plural pronoun (nominative or accusative); *bizin*, for plurality of possessed item

118	**Lote**	WOL – NNGL – NVL	plural word	Pearson & Berg (2008: 33–34)	*mur*, can also give the meaning 'the group of', or following a name as an associative plural marker.
119	**Manam**	WOL – NNGL – Schouten	no plural	Lichtenberk (1983: 107, 266–268)	demonstratives and adjectives can take plural suffix
120	**Kairiru**	WOL – NNGL – Schouten	plural word	Wivell (1981)	*rri*, as third person plural pronoun, on various spots
121	**Tobati**	WOL – NNGL – Sarmi-Jayapura Bay	no plural	Donohue (2002: 192)	
122	**Maisin**	WOL – Papuan Tip linkage (PTL) – Nuclear Papuan Tip linkage (NPTL)	plural suffix	Frampton (2014: 68)	*-e/-r*, only for human nouns, general nouns do not have plural marking, reduplication can be noted but rare.
123	**Tawala**	WOL – PTL – NPTL	reduplication	Ezard (1997: 20)	mostly for human
124	**Koluwawa**	WOL – PTL – NPTL	plural suffix	Guderian & Guderian (2002: 16–18)	for animate nouns
125	**Gapapaiwa**	WOL – PTL – NPTL	no plural	McGuckin (2002: 300)	reduplication is confined to one small class of nouns referring to humans, including kinship terms. Number can be marked on verbs and adjectives
126	**Sinaugoro**	WOL – PTL – Peripheral Papuan Tip	plural clitic	Tauberschmidt (1999: 62–63)	*=ri*, as third person plural possessive pronoun

127	**Motu**	WOL – PTL – Peripheral Papuan Tip	no plural	Lister-Turner & Clark (n.d.: 27, 30)	demonstratives and articles can encode plurality. Very few nouns indicate by accent or reduplication the first syllable.
128	**Yapese**	Yapesic	plural word	Jensen (1977: 152–155)	*gäl* ‘dual’; *pi’* ‘plural’ used in noun phrases without classifiers; *yuu* ‘plural ’used before classifiers

References

Adelaar, K. Alexander. 2005a. The Austronesian languages of Asia and Madagascar: A historical perspective. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*. London/New York: Routledge.

Adelaar, K. Alexander. 2005b. *Salako or Badameà: Sketch grammar, text and lexicon of a Kanayatn dialect in West Borneo*. Wiesbaden: Harrassowitz Verlag.

Adriani, N. A. & M. L. Adriani-Gunning. 1908. *Hoofdstukken uit de spraakkunst van het Tontemboansch*. 's Gravenhage: Martinus Nijhoff.

Adul, M. Asfandi, Syukrani Maswan & A. Yazidi. 1990. *Morfologi dan sintaksis bahasa Bulungan*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Allen, Janet L. 2014. *Kankanaey: A role and reference gramamr analysis*. Dallas, Texas: SIL International.

Anceaux, J. C. 1987. *Wolio dictionary (Wolio-English-Indonesian)*. Dordrecht: Foris Publications.

Anceaux, J. C. 1988. *The Wolio language: Outline of grammatical description and texts*. Dordrecht: Foris Publications.

Baird, Louise. 2002. *A grammar of Kéo: An Austronesian language of East Nusantara*. PhD Thesis, The Australian National University.

Bell, Alan. 1978. Language samples. In Joseph H. Greenberg (ed.), *Universals of human language, Volume 1: Method and theory*. Stanford: Stanford University Press.

Berg, René van den & Peter Bachet. 2006. *Vitu grammar sketch*. Ukarumpa, Papua New Guinea: SIL Printing Press.

Berg, René van den & Robert L Busenitz. 2012. *A grammar of Balantak a language of Eastern Sulawesi*. SIL International.

Blust, Robert A. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *Working Papers in Linguistics of the University of Hawai'i* 9(2). 1–15.

Blust, Robert A. 1978. Eastern Malayo-Polynesian: a subgrouping argument. In S. A. Wurm & Lois Carrington (eds.), *Second International Conference on Austronesian Linguistics: proceedings, Fascicle I, Western Austronesian*, 181–234. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Blust, Robert A. 1988. Sketches of the morphology and phonology of Bornean languages 2: Mukah (Melanau). *Pacific Linguistics. Series A. Occasional Papers* 78. 151–216.

Blust, Robert A. 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32(2). 241–293.

Blust, Robert A. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected Papers from the Eighth International Conference on Austronesian Linguistics, Symposium Series of the Institute of Linguistics (Preparatory Office)*, 31–94. Taipei: Academia Sinica.

Blust, Robert A. & Stephen Trussel. 2010. Austronesian comparative dictionary. (www.trussel2.com/ACD) (Accessed 2016-11-03.)

Bolton, Rosemary Ann. 1990. *A preliminary description of Nuaulu phonology and grammar*. MA Thesis, The University of Texas at Arlington.

Bowden, John. 2001. *Taba: Description of a South Halmahera languages*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Brickell, Timothy C. 2014. *A grammatical description of the Tondano (Toundano) language*. PhD Thesis, La Trobe University.

Brownie, John & Marjo Brownie. 2007. *Mussau grammar essentials*. Ukarumpa, Papua New Guinea: SIL-PNG Academic Publications.

Bugenhagen, Robert D. 1995. *A grammar of Mangap-Mbula: An Austronesian language of Papua New Guinea*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Campbell, Philip J. 1989. *Some aspects of Pitu Ulunna Salu grammar: A typological approach*. MA Thesis, The University of Texas at Arlington.

Clayre, Beatrice M. & L. E. Cubit. 1974. An outline of Kayan grammar. *The Sarawak Museum Journal* 22(43). 43–91.

Clynes, Adrian. 1995. *Topics in the phonology and morphosyntax of Balinese based on a dialect of Singaraja, North Bali*. PhD Thesis, The Australian National University.

Connell, Timothy M. 2013. *A grammar sketch of Matéq: A Land Dayak language of West Kalimantan, Indonesia*. MA Thesis, University of Canterbury.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Coward, David Forrest. 1990. *An introduction to the grammar of Selaru*. MA Thesis, The University of Texas at Arlington.

Crowley, Terry. 2002. Mwotlap. In John Lynch, Malcolm Ross & Terry Crowley (eds.),

The Oceanic languages, 587–598. London/New York: Routledge.

Crowley, Terry. 2006. *Tape: a declining language of Malaluka (Vanuatu)*. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Daguman, Josephine Sanicas. 2013. *A grammar of Northern Subanen*. Muenchen: LINCOM EUROPA.

Daigle, Benjamin T. 2015. *A grammar sketch of Batuley: An Austronesian language of Aru, eastern Indonesia*. MA Thesis, Leiden University.

Davies, William D. 2010. *A grammar of Madurese*. Berlin/New York: De Gruyter Mouton.

Davis, Karen. 2003. *A grammar of the Hoava language, Western Solomons*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Dillon, John Albert. 1994. *A grammatical description of Tatana'*. MA Thesis, The University of Texas at Arlington.

Donohue, Mark. 1999a. *A grammar of Tukang Besi*. Berlin/New York: Mouton de Gruyter.

Donohue, Mark. 1999b. *Warembori gramamr sketch*. Unpublished Manuscript.

Donohue, Mark. 2002. Tobati. In John Lynch, Malcolm Ross & Terry Crowley (eds.), *The Oceanic languages*, 186–203. London/New York: Routledge.

Donohue, Mark & Charles E Grimes. 2008. Yet more on the position of the languages of eastern Indonesia and east Timor. *Oceanic Linguistics* 47(1). 114–158.

Dryer, Matthew S. 1986. Word order consistency and English. In Scott DeLancey & Russel Tomlin (eds.), *Proceedings of the Second Annual Pacific Linguistics Conference*, 97–106. Eugene, Oregon: University of Oregon.

Dryer, Matthew S. 1988. Object-verb order and adjective-noun order: Dispelling a myth. *Lingua*(74). 185–217.

Dryer, Matthew S. 1989a. Plural words. *Linguistics* 27(5). 865–896.

Dryer, Matthew S. 1989b. Large linguistic areas and language sampling. *Studies in Language* 13(2). 257–292.

Dryer, Matthew S. 2007. Noun phrase structure. In Timothy Shopen (ed.), *Language typology and syntactic description. Complex constructions, vol 2*, 151–205. Cambridge: Cambridge University Press.

Dryer, Matthew S. 2013a. Coding of nominal plurality. 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/33) (Accessed 2016-10-12.)

Dryer, Matthew S. 2013b. *A grammatical description of Kara-Lemakot*. Canberra:

Asia-Pacific Linguistics, College of Asia and the Pacific, The Australian National University.

Durie, Mark. 1985. *A grammar of Acehnese: On the basis of a dialect of north Aceh*. Dordrecht/Cinnaminson: Foris Publications.

Eckermann, W. 2007. *A descriptive grammar of the Bukawa language of the Morobe Province of Papua New Guinea*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Edwards, Owen. 2016. *Metathesis and unmetathesis: Parallelism and complementarity in Amarasi, Timor*. PhD Thesis, The Australian National University.

Elbert, Samuel H. & Mary Kawena Pukui. 1979. *Hawaiian grammar*. Honolulu: The University Press of Hawaii.

Elkins, Richard E. 1970. *Major grammatical patterns of Western Bukidnon Manobo*. University of Oklahoma: Summer Institute of Linguistics.

Engelenhoven, Aone van. 2004. *Leti: a language of Southwest Maluku*. Leiden: KITLV Press.

Esser, Samuel J. & David Mead. 2011. *Phonology and morphology of Mori*. SIL International.

Ezard, Bryan. 1997. *A grammar of Tawala: An Austronesian language of the Milne Bay area, Papua New Guinea*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Ferrell, Raleigh. 1969. *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. Taipei: Academia Sinica, Institute of Ethnology.

Forsberg, Vivian M. 1992. A pedagogical grammar of Tboli. (Ed.) Fe T. Otanes & Hazel Wrigglesworth. *Studies in the Philippine Linguistics* 9(1). Manila: Linguistic Society of the Philippines. 1–110.

Frampton, Joanna Margaret. 2014. *Maisin: A grammatical description of an Oceanic language in Papua New Guinea*. MA Thesis, The University of Otago.

François, Alexandre. 2002. *Araki: A disappearing language of Vanuatu*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

François, Alexandre. 2015. Trees, waves and linkages: Models of language diversification. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 161–189. London/New York: Routledge.

Frostad, Benedicte Haraldstad. 2013. *A grammar of Ughele: An oceanic language of Solomon Islands.* Utrecht: LOT.

Frowein, Friedel Martin. 2011. *A grammar of Siar, an Oceanic language of New Ireland Province, Papua New Guinea*. PhD Thesis, La Trobe University.

Fukuda, Takashi. 1997. A discourse-oriented grammar of Eastern Bontoc. *Studies in the Philippine Linguistics* 10(1). Manila: Summer Institute of Linguistics. 1–116.

Geraghty, Paul. 2002. Nadrogâ. In John Lynch, Malcolm Ross & John Lynch (eds.), *The Oceanic languages*, 833–847. London/New York: Routledge.

Goudswaard, Nelleke. 2005. *The Begak (Ida'an) language of Sabah*. Utrecht: LOT.

Greenberg, Joseph H. 1963. Some universals of grammar, with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Univerals of language*, 58–90. Cambridge, Massachusetts: The M.I.T. Press.

Grimes, Charles E. 1991. *The Buru Language of Eastern Indonesia*. Phd Thesis, The Australian National University.

Gudai, Darmansyah H. 1985. *A grammar of Maanyan: A language of Central Kalimantan*. PhD Thesis, The Australian National University.

Guderian, Brad & Toni Guderian. 2002. *Koluwawa grammar essentials*. Summer Institute of Linguistics.

Guérin, Valérie. 2011. *A grammar of Mavea: An Oceanic language of Vanuatu*. Honolulu: University of Hawai'i Press.

Hafford, James A. 2015. *Wuvulu grammar and vocabulary*. PhD Dissertation, The University of Hawai'i at Mānoa.

Hamel, Patricia J. 1994. *A grammar and lexicon of Loniu, Papua New Guinea*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2016. Glottolog 2.7. Jena: Max Planck Institute for the Science of Human History. (http://glottolog.org) (Accessed 2016-09-22.)

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Heuvel, Wilco van den. 2006. *Biak: Description of an Austronesian language of Papua*. Utrecht: LOT.

Hill, Deborah. 2011. *Longgu grammar*. Muenchen: LINCOM EUROPA.

Hinton, Bryan Douglas. 1991. *Aspects of Tugun phonology and syntax*. MA Thesis, The University of Texas at Arlington.

Holzknecht, Susanne. 1986. A morphology and grammar of Adzera (Amari dialect), Morobe Province, Papua New Guinea. *Papers in New Guinea Linguistics No. 24*, 77–166. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Hull, Geoffrey. 2003. *Southern Mambai*. Díli, Timor Leste: Instituto Nacional de Linguística, Universidade Nacional Timor Lorosa'e.

Hull, Geoffrey & Lance Eccles. 2001. *Tetum reference grammar*. Winston Hills, N.S.W., Australia: Sebastião Aparício da Silva Project.

Jackson, Jason Alexander Johann. 2014. *A grammar of Irarutu, a language of West Papua, Indonesian, with historical anlaysis*. PhD Dissertation, University of Hawai'i at Mānoa.

Jauncey, Dorothy G. 2011. *Tamambo, the language of West Malo, Vanuatu*. Canberra: Pacific Linguistics, College of Asia and the Pacific, The Australian National University.

Jensen, John Thayer. 1977. *Yapese reference grammar*. Honolulu: University of Hawaii Press.

Johnston, Raymond Leslie. 1980. *Nakanai of New Britain: the grammar of an Oceanic language*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Josephs, Lewis S. 1975. *Palauan reference grammar*. Honolulu: The University Press of Hawaii.

Jukes, Anthony. 2006. *Makassarese (basa Mangkasara'): A description of an Austronesian language of South Sulawesi*. PhD Thesis, The University of Melbourne.

Kibort, Anna & Greville G. Corbett. 2008. Grammatical Features Inventory: Number. University of Surrey. (http://dx.doi.org/10.15126/SMG.18/1.02) (Accessed 2016-10-05.)

Kikusawa, Ritsuko. 2015. The Austronesian language family. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 657–674. London/New York: Routledge.

Klamer, Marian. 1998. *A grammar of Kambera*. Berlin/New York: Mouton de Gruyter.

Kluge, Angela. 2014. *A grammar of Papuan Malay*. Utrecht: LOT.

Kratochvíl, František. 2007. *A grammar of Abui: A Papuan language of Alor*. Utrecht: LOT.

Kroeger, Paul. 2005. Kimaragang. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 397–428. London/New York: Routledge.

Laidig, Wyn D. & Carol J. Laidig. 1991. *Larike grammar*. Ambon: UNPATTI - SIL.

Larish, Michael D. 2005. Moken and Moklen. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 513–533. London/New York: Routledge.

Lehmann, Winfred P. 1973. A structural principle of languages and its implications. *Language* 49. 47–66.

Lichtenberk, Frantisek. 1983. *A grammar of Manam*. Honolulu: University of Hawaii Press.

Lister-Turner, R. & J. B. Clark. *A grammar of the Motu language of Papua*. Sydney, New South Wales: A. H. Pettifer.

Lovegren, Jesse, Alice Mitchell & Natsuko Nakagawa. 2015. *The Wala language of Malaita, Solomon Islands*. Canberra: Asia-Pacific Linguistics, College of Asia and the Pacific, The Australian National University.

Lynch, John. 2000. *A grammar of Anejom̃.* Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Lynch, John. 2002. Cèmuhî. In John Lynch, Malcolm Ross & Terry Crowley (eds.), *The Oceanic languages*, 753–764. London/New York: Routledge.

Lynch, John, Malcolm Ross & Terry Crowley. 2002. *The Oceanic languages*. London/New York: Routledge.

McCracken, Chelsea Leigh. 2001. *A Grammar of Belep*. PhD Thesis, Rice University.

McGuckin, Catherine. 2002. Gapapaiwa. In John Lynch, Malcolm Ross & Terry Crowley (eds.), *The Oceanic languages*, 297–321. London/New York: Routledge.

Miller, Amy. 2001. *A grammar of Jamul Tiipay*. Berlin/New York: Mouton de Gruyter.

Miller, Jeanne & Helen Miller. 1976. *Mamanwa grammar*. Manila: Summer Institute of Linguistics.

Miller, Mark T. 2007. *A grammar of West Coast Bajau*. PhD Thesis, The University of Texas at Arlington.

Mintz, Malcolm W. 1971. *Bikol grammar notes*. Honolulu: University of Hawaii Press.

Mosel, Ulrike & Even Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Scandinavian University Press.

Mosel, Ulrike & Yvonne Thiessen. 2005. *The Teop sketch grammar*. Unpublished Manuscript.

Musgrave, Jill. 2007. *A grammar of Neve'ei, Vanuatu*. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Nababan, P. W. J. 1981. *A grammar of Toba-Batak (Materials in languages of Indonesia, No. 6)*. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Næss, Åshild & Even Hovdhaugen. 2011. *A grammar of Vaeakau-Taumako*. Berlin/Boston: De Gruyter Mouton.

Nagaya, Naonori. 2012. *The Lamaholot language of Eastern Indonesia*. PhD Thesis, Rice University.

Napsin, Syahrul, Zainal Abidin Naning, Slamet Abdullah, Sjafran Sjamsuddin, Mohammad Arsyad & Tarmizi. 1980. *Morfologi dan sintaksis bahasa Rejang*.

Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Nitsch, Wilhelm. 2009. Some Ayta Abenlen grammar. *SIL Language and Culture Documentation and Description 3*. SIL International.

Noor Alifah, binti Abdullah. 2004. *Struktur bahasa Belait*. Berakas, Brunei Darussalam: Dewan Bahasa dan Pustaka, Kementerian Kebudayaan, Belia and Sukan.

Olson, Kenneth S. 2001. *The phonology and morphology of Mono*. PhD Dissertation, The University of Chicago.

Osumi, Midori. 1995. *Tinrin grammar*. Honolulu: University of Hawai'i Press.

Palmer, Bill. 2009. *Kokota grammar*. Honolulu: University of Hawai'i Press.

Pawley, Andrew. 1999. Chasing rainbows: implications for the rapid dispersal of Austronesian languages for subgrouping and reconstruction. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected Papers from the Eighth International Conference on Austronesian Linguistics, Symposium Series of the Institute of Linguistics (Preparatory Office)*. Taipei: Academia Sinica.

Pearce, Elizabeth. 2015. *A grammar of Unua*. Boston/Berlin/Munich: De Gruyter Mouton.

Pearson, Gerg & René van den Berg. 2008. *Lote grammar sketch*. Ukarumpa, Papua New Guinea: SIL-PNG Academic Publications.

Plungian, Vladimir A. 1995. *Dogon (Languages of the world: Materials 64)*. Muenchen: Lincom Europa.

Poerwadi, Petrus, Dunis Iper & A. Purwaka. 2003. *Morfologi bahasa Seruyan*. Jakarta: Pusat Bahasa, Departemen Pendidikan Nasional.

Porter, Doris. 1977. *A Tboli grammar*. Manila: Linguistic Society of the Philippines.

Prentice, D. J. 1971. *The Murut languages of Sabah*. Canberra: Pacific Linguistics, Department of Linguistics, The Australian National University.

Price, Daniel Charles. 2007. *Bundu Dusun sketch grammar*. Unpublished Manuscript.

Quick, Phil. 2007. *A grammar of the Pendau language of central Sualwesi, Indonesia*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Rasoloson, Janie & Carl Rubino. 2005. Malagasy. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 456–488. London/New York: Routledge.

Rehg, Kenneth L. 1981. *Ponapean reference grammar*. Honolulu: The University Press of Hawaii.

Rijkhoff, Jan & Dik Bakker. 1998. Language sampling. *Linguistic Typology* 2. 263–

314.

Rijkhoff, Jan, Dik Bakker, Kees Hengeveld & Peter Kahrel. 1993. A method of language sampling. *Studies in Language* 17(1). 169–203.

Robinson, Laura C. 2011. *Dupaningan Agta: Grammar, vocabulary, and texts*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Roddy, Kevin M. 2007. *A sketch grammar of Satawalese, the language of Satawal Island, Yap State, Micronesia*. MA Thesis, The University of Hawai'i.

Ross, Malcolm D. 1995. Some current issues in Austronesian linguistics. In Darrell Tryon, Malcolm D. Ross, Charles E. Grimes, Adrian Clynes & K. A. Adelaar (eds.), *Comparative omparative Austronesian dictionary. An introduction to Austronesian studies*. Berlin/New York: Mouton de Gruyter.

Ross, Malcolm. 2002. Jabêm. In John Lynch, Malcolm Ross & Terry Crowley (eds.), *The Oceanic languages*, 270–296. London/New York: Routledge.

Ross, Malcolm. 2009. Proto Austronesian verbal morphology: a reappraisal. In K. Alexander Adelaar & Andrew Pawley (eds.), *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 295–326. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Ross, Malcolm. 2012. In defense of Nuclear Austronesian (and against Tsouic). *Language and Linguistics* 13(6). 1253–1330.

Rubino, Carl Ralph Galvez. 1997. *A reference grammar of Ilocano*. PhD Dissertation, University of California, Santa Barbara.

Ruffolo, Roberta. 2004. *Topics in the morpho-Syntax of Ibaloy, Northern Philippines*. PhD Thesis, The Australian National University.

Samsuri. 1965. *Introduction to a Rappang Buginese grammar*. PhD Thesis, Indiana University.

Sato, Hiroko. 2013. *Grammar of Kove: An Austronesian language of the West New Britain province, Papua New Guinea*. PhD Dissertation, The University of Hawai'i at Manoa.

Satun, Aidy Ruslan, R. M. Arif, Lamsari, Kasmansyah & Erhamma. 1985. *Struktur bahasa Lampung*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Scebold, Robert A. 2003. *Central Tagbanwa: A Philippine language on the brink of extinction*. Manila: Linguistic Society of the Philippines.

Schachter, Paul & Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.

Schneider, Cynthia. 2010. *A grammr of Amba: A language of Pentecost Island, Vanuatu*.

Canberra: Pacific Linguistics, College of Asia and the Pacific, The Australian National University.

Schokkin, Gerda Hendrike. 2014. *A grammar of Paluai: The language of Balluan Island, Papua New Guinea*. PhD Thesis, James Cook University.

Silzer, Peter James. 1983. *Ambai: An Austronesian language of Irian Jaya, Indonesia*. PhD Thesis, The Australian National University.

Sneddon, James Neil. 1975. *Tondano phonology and grammar*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Sneddon, James Neil, Alexander Adelaar, Dwi Noverini Djenar & Michael C. Ewing. 2010. *Indonesian: A comprehensive grammar*. London/New York: Routledge.

Sofyan, Inghuong Alias, Syahruddin Kaseng, M. Sikki & Patuko Pepy. 1979. *Morfologi dan sintaksis bahasa Kaili*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Stober, Scot F. 2013. *Mato Grammar Sketch*. Ukarumpa, Papua New Guinea: SIL-PNG Academic Publications.

Suharno, Ignatius. 1982. *A descriptive study of Javanese*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Sunihati, Anak Agung Dewi, I. Made Denes, Ida Ayu Mirah Purwiati & Ni Luh Komang Candrawati. 1997. *Struktur bahasa Donggo*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Svelmoe, Gordon & Thelma Svelmoe. 1974. *Notes on Mansaka grammar*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

Tanangkingsing, Michael. 2009. *A functional reference grammar of Cebuano*. PhD Dissertation, National Taiwan University.

Tauberschmidt, Gerhard. 1999. *A grammar of Sinaugoro: An Austronesian language of the Central Province of Papua New Guinea*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Teng, Stacy Fang-Ching. 2007. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. PhD Thesis, The Australian National University.

Tetelepta, J., E. A. Apituley, J. Kasihuw, E. Wattimury & I. Talakua. 1985. *Struktur bahasa Kei*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Thieberger, Nicholas. 2006. *A grammar of South Efate: An Oceanic language of Vanuatu*. Honolulu: University of Hawai'i Press.

Tjia, Johnny. 2007. *A grammar of Mualang*. Utrecht: LOT.

Tomlin, Russell. 1986. *Basic constituent orders: functional principles*. London: Croom-

Helm.
Topping, Donald M. 1973. *Chamorro reference grammar*. Honolulu: The University Press of Hawaii.
Tryon, D. T. 1968. *Dehu grammar*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
Tuuk, H. N. van der. 1971. *A grammar of Toba Batak*. The Hague: Martinus Nijhoff.
Vaa, Anders. 2013. *A grammar of Engdewu: An Oceanic language of the Solomon Islands*. PhD Thesis, Oslo University.
Vari-Bogiri, Hannah. 2011. *Phonology and morpho-syntax of Raga, Vanuatu*. PhD Thesis, The University of the South Pacific.
Wang, Peter, Robert Hunt, Jeff McGriff & Richard E. Elkins. 2006. *The Grammar of Matigsalug Manobo*. Summer Institute of Linguistics.
Wattimury, E., Ny. A. Haulussy & J. Pentury. 1996. *Struktur bahasa Alune*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.
Williams-van Klinken, Catharina, John Hajek & Rachel Nordlinger. 2002. *Tetun Dili: A grammar of an East Timorese language*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
Wivell, Richard. 1981. *Kairiru grammar*. MA Thesis, The University of Auckland.
Zeitoun, Elizabeth. 2007. *A grammar of Mantauran (Rukai)*. Taipei: Institute of Linguistics, Academia Sinica.