

TYOLOGY OF NUMERAL SYSTEMS

Bernard Comrie
Max Planck Institute for Evolutionary Anthropology and
University of California Santa Barbara
<comrie@eva.mpg.de>

A. Restricted systems, with little or no internal structure

- | | | | | |
|----|---|------------------------------|-------------------|--------------------------|
| 1. | No numerals | | <i>Pirahã</i> | |
| 2. | 1 | (ŋa)wumbawa | <i>Mangarayi</i> | |
| | 2 | ŋabaranwa | | |
| | 3 | ŋabaɭawa | | |
| 3. | 1 | guman | <i>Yidiny</i> | |
| | 2 | jambul | | |
| | 3 | dagul | | |
| | 4 | yunggan.gunyjii or mugungabi | | |
| | 5 | mala | | |
| 4. | 1 | towenyxa | <i>Hixkaryana</i> | |
| | 2 | asako | | |
| | 3 | osorwawo | | |
| | 4 | towtinke | | ‘its brother twice over’ |
| | 5 | kamori irakay (o) me | | ‘half of our hands’ |
| | 10 | kamothiri tkatxehkaxe ro | | ‘our hands completely’ |
| 5. | This seems to be one of the few areas in linguistics where present-day languages provide direct insight into the evolution of language. Speakers of languages with restricted systems, such as Australian languages, typically did not engage traditionally in counting. The number of entities was arrived at by “subitizing”, i.e. immediately recognizing the number, as is possible up to around 5. | | | |

The Hixkaryana system suggests that the development of higher counts may have involved gaps, with higher round numbers developing before some lower numbers (unless, of course, Hixkaryana has lost the original numerals 6–10).

Names of numbers, especially beyond the first couple, often betray an origin in body-part representation, as with the Hixkaryana terms for (4,)5, and 10.

B. Simple systems with addition only

- | | | | | |
|----|---|---------|-------|---------------|
| 6. | 1 | paŋ | | <i>Haruai</i> |
| | 2 | mos | | |
| | 3 | mos paŋ | 2 + 1 | |
| | 4 | mos mos | 2 + 2 | |

C. More complex systems using multiplication and addition applied to a base

7. General pattern:
For base b : $(n \times b) + m$ (where $m < b$)

Decimal (base 10)

8. wǔ-shí sì *Chinese*
five-ten four
 $54 [(5 \times 10) + 4]$

Vigesimal (base 20)

9. kəlgən-qlekken məngətkən ŋireq parol *Chukchi*
fifteen-twenty ten two left
 $312 [(15 \times 20) + (10 + 2)]$

Base 60

10. èna ma gàati dàimita mutò *Ekari*
one and ten and sixty
 $71 [60 + (10 + 1)]$

11. muto wii
sixty four
 $240 [4 \times 60]$

Base 32

12. ìfò wǎdhì *Ngiti*
four thirtytwo
 $128 [4 \times 32]$

Base 12

13. ba-kuru ba-ba ná |-ā| |-bā| *Biom*
PL-twelve PL-two plus two
 $26 [(2 \times 12) + 2]$

Base 8

14. kanuje? tehiuŋ rnu? *Northern Pame*
three eight three
 $27 [(3 \times 8) + 3]$

Base 6

15. swabra ptae ynaoaemy ntamnao *Kanum*
five thirtysix two three.six
 $200 [(5 \times 6^2) + (3 \times 6) + 2]$

New Guinea Highland body-part counting systems (bases 18-74)

16.	1	little finger	23	24	46	47	<i>Kobon</i>
	2	ring finger	22	25	45	48	
	3	middle finger	21	26	44	49	
	4	forefinger (index finger)	20	27	43	50	
	5	thumb	19	28	42	51	
	6	wrist	18	29	41	52	
	7	forearm	17	30	40	53	
	8	inside of elbow	16	31	39	54	
	9	biceps	15	32	38	55	
	10	shoulder	14	33	37	56	
	11	collarbone	13	34	36	57	
	12	hole above breastbone		35		58	

Kobon forms are, in order: wañig nöbö, igwo, igwo aŋ nöbö, igwo milö, mamid, kagoŋ, mudun, raleb, ajip, siduŋ, agip, mögan

17. Hypothesis: Arithmetic bases of numeral systems have either a somatic or a commercial (transactional) origin; lower bases are typically somatic, higher bases commercial, but New Guinea Highland body-part counting systems have relatively high somatic-origin bases.

- 10 fingers
- 20 fingers and toes; each finger twice (two phalanges/knuckles)
- 8 spaces between fingers (attested for some California languages)
- 12 phalanges or knuckles of fingers (excluding thumbs)

For higher bases with a commercial origin, cf. English score '20', which in some varieties has made it into the numeral system.

18. Distribution of different bases across a sample of languages of the world

[See: <http://wals.info/feature/131>.]

D. Idiosyncrasies relating to bases

Portmanteau forms

- | | | |
|-----|--|----------------|
| 19. | sorok
forty
40 [expected 4 x 10] | <i>Russian</i> |
| 20. | eleven
11 [expected 10 + 1] | <i>English</i> |

21. *Balinese (cp. also Javanese, Madurese)*
- | | | |
|-----|---------------|--|
| 25 | se-lae | ‘one thread (of Chinese coins)’ |
| 45 | se-timan | ‘one opium packet (costing 45 Chinese coins)’ |
| 50 | se-ket | ‘one tie (i.e. two threads of 25 Chinese coins)’ |
| 75 | telung benang | ‘three threads (of Chinese coins)’ |
| 200 | s-atak | ‘one bundle of 200 Chinese coins’ |
| 400 | s-aman | ‘one gold (coin worth 400 Chinese coins)’ |
| 900 | sanga | [etymology unclear] |

22. Compare less spectacular irregularities

fif-teen (*five-teen)
 five-ten
 15

English

23. twenty; twelve

24. In Hindi, arguably all the numerals 1–100 are irregular

Hindi

	0	1	2	3	4	5	6	7	8	9
-		ek	do	tīn	cār	pāñc	chah	sāt	āṭh	nau
10	das	gyārah	bārah	terah	caudah	pandrah	solah	satrah	aṭhārah	unnīs
20	bīs	ikkīs	bāīs	teīs	caubīs	paccīs	chabbīs	sattāīs	aṭṭāīs	untīs
30	tīs	ikattīs	battīs	taiṁtīs	cauṁtīs	paiṁtīs	chattīs	saiṁtīs	aṛtīs	untālīs
40	cālīs	iktālīs	bayālīs	taiṁtālīs	cavālīs	paiṁtālīs	chiyālīs	saiṁtālīs	aṛtālīs	uncās
50	pacās	ikyāvan	bāvan	tirpan	cauvan	pacpan	chappan	sattāvan	aṭṭhāvan	unsaṭh
60	sāṭh	iksāṭh	bāsāṭh	tirsāṭh	cauṁsāṭh	paiṁsāṭh	chiyāsāṭh	sarsāṭh	aṛsāṭh	unhattar
70	sattar	ikhattar	bahattar	tihattar	cauhattar	pachattar	chihattar	sathattar	aṭṭhattar	unyāsī
80	assī	ikyāsī	bayāsī	tirāsī	caurāsī	pacāsī	chiyāsī	sattāsī	aṭṭhāsī	navāsī
90	nave	ikyānve	bānve	tirānve	caurānve	pacānve	chiyānve	sattānve	aṭṭhānve	ninyānve

Isolated “bases”

25. quatre-vingt-douze
 four-twenty-twelve
 92 [(4 x 20) + 12]

French

26. deu-naw
 two-nine
 18 [2 x 9]

Welsh

Overrunning

27. disaṭ-nocti
 10-teen
 20 [10 + 10]

Polabian

21 is jadān disaṭ-nocti, i.e. [1 + 20], not [11-teen]

28. soixante-dix
 sixty-ten
 70 [60 + 10]

French

29. soixante-douze
sixty-twelve
72 [60 + 12]

30. soixante-dix-sept
sixty-ten-seven
77

E. Exponentiation and other higher bases

31. 10^1 10^2 10^3 10^6 *English*
ten hundred thousand million

Absence of exponentiation

32. qliq-qlikkin *Chukchi*
twenty-twenty
400 (20 x 20) – highest numeral in traditional system

Effectively monomorphemic series of bases

33. 10 dáśa- *Sanskrit*
100 śatá-
1,000 sahásra-
10,000 ayúta-
100,000 lakṣá-
1,000,000 prayúta-
 10^7 kóṭi-
 10^8 arbudá-
 10^9 mahārbuda-
 10^{10} kharvá-
 10^{11} nikharva-

Note current Indian English: lakh 10^5 ; crore 10^7

34. shí 十 10^1 *Chinese*
bǎi 百 10^2
qiān 千 10^3
wàn 万 萬 10^4
yì 亿 億 10^8
zhào 兆 10^{12}
jīng 京 10^{16} etc.

This system is used in Chinese, Japanese, Korean.

35. kal 20 *Yucatec*
bak 400 (20^2)
pic 8000 (20^3)
calab 160,000 (20^4)

kinchil	3,200,000 (20 ⁵)
alau	64,000,000 (20 ⁶)

(Semi-)productive systems

36.			long scale	short scale	<i>English</i>
	million	first	10 ⁶	10 ⁶	
	billion	second	10 ¹²	10 ⁹	
	trillion	third	10 ¹⁸	10 ¹²	
	quadrillion	fourth	10 ²⁴	10 ¹⁵	
	[general pattern]	nth	10 ⁶ⁿ	10 ³⁽ⁿ⁺¹⁾	

See <http://www.isthe.com/chongo/tech/math/number/howhigh.html> for a proposal on how to count indefinitely high using (pseudo-)Latin prefixes.

37. Contrast innovative / computer

kilo-	10 ³	<i>English / International</i>
mega-	10 ⁶	
giga-	10 ⁹	
tera-	10 ¹²	
peta-	10 ¹⁵	
exa-	10 ¹⁸	
zetta-	10 ²¹	
yotta-	10 ²⁴	

Sequence of bases that are not (all) powers of a single base

38.	cxr-as	otxm-oc-da-cxra-met'i	<i>Georgian</i>
	nine-hundred	four-twenty-and-nine-teen	
	999 [(9 x 100) + (4 x 20) + (10 + 9)]		
	Bases: (10,) 20, 100		

39.	kàmpwòd	ɲkwuu	sicyɛɛré	'ná	béé-tàànre	ná	ké	'ná	báár-ìcyèèrè	<i>Supyire</i>
	fourhundred	eighty	four	and	twenty-three	and	ten	and	five-four	
	799 [i.e. 400 + (4 x 80) + (3 x 20) + {10 + (5 + 4)}]									
	Bases: (5, 10,) 20, 80, 400									

Alternating bases

40.	10	désat	10	<i>Resian Slovene, etc. (e.g. some Albanian)</i>
	20	dwísti	2 x 10	
	30	trásti	3 x 10	
	40	dwákrat dwísti	2 x 20	
	50	patardú	5 x 10	
	60	tríkrat dwísti	3 x 20	

F. Other arithmetic processes

Subtraction

41. un-de-viginti *Latin*
one-from-twenty
19 [20 – 1]

Division (actually: multiplication by fraction)

42. hanner cant *Welsh*
half hundred
50 [$\frac{1}{2} \times 100$]

Subtraction and addition

43. dɔŋas' bən's'aŋ ʔkiʔ *Ket*
thirty without hundred
70 [100 – 30]

44. qus'am ʌɣam dɔŋas' bən's'aŋ ʔkiʔ
one left.over thirty without hundred
71 [(100 – 30) + 1; NB: not 100 – (30 + 1)]

45. Successive approximation, cf. time expressions in some languages
drie (uur) *Dutch*
three hour
03:00

46. half drie
half three
02:30

47. vijf over half drie
five after half three
02:35

Overcounting

48. halv-tred-sinds-tyve *Danish*
half-third-times-twenty
50 [half of the third twenty]

Now usually: halvtreds

49. paüne tini šata *Oriya*
three.quarters three hundred
275 [three quarters of the third hundred]

G. Ordering of constituents

From larger to smaller

50. sān-bǎi wǔ-shí sì Chinese
three-hundred five-ten four
354 [i.e. 300 + 50 + 4]

From smaller to larger

51. efatra amby dima-mpolo sy telo-njato Malagasy (Standard)
four plus five-ten and three-hundred
354 [i.e. 4 + 50 + 300]

From smaller to larger for smaller combinations, from larger to smaller for larger combinations

52. drei-hundert-vier-und-fünf-zig German
three-hundred-four-and-five-ten
354 [i.e. 300 + 4 + 50]

53. zwei-hundert-sechs-und-fünf-zig-tausend-drei-hundert-vier-und-sieb-zig
two-hundred-six-and-five-ten-thousand-three-hundred-four-and-seven-ty
256 374 [i.e. (200 + 6 + 50) x 1000 + (300 + 4 + 70)]

54. order smaller-larger (contrasting with higher larger-smaller)
Spanish through 15 (quin-ce, cf. 16 diec-i-séis) Spanish
Italian through 16 (se-dici, cf. 17 dici-as-sette) Italian
English through 19 (nine-teen) English
German through 99 (neun-und-neun-zig) German

From larger to smaller for smaller combinations, from smaller to larger for larger combinations

55. limam-polo roe amby, amby telon-jato Malagasy (Nosy Be)
five-ten two plus plus three-hundred
'352 [i.e. 50 + 2 + 300]'

56. Hypothesis: The order from larger to smaller is preferred because it gives earlier recognition of the approximate quantity involved, i.e. in 354 the 300 is more significant than the 4. Local inversion of lower positions (e.g. ten and units) is minimally disruptive. So we expect prevalence of the order from larger to smaller, with possible local inversion of the lower positions.

H. Ambiguity

Parsing ambiguities

57. a million and a half (apples) English
(i) 1½ million, i.e. 1,500,000
(ii) 1,000,000 ½

58. un fil a thri ugain o asynod *Biblical Welsh*
 one thousand and three twenty of asses
 (i) 1060 asses
 (ii) 61,000 asses
59. deuddeg a thri ugain mil o eidionau
 twelve and three twenty thousand of cattle
 72,000 cattle
60. saith mil a phedwar ugain mil
 seven thousand and four twenty thousand
 87,000

Abbreviation

61. pèet-phan hòk *Thai*
 eight-thousand six
 (i) 8006
 (ii) 8600
 (iii) *8060
62. m~~u~~an cèt
 ten.thousand seven
 (i) 10,007
 (ii) 17,000
 (iii) *10,700
 (iv) *10,070
63. pèet-phan hòk-róy
 eight-thousand six-hundred
 8600
64. (a) s~~ā~~n-b~~ǎ~~i liù-shí *Chinese*
 three-hundred six-ten
 (b) s~~ā~~n-b~~ǎ~~i liù
 three-hundred six
 360
65. s~~ā~~n-b~~ǎ~~i líng liù
 three-hundred zero six
 306

Obligatory use of *líng*, creating a partial place-value system, avoids ambiguity.

66. s~~ā~~n-qiān líng liù
 three-thousand zero six
 3006 (older: s~~ā~~n-qiān líng líng liù)

67. sān-qiān líng liù-shí
 three-thousand zero six-ten
 3060

Diachronic merger

68. thirty, thirteen *English*
69. 7 sāt *Northern Mansi*
 100 sāt > janig sāt (janig 'big')

Semantic change

70. billion *British English*
 (i) older, "long scale" 10^{12}
 (ii) newer, "short scale" = US English 10^9
 (long scale: 10^{6n} ; short scale: $10^{3(n+1)}$)

Specialized use

71. bak *Mayan*
 usually 400
 but 360 days (long calendar)
72. kilo- *English / International*
 1000
 but: kilobyte '1024 bytes'

Body part systems

73. siduŋ 'shoulder' = 10, 14, 33, 37, 56, 60 *Kobon*

I. Internal structure and psychological reality

74. Issues at interface with psychomathematics (psychoarithmetic)
- a) To what extent are problems caused by having a linguistic representation that does not correspond to the arithmetic notation, e.g., assuming a decimal notation in descending order
 (i) if a formation is non-decimal, e.g. French quatre-vingt-onze, literally 'four twenties and eleven', for 91?
 (ii) if the order is (partially) inverted, e.g. German sechsfundfünfzig, literally 'six and fifty', for 56?
- b) When people do arithmetic, do they operate with linguistic representations of numerals (in which case problem (a) might arise) or do they operate with abstract quantities (in which case it should not)?
75. Writing 'six and fifty' for 56 Dutch zesenvijftig
 German sechsfundfünfzig

Dutch speakers typically write the 6 and then go back to “fill in” the 5.
 German speakers typically write the 5 then the 6, and are explicitly instructed to do this at school.

76. But note that in some instances the overall structure of the numeral system means that the formal structure of individual numerals can be overridden.

77.	Cardinal	Ordinal	Danish
10	ti	tiende	
20	tyve	tyvende	
30	tred(i)ve	tred(i)vtte	
40	fyrre	fyrretyvende	
50	halvtreds	halvtredsindstyvende	cf. tredje 3rd
60	tres	tresindstyvende	cf. tre 3
70	halvfjerds	halvfjerdsindstyvende	cf. fjerde 4th
80	firs	firsindstyvende	cf. fire 4
90	halvfems	halvfemsindstyvende	cf. femte 5th
100	(et) hundred(e)	hundrede	

Some Basic References

- Barriga Puente, Francisco. 1998. *Los sistemas de numeración indoamericanos: un enfoque areotipológico*. México: Universidad Nacional Autónoma de México.
- Comrie, Bernard. 2005a. 'Endangered numeral systems'. In Jan Wohlgemuth and Tyko Dirksmeyer (eds.): *Bedrohte Vielfalt: Aspekte des Sprach(en)tods [Endangered diversity: Aspects of language death]*, 203–230. Berlin: Weißensee Verlag.
- Comrie, Bernard. 2005b. 'Numeral bases'. In Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie (eds.), *The world atlas of language structures*, 530–533. Oxford: Oxford University Press. (Electronic version: <http://wals.info/feature/131>.)
- Everett, Daniel L. 2005. 'Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language'. *Current Anthropology* 46: 621–646.
- Greenberg, Joseph H. 1978. 'Generalizations about numeral systems'. In Joseph H. Greenberg, Charles A. Ferguson and Edith A. Moravcsik (eds.), *Universals of human language, Volume 3: Word structure*, 249–295. Stanford: Stanford University Press. Reprinted in Keith Denning and Suzanne Kemmer (eds.), *On language: Selected writings of Joseph H. Greenberg*, 271–309. Stanford, 1990: Stanford University Press.
- Hanke, Thomas. 2005. *Bildungsweisen von Numeralia: eine typologische Untersuchung*. Berlin: Weißensee Verlag.
- Hurford, James R. 1975. *The linguistic theory of numerals*. Cambridge: Cambridge University Press.
- Hurford, James R. 1987. *Language and number: The emergence of a cognitive system*. Oxford: Blackwell.
- Meeuwissen, Marjolein. 2004. *Producing complex spoken numerals for time and space*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Owens, Kay. 2001. 'The work of Glendon Lean on the counting systems of Papua New Guinea and Oceania'. *Mathematics Education Research Journal* 13: 47–71.
- Yasugi, Yoshiho. 1995. *Native Middle American languages: An areal-typological perspective*. Osaka: National Museum of Ethnology.

Index of Languages

Balinese	BAN	W Mal.-Pol., Austronesian	Bali, Indonesia
Biom	BOM	Benue-Congo, Niger-Congo	Plateau State, Nigeria
Chinese	CMN	Sinitic, Sino-Tibetan	China
Chukchi	CKT	Chukotko-Kamchatkan	Chukotka, Russia
Danish	DAN	Germanic, Indo-European	Denmark
Dutch	NLD	Germanic, Indo-European	Netherlands; etc.
Ekari	EKG	Wissel Lakes-Kemandoga, Trans-New Guinea	Papua, Indonesia
English	ENG	Germanic, Indo-European	England; USA; etc.
French	FRA	Italic, Indo-European	France; etc.
Georgian	KAT	Kartvelian	Rep. of Georgia
German	DEU	Germanic, Indo-European	Germany; etc.
Haruai	TMD	Piawi	Madang Pr., Papua New Guinea
Hindi	HIN	Indo-Aryan, Indo-European	North-Central India
Hixkaryana	HIX	Cariban	Amazonas, Brazil
Italian	ITA	Italic, Indo-European	Italy
Japanese	JPN	Japanese-Ryukyuan	Japan
Kanum	(KCD)	Morehead and Upper Maro Rivers	Papua, Indonesia
Ket	KET	Yeniseian	W. Siberia, Russia
Kobon	KPW	Kalamitic, Trans-New Guinea	Madang Pr., Papua New Guinea
Latin	LAT	Italic, Indo-European	Rome [extinct]
Malagasy, Nosy Be	(SKG)	W Mal.-Pol., Austronesian	NW Madagascar
Malagasy, Stand.	PLT	W Mal.-Pol., Austronesian	Madagascar
Mangarayi	MPC	Mangarayi, Australian [language family]	Northern Terr., Australia
Mayan			Mesoamerica
Ngiti	NIY	Central Sudanic, Nilo- Saharan	Oriental Pr., DR Congo
Northern Mansi	(MNS)	Ugric, Uralic	W. Siberia, Russia
Northern Pame	PMQ	Pamean, Otomanguean	San Luis Potosí, Mexico
Oriya	ORI	Indo-Aryan, Indo-European	Orissa, India
Pirahã	MYP	Mura	Amazonas, Brazil
Polabian	POX	Slavic, Indo-European	Germany [extinct]
Resian Slovene	(SLV)	Slavic, Indo-European	Italy
Russian	RUS	Slavic, Indo-European	Russia
Sanskrit	SAN	Indo-Aryan, Indo-European	India [extinct]
Spanish	SPA	Italic, Indo-European	Spain; Latin America
Supyire	SPP	Gur, Niger-Congo	Mali
Thai	THA	Tai-Kadai	Thailand
Welsh	CYM	Celtic, Indo-European	Wales, United Kingdom
Yidiny	YII	Pama-Nyungan, Australian	Queensland, Australia
Yucatec	YUA	Mayan	Yucatán, Mexico