

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
- 02** Methodology
- 03** Base systems
- 04** Special expression strategies
- 05** Conclusions



01

Introduction



Austronesian

- 1200+ languages

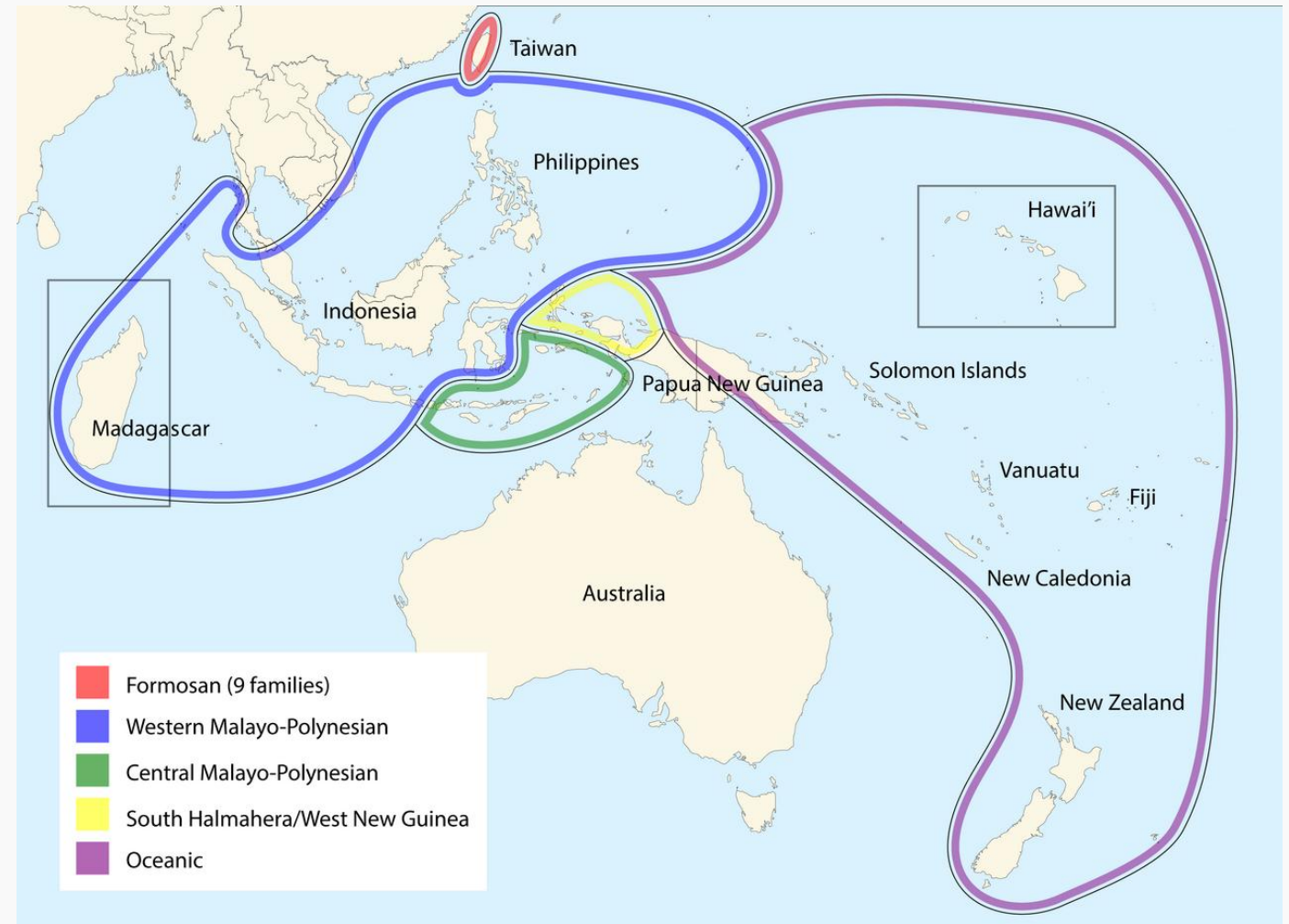
Proto-Austronesian

- | | |
|----|-----------|
| 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>

Introduction



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

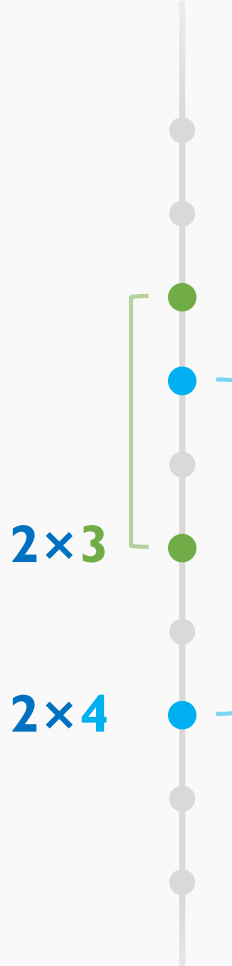
- Li (2024): Formosan numerals — predominantly decimal, quinary also attested

Motivating Observation

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

Research Gap

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



	Seediq
1	kiŋal
2	daha
3	teru
4	sepac
5	rima
6	mumuteru
7	mupitu
8	mumusepac
9	muŋari
10	maxal

| 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 Sedeq (Taroko)	qutux kiqal	saziq daha	ciwal teru	payat sepac	mayal rima	tizyu? mumuteru	pitu? mupitu	spat mumusepac	qeru? muqari	mpuw maxal
Tsouic										
Rukai Taloma	əa	Dosa	toLó	soʔátə	Limá	ənámə	ʔitó	vaLó	baǵátə	poLóko
<i>Bud</i>	iθa	Dósa	tóLo	səpátə	Líma	ənámə	pító	vaLo	bǵátə	póLoko
<i>Maga</i>	Duka	Dúsa	túru	pátə	ríma	nimi	pítu	váru	vǵátí	prúku
<i>Tona</i>	nága	dóθa	tóo	pátə	imá	námə	pító	vaó	vǵátə	pó:ko
<i>Mantauran</i>	náka	nosa	toLo	pátə	Lima	námə	pító	vaLo	vǵátə	poLoko
Tsou	cóni	yúso	túyu	sáptə	eímo	nómə	pítu	vóyu	sío	máskə
Kanakanabu	caáni	cuúsa	tuúlu	suúpata	líima	n'ámə	piítu	áalu	síia	máana
Saaroa	caani	suua	tuulu	paatə	kulima	kənámə	kupitu	kualu	kusia	kumaatə
Paiwanic										
Bunun	tasʔaʔ	dusaʔ	tau	paat	himaʔ	nuum	pituʔ	vauʔ	sivaʔ	macʔan
Paiwan	ita	ùsa	cal. 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	rosaʔ	tooʔ	šəpat	asəb	sayboši:	sayboši o ʔähäʔ	kəspat	äähʔhäʔ	lagpəz
Ami	cicay	tusaʔ	tuluʔ	spat	limaʔ	ʔnəm	pituʔ	faLuʔ	siwaʔ	guʔtəp
Sinicized										
Kavalan	ʔissaʔ	zusaʔ	tuLuʔ	spát	Limaʔ	ʔənəm	pituʔ	waLuʔ	siwaʔ	staRay
Pazeh	ʔidaʔ	dusaʔ	turuʔ	supat	xasap	xasəbuzaʔ	xasəbiltusaʔ	xasəbilturuʔ	xasəbitsupat	ʔisit
Thao	ta:haʔ	tu:saʔ	to:roʔ	špa:t	ri:maʔ	kato:roʔ	pi:tuʔ	kəspat	tana:θuʔ	ma:kθin
+Ketangalan	təsa	Lusa	təsu:	špat	tsima	anum	pitu	watsu	siwa	Labatan
+Taokas	ta:nu	rua	tuLu	lepat	hasap	tahap	jweto	mahəlpəp	tanəsu	taisid
+Papora	tanu	nia	tuLu	nəpat	nema	nenim	pitu	məhal	mesia	metsi
+Babuza	patə	ra	to:ro	nəpat	pa:ab	pa:an	pa:to	ma:pa:at	ta:pa:pa	tsə:it

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

A series of five circles at the top right of the slide. The third circle from the left is solid dark gray, while the others are white with dark gray outlines.

03

Base systems

A small dark gray triangle pointing upwards, centered on the bottom horizontal line.

Base systems

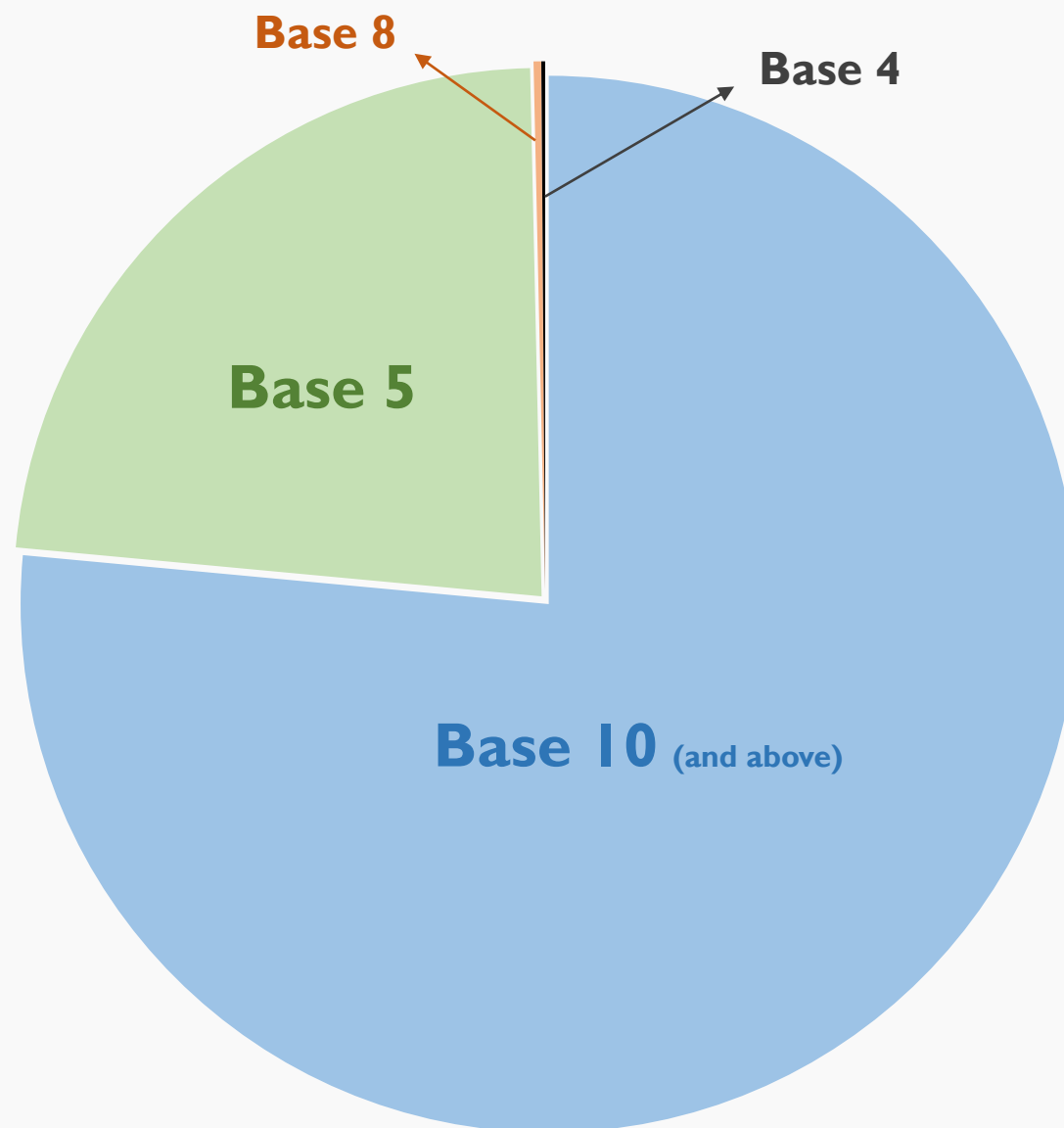
in 900 Austronesian Languages

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	Λu	tohats
5	[4] be kubua	nolim	mƏniΛm	tolima
6	[4] be areg di ru	elmoke	mΛndosim	tönomo
7	[4] be areg di tuol	elmokip	tiΛmgΛtrɔnduk	tohit
8	kiki ru	elmoyok	rɔnduk	toal
9	[8] be kubua	elmopeng	rɔnduk mΛnitiΛm	tosie
10		esugul	rɔnduk mΛnirɔsim	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	Λu	tohats
5	[4] be kubua	nolim	mƏniΛm	tolima
6	[4] be areg di ru	elmoke	mΛndosim	tönomo
7	[4] be areg di tuol	elmokip	tiΛmgΛtrɔnduk	tohit
8	kiki ru	elmoyok	rɔnduk	toal
9	[8] be kubua	elmopeng	rɔnduk mΛnititiam	tosie
10		esugul	rɔnduk mΛnirɔsim	maloto

Others

Non-base numeral constructions

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]

Sirasira

1	tangkua
2	iruk
3	irikiruk ena tango mangan
4	[2] da [2]
5	[2] da [2] a mangan
6	[2] a mangan [2] a mangan
7	
8	
9	
10	

no data

Base systems

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	
↓	8	
↓	9	
↓	10	

no data



04

Special expression strategies



Special expression strategies

Special expression strategies

Overview

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

Addition

Special expression strategies

$$4=2+2$$

Base 5		Base 10
Gapapaiwa		Fagululu
1	sago	ta'aga
2	ruwa	luwa
3	aroba	toi
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

Special expression strategies

	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

Special expression strategies

		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 dja ha:hu	anoh	nong
10-3	●	7 djo doloh	ndro teloh	hitu
10-2	●	8 djo lue	ndro rueh	walu
10-1	●	9 djo ʔeri	ndro sih	ta siya
	●	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

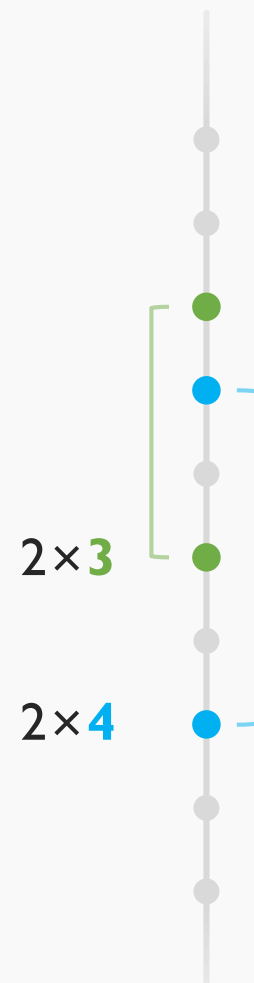
Greenberg, Joseph H (1978)

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.

		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	dja ha:hu	anoh	nong
10-3	7	djo doloh	ndro teloh	hitu
10-2	8	djo lue	ndro rueh	walu
10-1	9	djoʔ eri	ndro sih	ta siya
	10	rono	sunguah	hulu

Multiplication

Special expression strategies



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

Multiplication

Hierarchy Hypothesis II (original)

2x3	2x4	Thao
	2x4	Ujir

	Thao	Ujir
1	ta:ha?	set
2	tu:ša?	rúa
3	to:ro?	láti
4	špa:t	ká
5	ri:ma?	líma
6	ka ^{to:ro?}	dúau
7	pi:tu?	dubusam
8	ka ^{špat}	karúa
9	tana:θu?	τέρα
10	ma:kθin	uysía

Multiplication II

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

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

Multiplication II

I

Roro

1

2

3

4

5

6

7

8

9

10

hamomo

rua

aihau

bani

ima

abaihau

abaihau hamomo

ababani

ababani hamomo

harau haea

2×3

$2 \times 3 + 1$

2×4

$2 \times 4 + 1$

-I

Keapara

1

2

3

4

5

6

7

8

9

10

kopuna

lualua

koikoi

vaivai

imaima

kaulakoi

mapere kaulavai

kaulavai

mapere ka galena

galana

2×3

$2 \times 4 - 1$

2×4

$10 - 1$

Hierarchy Hypothesis

Special expression strategies

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

Special expression strategies

	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

Special expression strategies

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

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



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 $3 \times 2=6$, it also has $4 \times 2=8$.



— 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>