Emory Undergraduate Linguistics Conference (EULC) 7

Beyond Decimal

Typological Patterns and Hierarchies in Numeral Formation (1–10) across 900 Austronesian Languages

LU, NAI-WEI

Emory, April 17, 2026

Outline

01

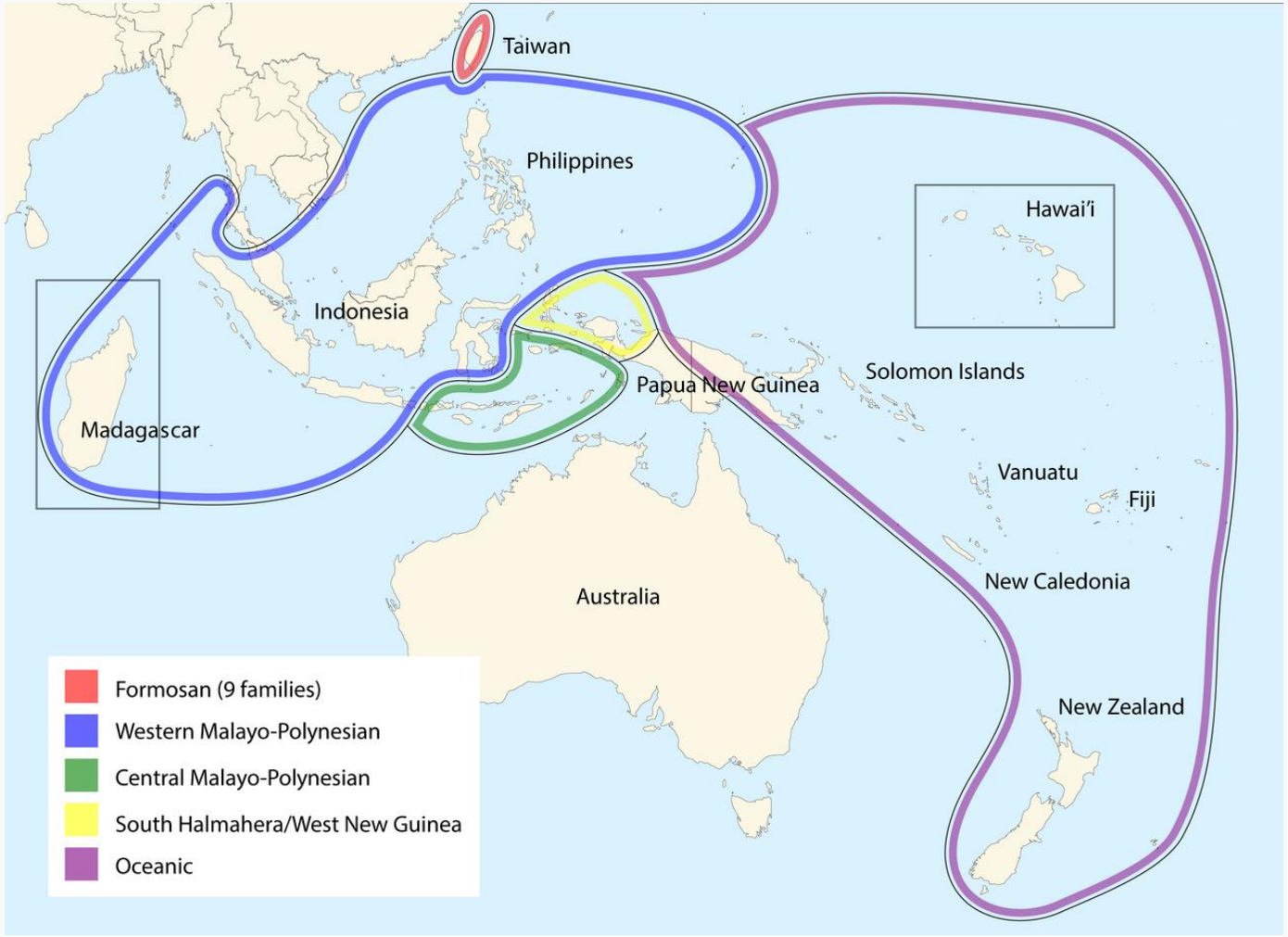
Introduction

Austronesian

- 1200+ languages

	Proto-Austronesian
1	*esa/isa
2	*duSa
3	*telu
4	*Sepat
5	*lima
6	*enem
7	*pitu
8	*walu
9	*Siwa
10	*sa-puluq

[The Numbers List]. https://www.zompist.com/numbers.shtml



By Stefano Coretta - Own work. Base map from https://www.amcharts.com/svg-maps/?map=worldIndia, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=42753023

Motivation

Prior Work

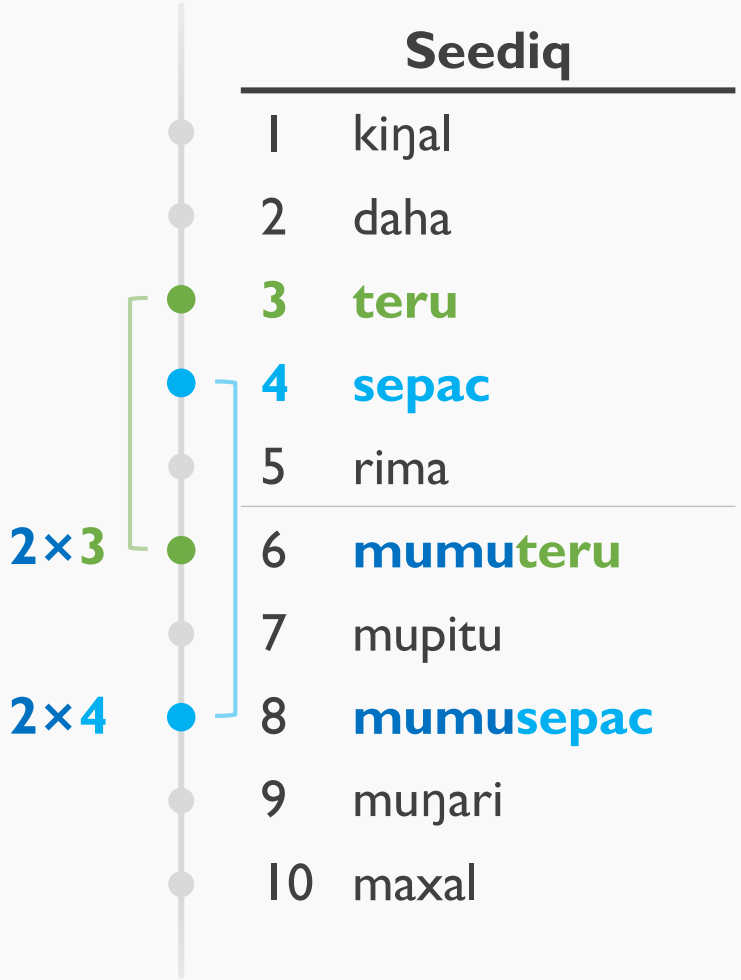
- Li (2024): Formosan numerals — predominantly decimal, quinary also attested

Motivating Observation

- Seediq: 6 = 2×3, 8 = 2×4
- Are such strategies attested across the wider Austronesian family?

Research Gap

- No systematic typological survey of special expression strategies across non-Formosan languages



Research Questions & Aims

How **numerals 1–10** are expressed in Austronesian languages outside of Taiwan?

1. What **base systems** these languages employ and their distributional frequency?

2. Whether **special expressions** for 6–9 exist beyond what the base system predicts?

02

Methodology

Typological approach

- lexical lists of numerals **1–10**
- documented in **the Numbers List** (Rosenfelder, Mark.)
- **900** Austronesian languages https://www.zompist.com/numbers.shtml

Austronesian										
+Proto-Austronesian	*esa/isa	*duSa	*telu	*Sepat	*lima	*enem	*pitu	*walu	*Siwa	*sa-puluq
Atayalic										
Atayal	qutux	saziŋ	ciwal	payat	maɣal	tizyuʔ	pituʔ	spat	qeruʔ	mpuw
Sedeq (Taroko)	kiŋal	daha	teru	sepac	rima	mumuteru	mupitu	mumusepac	muŋari	maxal
Tsouic										
Rukai Taloma	əa	Dosa	toLó	soʔátə	Limá	ənámə	ʔitó	vaLó	baŋátə	poLóko
Bud	iθa	Dósa	tóLo	səpátə	Líma	ánəmə	píto	vaLo	báŋatə	póLoko
Maga	Duka	Dúsa	túru	pátə	ríma	nɨmɨ	pítu	váru	vŋátɨ	prúku
Tona	nága	dóθa	tóo	pátə	imá	námə	píto	vaó	vaŋátə	pó:ko
Mantauran	nə́ka	nosa	toLo	patə	Lima	nəmə	pito	vaLo	vaŋatə	poLoko
Tsou	cóni	yúso	túyu	sə́ptə	eímo	nómə	pítu	vóyu	sío	máskə
Kanakanabu	caáni	cuúsa	tuúlu	suúpata	liíma	n´əmə	piítu	áalu	síia	máanə
Saaroa	caani	suua	tuulu	paatə	kulima	kənəmə	kupitu	kualu	kusia	kumaatə
Paiwanic										
Bunun	tasʔaʔ	dusaʔ	tau	paat	himaʔ	nuum	pituʔ	vauʔ	sivaʔ	macʔan
Paiwan	ita	ùsa	cəl. u	səpac	l. ima	unəm	picu	al. u	siva	ta-pul. uq
Puyuma	saʔ	Dua	təLu	pat	Lima	nəm	pitu	waLu	iwa	puLuʔ
Saisiyat	ʔähäʔ	rošaʔ	tooʔ	šəpat	asəb	sayboši:	sayboši o ʔähäʔ	kašpat	ääʔhäʔ	laŋpəz
Ami	cɨcay	tusaʔ	tuluʔ	spat	limaʔ	ʔnəm	pituʔ	faluʔ	siwaʔ	ŋuʔtəp
Sinicized										
Kavalan	ʔissaʔ	zusaʔ	tuLuʔ	spát	Limaʔ	ʔənəm	pituʔ	waLuʔ	siwaʔ	staRay
Pazeh	ʔidaʔ	dusaʔ	turuʔ	supat	xasəp	xasəbuzaʔ	xasəbiłdusaʔ	xasəbiłturuʔ	xasəbiłsupat	ʔisit
Thao	ta:haʔ	tu:šaʔ	to:roʔ	špa:t	ri:maʔ	kato:roʔ	pi:tuʔ	kašpat	tana:θuʔ	ma:kθin
+Ketangalan	tsa	Lusa	tsʰu:	špat	tsima	anum	pitu	watsu	siwa	Labatan
+Taokas	ta:nu	rua	tuLu	lepat	hasap	tahap	jweto	mahalpat	tanasu	taisid
+Papora	tanu	nia	tuLu	nepat	nema	nenim	pitu	mahal	mesia	metsi
+Babuza	natta	roa	torro	naspat	naxab	natap	naito	maaspat	tannaxo	tsxiet

Morphological analysis

- All 900 languages are compiled into a **database** with expression strategies annotated accordingly.

A	B	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE
編號	語言名稱	4=2+2	4=2*2	6=4+2	7=4+3	6=5+1	7=5+2	8=5+3	9=5+4	10=5+5	10=2*5	7=6+1	7=8-1	8=6+2	8=7+1	9=8+1	6=10-4	7=10-3	8=10-2	9=10-1
733	Nauruan					0	0	0	0			0								0
734	Gilbertese (Kiribati)					0	0	0	0			0								0
735	*Fabre 1847*					0	0	0	0			0								0
736	Marshallese					0	0	0	0			0								0
737	Kusaiean (Kosraean)					0	0	0	0			0								0
738	Mokilese					0	0	0	0			0								0
739	Ponapean					0	0	0	0			0								0
740	Trukese					0	0	0	0			0								0
741	Puluwat					0	0	0	0			0								0
742	Carolinian					0	0	0	0			0								0
743	Ulithian					0	0	0	0			0								0
744	*Sonsorol*					0	0	0	0			0								0
745	*Pulo Anna*					0	0	0	0			0								0
746	*Merir*					0	0	0	0			0								0
747	*Tobi*					0	0	0	0			0								0
748	*Woleai*					0	0	0	0			0								0
749	Mapia					0	0	0	0			0								0
750	Qae (Visale)					0	0	0	0			0								0
751	Di (Vaturanga, Ghara)					0	0	0	0			0								0

03

Base systems

Base systems

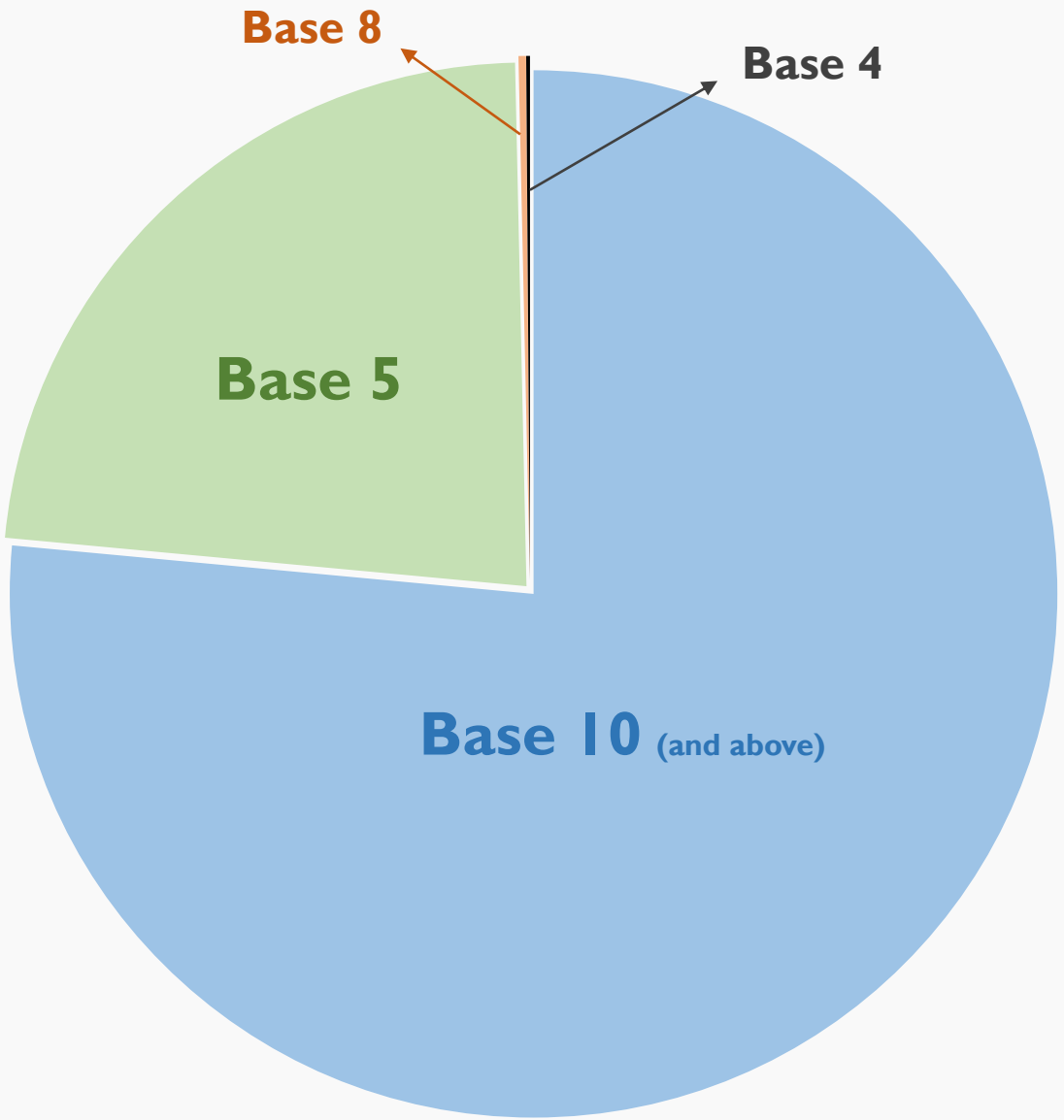
in 900 Austronesian Languages

Base 10 (and above)	75%~80%
Base 5	20%~25%
Base 8	<0.3% (2 languages)
Base 4	<0.2% (1 languages)

~~Base 6~~

~~Base 7~~

~~Base 9~~



Base systems

Examples of different base systems (4, 5, 8, 10)

	Base 4	Base 5	Base 8	Base 10
	Bam	**Arawe (Solong)**	**Yotafa (Tobati)**	**Halia**
1	tini	noneke	tei	toa
2	ru	nokip	rɔs	huol
3	tual	nomoyok	tor	topisa
4	**kiki**	nopeng	ʌu	tohats
5	**[4]** be kubua	nolim	mənіʌm	tolima
6	**[4]** be areg di ru	**elmo**ke	mʌndosim	tönomo
7	**[4]** be areg di tuol	**elmo**kip	tiʌmgʌɬrɔnduk	tohit
8	**kiki** ru	**elmo**yok	**rɔnduk**	toal
9	**[4]** be kubua	**elmo**peng	**rɔnduk** mʌnitiʌm	tosie
10		esugul	**rɔnduk** mʌnirɔsim	maloto

Base systems

Examples of different base systems (4, 5, 8, 10)

	Base 4	Base 5	Base 8	Base 10
	Bam	**Arawe (Solong)**	**Yotafa (Tobati)**	**Halia**
1	tini	noneke	tei	toa
2	ru	nokip	rɔs	huol
3	tual	nomoyok	tor	topisa
4	kiki	nopeng	ʌu	tohats
5	[4] be kubua	nolim	məniʌm	tolima
6	[4] be areg di ru	elmoke	mʌndosim	tönomo
7	[4] be areg di tuol	elmokip	tiʌmgʌɬrɔnduk	tohit
8	kiki ru	elmoyok	rɔnduk	toal
9	[4] be kubua	elmopeng	rɔnduk mʌnitiʌm	tosie
10		esugul	rɔnduk mʌnirɔsim	maloto

Others

Non-base numeral constructions

	Sissano	Sirasira
1	pontanen	tangkua
2	**entin**	**iruk**
3	entin e pontenen	irikiruk ena tango mangan
4	[2] ke [2]	[2] da [2]
5	[2] ke [2] ke [1]	[2] da [2] a mangan
6	[2] ke [2] ke [2]	[2] a mangan [2] a mangan
7	[2] ke [2] ke [2] [1]	no data
8	[2] ke [2] ke [2] ke [2]	
9	[2] ke [2] ke [2] ke [2] ke [1]	
10	[2] ke [2] ke [2] ke [2] ke [2]	

Others

Non-base numeral constructions

6=2+2+2

Sissano

1	**pontanen**
2	**entin**
3	**entin** e **pontenen**
4	[2] ke [2]
5	[2] ke [2] ke [1]
6	[2] ke [2] ke [2]
7	[2] ke [2] ke [2] [1]
8	[2] ke [2] ke [2] ke [2]
9	[2] ke [2] ke [2] ke [2] ke [1]
10	[2] ke [2] ke [2] ke [2] ke [2]

6=2+1+2+1

Sirasira

1	**tangkua**
2	**iruk**
3	irik**iruk** ena tango **mangan**
4	[2] da [2]
5	[2] da [2] a **mangan**
6	[2] a **mangan** [2] a **mangan**
7	no data
8	
9	
10	

04

Special expression strategies

Special expression strategies

Overview

- Addition
- Subtraction
- Multiplication
- ~~Division~~
  (no attested cases found in the dataset)

Addition

4=2+2

		Base 5	Base 10
		Gapapaiwa	Fagululu
	1	sago	ta‘aga
	2	**ruwa**	**luwa**
	3	aroba	toi
2+2	4	**ruwa** ma **ruwa**	**luwa** ta **luwa**
	5	mikovi	faibi
	6	[5] ma sago	sikisi
	7	[5] ma ruwa	seven
	8	[5] ma aroba	eiti
	9	[5] ma ruwa ma ruwa	naen
	10	ima ruma	ten

Subtraction

	Levei (Khehek)	Gele	Taliabo
1	o:ri	sih	siya
2	luo	rueh	howo
3	taloh	teloh	tolu
4	ha:hu	hahu	nga
5	li:me	limueh	lima
6	djaha:hu	anoh	nong
7	djodoloh	ndroteloh	hitu
8	djolue	ndrorueh	walu
9	djoʔeri	ndrosih	tasiya
10	rono	sunguah	hulu

Subtraction

		Levei (Khehek)	Gele	Taliabo
	1	o:ri	sih	siya
	2	luo	rueh	howo
	3	taloh	teloh	tolu
	4	ha:hu	hahu	nga
	5	li:me	limueh	lima
10-4	6	djaha:hu	anoh	nong
10-3	7	djodoloh	ndroteloh	hitu
10-2	8	djolue	ndrorueh	walu
10-1	9	djoʔeri	ndrosih	tasiya
	10	rono	sunguah	hulu

Subtraction

Hierarchy Hypothesis I

10-4	10-3, 10-2	10-1	Levei (Khehek)
	10-3, 10-2	10-1	Gele
		10-1	Taliabo

Greenberg, Joseph H (1978)

If a number n is expressed by subtraction as y- x , then every number z (z > y > n) is also expressed subtractively and with y as the minuend.

10-4
10-3
10-2
10-1

	Levei (Khehek)	Gele	Taliabo
1	o:ri	sih	siya
2	luo	rueh	howo
3	taloh	teloh	tolu
4	ha:hu	hahu	nga
5	li:me	limueh	lima
6	djaha:hu	anoh	nong
7	djodoloh	ndroteloh	hitu
8	djolue	ndrorueh	walu
9	djoʔeri	ndrosih	tasiya
10	rono	sunguah	hulu

Multiplication

		Thao	Ujir
	1	ta:haʔ	set
	2	tu:šaʔ	rúa
	3	**to:roʔ**	láti
	4	**špa:t**	**ká**
	5	ri:maʔ	líma
2×**3**	6	ka**to:roʔ**	dúau
	7	pi:tuʔ	dubusam
2×**4**	8	ka**špat**	**ka**rúa
	9	tana:θuʔ	t́ɛ́ra
	10	ma:kθin	uysía

Multiplication

Hierarchy Hypothesis II (original)

2×3	2×4
	2×4

Thao

Ujir

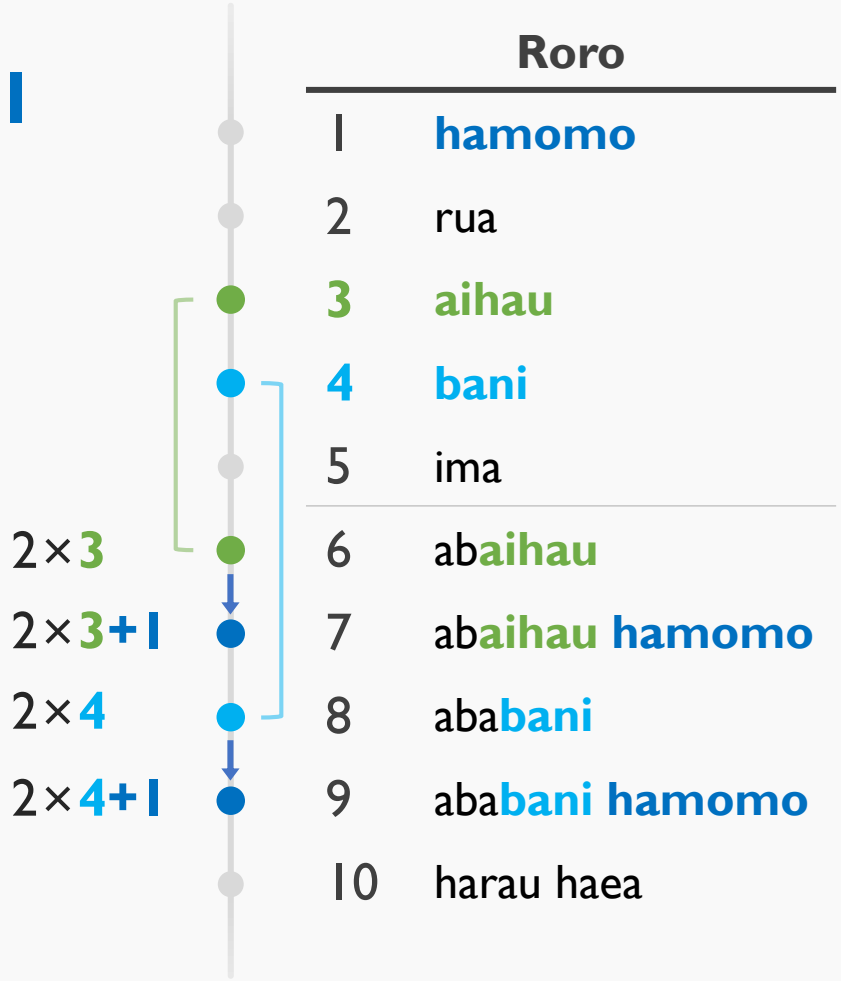
		Thao	Ujir
	1	ta:haʔ	set
	2	tu:šaʔ	rúa
	3	to:roʔ	láti
	4	špa:t	ká
	5	ri:maʔ	líma
2×3	6	kato:roʔ	dúau
	7	pi:tuʔ	dubusam
2×4	8	kašpat	karúa
	9	tana:θuʔ	t́ɛ́ra
	10	ma:kθin	uysía

Multiplication II

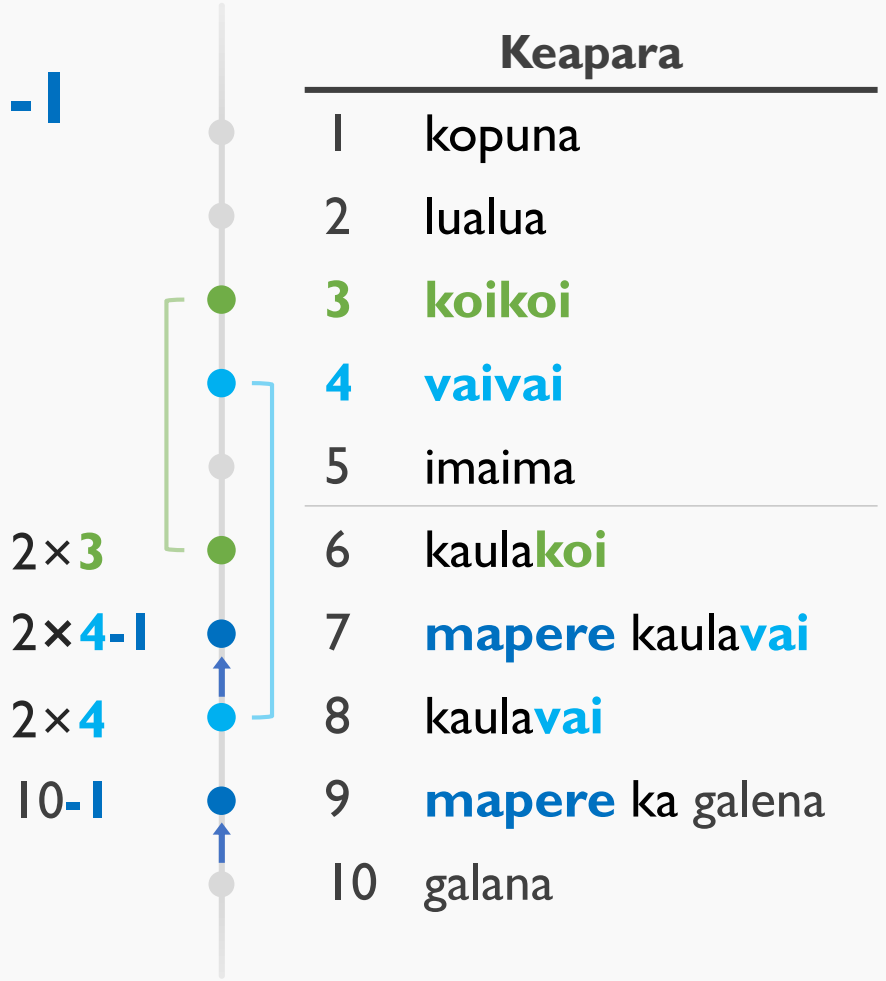
	Roro	
1	hamomo	
2	rua	
3	aihau	
4	bani	
5	ima	
6	abaihau	2×3
7	abaihau hamomo	
8	ababani	2×4
9	ababani hamomo	
10	harau haea	

	Keapara	
1	kopuna	
2	lualua	
3	koikoi	
4	vaivai	
5	imaima	
6	kaulakoi	2×3
7	mapere kaulavai	
8	kaulavai	2×4
9	mapere ka galena	
10	galana	

Multiplication II



Roro	
1	**hamomo**
2	rua
3	**aihau**
4	**bani**
5	ima
6	ab**aihau**
7	ab**aihau hamomo**
8	ab**abani**
9	ab**abani hamomo**
10	harau haea



Keapara	
1	kopuna
2	lualua
3	**koikoi**
4	**vaivai**
5	imaima
6	kaula**koi**
7	**mapere** kaula**vai**
8	kaula**vai**
9	**mapere** ka galena
10	galana

Hierarchy Hypothesis

Hierarchy Hypothesis I

Subtraction

10-4	10-3, 10-2	10-1	2 languages
	10-3, 10-2	10-1	21 languages
		10-1	10 languages

Hierarchy Hypothesis II

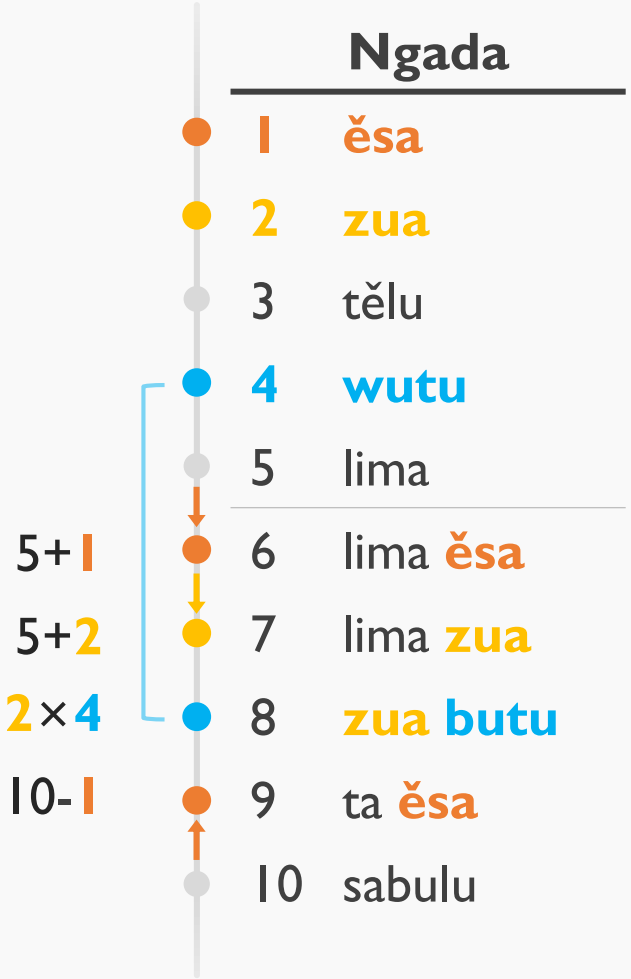
Multiplication

2×3	2×4	9 languages
	2×4	14 languages

Combination of Addition, Subtraction and Multiplication

	Ngada
1	ěsa
2	zua
3	tělu
4	wutu
5	lima
6	lima ěsa
7	lima zua
8	zua butu
9	ta ěsa
10	sabulu

Combination of Addition, Subtraction and Multiplication



Ngada demonstrates that multiple strategies can coexist within a single language

Hierarchy Hypothesis I

Subtraction

10-4	10-3, 10-2	10-1
	10-3, 10-2	10-1
		10-1

Hierarchy Hypothesis II

Multiplication

2×3	2×4
	2×4

05
Conclusions

Conclusions

Typological Diversity

Despite sharing a common ancestor, the 900 languages show **substantial variation** in numeral systems, reflecting the diverse ways humans conceptualize number.

Beyond Base Systems

While decimal systems predominate, Austronesian languages also use **addition**, **multiplication**, and **subtraction** to express numerals 4–9 — but no division. Numeral formation is not solely determined by base.

Implicational Hierarchies

Two implicational generalizations hold: if a language has $10-4=6$, it also has $10-3=7$, $10-2=8$, $10-1=9$; and if it has $3\times2=6$, it also has $4\times2=8$.

THANKS

LU, NAI-WEI
weiweiaakid@gmail.com

References

- Greenberg, Joseph H. 1978. Generalizations about numeral systems. In Joseph H. Greenberg et al. (eds.), Universals of Human Language, vol. 3: Word Structure, 249–295. Stanford: Stanford University Press.
- Li, Paul Jen-kuei. 2024. Numeral systems in Formosan languages. In Paul Jen-kuei Li, Elizabeth Zeitoun & Rik De Busser (eds.), Handbook of Formosan Languages: The Indigenous Languages of Taiwan (Brill's Handbooks in Linguistics 5), 343–370. Leiden: Brill. https://brill.com/edcollbook/title/63523
- Li, Paul Jen-kuei. 2006. Numerals in Formosan languages. *Oceanic Linguistics* 45(1): 133–152.
- Rosenfelder, Mark. n.d. Numbers from 1 to 10 in over 5000 languages [The Numbers List]. https://www.zompist.com/numbers.shtml