A. Restricted systems, with little or no internal structure

(1) No numerals

[Pirahã; Everett 2005: 623-627]

(2) 1 (ŋa)wumbawa
    2 ŋabaranwa
    3 ŋabaḷawa

[Mangarrayi; Merlan 1982: 92]

(3) 1 guman
    2 jambul
    3 dagul
    4 yunggan.gunyjii or mugungabi
    5 mala ‘face of hand’

[Yidiny; Dixon 1991: 224]

Note: The forms for 4 and above may not be fully conventionalized.

(4) 1 pūg (ma)
    2 xep xep
    3 eba-pūg 2 + 1
    4 ebadipdip 2 + 1 + 1
    5 pūg pōgbī ‘one hand’

[Munduruku; Pica et al. 2004; see also section B]

(5) 1 towenyxa
    2 asako
    3 osorwawo
    4 towtinke rye ‘its brother twice over’
    5 kamori irakay (o) me ‘half of our hands’
10 kamothiri tkatxehkaxe ro ‘our hands completely’

[Hixkaryána; Derbyshire 1979: 154-155]

Note: The forms for 4 and above are “non-basic” and “are also used without precision”.

(6) **subitize**
Subitizing is the rapid, accurate, and confident judgment of numbers performed for small numbers of items with a cut-off point around 4

B. Simple systems with addition only

(7) 1 paŋ
2 mös
3 mös paŋ 2 + 1
4 mös mös 2 + 2

[Haruai; Comrie: Own fieldwork]

(8) 1 añi (nôbô)
2 mõhöp
3 mõhau nõganj
4 mõhau mõhau

[Kobon; Davies 1981: 208]

(9) 1 bits
2 iru?
3 iru? da bits 2 + 1
4 iru? da iru? 2 + 2
5 iru? da iru? da bits 2 + 2 + 1
6 iru? da iru? da iru? 2 + 2 + 2
etc.
10 dzi bangi marafain da dzi bangi marafain sib ‘my hand half and my hand half completed’


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

(10) General pattern:
For base b: \((n \times b) + m\)
where \(m < b\)
and usually \(n < b\), or \(n \leq b\)
Decimal (base 10)

(11)   DBHelper  si
      five-ten   four
     54 [50 + 4]

[Mandarin Chinese]

Vigesimal (base 20)

(12)  kalgon-glekken mangatken njireq parol
      fifteen-twenty ten two left
     312 [(15 x 20) + (10 + 2)]

[Chukchi; Skorik 1961: 390]

Base 60

(13)  éna ma gati dåimita mutò
      one and ten and sixty
     71 [60 + (10 + 1)]

(14)  muto wii
      sixty four
     240 [4 x 60]

[Ekari; Drabbe 1952: 30]

Base 32

(15)  ifo wadhì
      four thirtytwo
     128 [4 x 32]


Base 12

(16)  ba-kuru ba-ba nā CL-ā CL-bā
      PL-twelve PL-two plus CL-this CL-two
     26 [(2 x 12) + 2]

[Berom; Bouquiaux 1970: 259]
Note: The last two elements -ā and -bā require a class prefix corresponding to the class of the head noun.

Base 8

(17) karnuʔ  tenhiuŋ  rnuʔ
three  eight  three
27 [(3 x 8) + 3]

[Northern Pame; Avelino 2006: 45]

Base 6

(18)  1  näbi
2  eda/yda
3  etha/ytho
4  asar
5  tabuthui
6  nibo
36  fta  \(6^2\)
216  taruba  \(6^3\)
1296  damno  \(6^4\)
7776  wărämâkâ  \(6^5\)
46656  wi  \(6^6\)
72  eda fta  \(2 \times 6^2\)
73  eda fta a näbi  \((2 \times 6^2) + 1\)
50  näbi fta a eda nibo a eda  \((1 \times 6^2) + (2 \times 6) + 2\)

[Komnzo; Döhling 218: 93-94; Yam-counting video: https://vimeo.com/54887315]

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

(19) wañig nöbö  little finger  1  23  24  46
igwo  ring finger  2  22  25  45
igwo aŋ nöbö  middle finger  3  21  26  44
igwo milō  index finger  4  20  27  43
mamid  thumb  5  19  28  42
kagoł  wrist  6  18  29  41
mudun  forearm  7  17  30  40
raleb  inside of elbow  8  16  31  39
ajip  biceps  9  15  32  38
siduŋ  shoulder  10  14  33  37
agip  collarbone  11  13  34  36
mögan  hole above breastbone  12  35
[Kobon; Davies 1981: 206-208]

Notes:
a) The second half of the body on each pass across the body is distinguished by böŋ 'one side of'.
b) The second pass across the body is distinguished by ŋin juöl adog da 'pulling out the hand, give back!', subsequent passes by modifying ŋin juöl with a numeral.
c) The system is "symmetrical", i.e. the body parts used on one side of the body (in Kobon, for 1–11) are used in reverse order for the other side of the body (Kobon 13–23).

Arithmetic properties (without any claims to psychological reality):

a) On the first pass across the body, any given number term n also denotes 24 – n, and similarly, mutatis mutandis, for subsequent passes across the body.
b) The system has a base 23.

| (20) | agilîn | little finger | 1 | 19 |
| agilîn rolyöbö | ring finger | 2 | 20 |
| wölö mil | middle finger | 3 | 21 |
| köño nigib | index finger | 4 | 22 |
| mömîd | thumb | 5 | 23 |
| urap cigib | wrist | 6 | 18 | 24 | 36 |
| mij | forearm | 7 | 17 | 25 | 35 |
| amîña| inside of elbow | 8 | 16 | 26 | 34 |
| mac | biceps | 9 | 15 | 27 | 33 |
| möib | shoulder | 10 | 14 | 28 | 32 |
| katlöi | collarbone | 11 | 13 | 29 | 31 |
| migan | hole above breastbone | 12 | 30 |

[Haruai; Comrie: Own fieldwork]

[Some numerals, like the corresponding body parts, have alternative forms; urap cigib and mac are properly bracelets worn at the given body part.]

Notes:
a) The second half of the body on each pass across the body is distinguished by adökwebö 'of that side'.
b) The second pass across the body is distinguished by höuölîb 'turning' or höbkalib 'returning'.
c) The end of the first pass across the body (i.e. 23) can be identified by saying padö kwo panyöbü dua 'housepost there one went', subsequent passes by replacing panyöbü 'one' with a higher numeral and using the plural verb form dumä.
d) The system is "asymmetrical", i.e. the body parts used on one side of the body (in Haruai, for 1–11) are not all used in reverse order for the other side of the body (cf. in particular the fingers).
<table>
<thead>
<tr>
<th>(20')</th>
<th>agiliŋ</th>
<th>little finger</th>
<th>1</th>
<th>19</th>
<th>37</th>
</tr>
</thead>
<tbody>
<tr>
<td>agiliŋ rolyöbô</td>
<td>ring finger</td>
<td>2</td>
<td>20</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>wölö mil</td>
<td>middle finger</td>
<td>3</td>
<td>21</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>kŏnö nigib</td>
<td>index finger</td>
<td>4</td>
<td>22</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>mömíd</td>
<td>thumb</td>
<td>5</td>
<td>23</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>urap cigib</td>
<td>wrist</td>
<td>6</td>
<td>18</td>
<td>24</td>
<td>36</td>
</tr>
<tr>
<td>mij</td>
<td>forearm</td>
<td>7</td>
<td>17</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>amiñaxib</td>
<td>inside of elbow</td>
<td>8</td>
<td>16</td>
<td>26</td>
<td>34</td>
</tr>
<tr>
<td>mac</td>
<td>biceps</td>
<td>9</td>
<td>15</td>
<td>27</td>
<td>33</td>
</tr>
<tr>
<td>möib</td>
<td>shoulder</td>
<td>10</td>
<td>14</td>
<td>28</td>
<td>32</td>
</tr>
<tr>
<td>katlöi</td>
<td>collarbone</td>
<td>11</td>
<td>13</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>migan</td>
<td>hole above breastbone</td>
<td>12</td>
<td>30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Arithmetic properties:

a) There is no consistent relation between the numerical values of the same body part on different sides of the body. For 6–18, the two add up to 24; for 1–5 and 19–23, the total varies from 20 to 28.

b) The base shifts from 23 on the first pass across the body to 18 on subsequent passes, i.e. passes end at 23, 41, 59, etc.


(21) 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.

(22) If the lowest multiplicative base is higher than about 12, and the system is not an extended body-part counting system, then the higher numbers below the lowest multiple base make use of an additive base, e.g. Georgian (lowest multiplicative base: 20); some languages use this even for a smaller lowest multiplicative base, e.g. Nauete (lowest multiplicative base: 10); see also Chukchi example (12) above.

(23) 1 ert-i -i nominative case (NOM)
2 or-i
5 xut-i
6 ekvi-s-i
9 cxra  NOM drops after a
10 at-i
11 t-ert-met’-i  Ø-one-teen-NOM
15 t-xut-met’-i
16 t-ekvi-s-met’-i
19 [t-]cxra-met’-i (t + c > c)
20 oc-i
21 oc-da-ert-i  twenty-and-one-NOM
30 oc-da-at-i  twenty-and-ten-NOM
31 oc-da-t-ert-met’-i  twenty-and-Ø-three-teen-NOM
40 oc-m-ooc-i  two-Ø-twenty-NOM


(24) 2 kai-rua  CL-two
3 kai-tolu  CL-three
5 kai-lima  CL-five
7 kai-lima resi kai-rua  CL-five and CL-two
8 kai-lima resi kai-tolu  CL-five and CL-three
10 welit-see  ten-one
12 welit-see resi kai-rua  ten-one and CL-two
17 welit-see resi kai-lima resi kai-rua  ten-one and CL-five and CL-two
20 welit-kai-rua  ten-CL-two

[Nauete; https://mpi-lingweb.shh.mpg.de/numeral/Naute.htm, with data provided by Geoffrey Hull in 1996; see also Schapper & Hammarström 2013: 429, and https://abvd.shh.mpg.de/austroasiatic/language.php?id=1365, the latter with data provided by Alexandre Veloso in 2016. There are differences of detail among the sources, not affecting the main point.]

Note: Nauete numerals 2-5 require a numeral classifier prefix, with kai- being the default used in counting.

(25) Distribution of different bases across a sample of languages of the world

[See: Comrie (2013)]

D. Idiosyncrasies relating to bases

Portmanteau forms

(26) sorok forty
40 [expected 4 x 10]

[Russian]

(27) eleven
11 [expected 10 + 1]

[English]

(28) 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]

[Balinese; Eiseman 1990: 162-168]

(29) Compare less spectacular irregularities

fif-teen (*five-teen)
five-ten
15
twenty; twelve

[English]

(30) In Hindi, arguably all the numerals 1–100 are irregular

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>ek</td>
<td>do</td>
<td>tūn</td>
<td>tān</td>
<td>pānic</td>
<td>chah</td>
<td>sāt</td>
<td>āth</td>
<td>nau</td>
</tr>
<tr>
<td>10</td>
<td>das</td>
<td>gaṭrah</td>
<td>bārah</td>
<td>terah</td>
<td>caudah</td>
<td>pandrah</td>
<td>solah</td>
<td>satrah</td>
<td>athārah</td>
<td>umni</td>
</tr>
<tr>
<td>20</td>
<td>bis</td>
<td>ikkis</td>
<td>bāis</td>
<td>teis</td>
<td>caubis</td>
<td>paccis</td>
<td>chabhis</td>
<td>sattās</td>
<td>aṭṭās</td>
<td>untis</td>
</tr>
<tr>
<td>30</td>
<td>tīs</td>
<td>ikattis</td>
<td>battis</td>
<td>taintis</td>
<td>cauntis</td>
<td>paimtis</td>
<td>chattis</td>
<td>sauntis</td>
<td>artis</td>
<td>untlās</td>
</tr>
<tr>
<td>40</td>
<td>cālīs</td>
<td>ikṭālis</td>
<td>bayālis</td>
<td>taintālis</td>
<td>cavālis</td>
<td>paimtālis</td>
<td>chiyālis</td>
<td>saintālis</td>
<td>artlālis</td>
<td>uncās</td>
</tr>
<tr>
<td>50</td>
<td>pacās</td>
<td>ikyāvan</td>
<td>bāvan</td>
<td>tirpan</td>
<td>cauvan</td>
<td>pacpan</td>
<td>chappan</td>
<td>sattāvan</td>
<td>athāvan</td>
<td>unsath</td>
</tr>
<tr>
<td>60</td>
<td>sāth</td>
<td>iḳsath</td>
<td>bāsat</td>
<td>tirsath</td>
<td>caumtsath</td>
<td>paimstath</td>
<td>chiyāsath</td>
<td>sarsath</td>
<td>arsat</td>
<td>unhattar</td>
</tr>
<tr>
<td>70</td>
<td>sattar</td>
<td>ikhattar</td>
<td>bāhattar</td>
<td>tilhattar</td>
<td>cauhattar</td>
<td>pachattar</td>
<td>chihhattar</td>
<td>sathhattar</td>
<td>athhattar</td>
<td>unyāsī</td>
</tr>
<tr>
<td>80</td>
<td>assā</td>
<td>ikyāsī</td>
<td>bayāsī</td>
<td>tirāsī</td>
<td>cauṛāsī</td>
<td>pacāsī</td>
<td>ciyāsī</td>
<td>sattāsī</td>
<td>atthāsī</td>
<td>navāsī</td>
</tr>
<tr>
<td>90</td>
<td>nāve</td>
<td>ikyānve</td>
<td>bānve</td>
<td>tirānve</td>
<td>cauṛānve</td>
<td>pacānve</td>
<td>chiyānve</td>
<td>sattānve</td>
<td>atthānve</td>
<td>ninyānve</td>
</tr>
</tbody>
</table>

[Hindi; McGregor 1972: 61-62; note that some numerals have minor variants]

Isolated “bases”

(31) quatre-vingt-douze
four-twenty-twelve
92 \([(4 \times 20) + 12]\)

[French]

(32) deu-naw
two-nine
18 \([2 \times 9]\)

[Welsh; King 1993: 113]

Overrunning

(33) disąt-nocti
10-teen
20 \([10 + 10]\)

Note: 21 is janū disąt-nocti, i.e. \([1 + 20]\), not \([11-teen]\)

[Polabian; Polański & Sehnert 1972: 52, 73]

(34) soixante-dix
sixty-ten
70 \([60 + 10]\)

(35) soixante-douze
sixty-twelve
72 \([60 + 12]\)

(36) soixante-dix-sept
sixty-ten-seven
77 \([60 + (10 + 7)], \text{rather than } [(60 + 10) + 7]\)

[French]

E. Exponentiation and other higher bases

(37) \(10^1\quad 10^2\quad 10^3\quad 10^6\)
ten\quad hundred\quad thousand\quad million

[English]

Absence of exponentiation

(38) qliq-qlikkin
twenty-twenty
400 (20 x 20) – highest numeral in traditional system

[Chukchi; Skorik 1961: 388, 391]

Effectively monomorphemic series of bases

(39) 10 dáśa
      100 śatá
      1,000 sahasra
      10,000 ayúta
      100,000 lakśá
      1,000,000 prayúta
      10⁷ kóti
      10⁸ arbudá
      10⁹ mahārbuda
      10¹⁰ kharvá
      10¹¹ nikharva

[Sanskrit; Whitney 1889: 177-178, who notes further “The series of decimal numbers may be carried still further; but there are great differences among the different authorities with regard to their names; and there is more or less of discordance even from ayúta on.”]

Note: Indian English currently uses: lakh 10⁵; crore 10⁷; in India, 133,435,360 would be written as 13,34,35,360 (13 crore, 34 lakh, 35 thousand, 360).

(40) wàn 万       萬   10⁴
      yì 亿         愈  10⁸
      zhào 兆       10¹²
      jīng 京       10¹⁶ etc.

[Mandarin Chinese. This system is used in Chinese, Japanese, Korean; but see also (92)]

(41) cem-pōhual-li one-twenty-ABS 20
     cen-tzon-tli one-four.hundred-ABS 400 (20²)
     cen-xiquipil-li one-eight.thousand-ABS 8000 (20³)

[Classical Nahuatl; Andrews 1975: 397-398, 464, 482, 484]

Note: pōhua- is also a verb ‘to count’; tzon- also means ‘hair’; xiquipil- also means ‘bag, sack’.

(Semi-)productive systems
(42) | million | first | $10^6$ | $10^6$ |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>billion</td>
<td>second</td>
<td>$10^{12}$</td>
<td>$10^9$</td>
</tr>
<tr>
<td>trillion</td>
<td>third</td>
<td>$10^{18}$</td>
<td>$10^{12}$</td>
</tr>
<tr>
<td>quadrillion</td>
<td>fourth</td>
<td>$10^{24}$</td>
<td>$10^{15}$</td>
</tr>
<tr>
<td>[general pattern]</td>
<td>nth</td>
<td>$10^{6n}$</td>
<td>$10^{3(n+1)}$</td>
</tr>
</tbody>
</table>

[English]


(43) Contrast innovative / metric

| kilo- | $10^3$ |
| mega- | $10^6$ |
| giga- | $10^9$ |
| tera- | $10^{12}$ |
| peta- | $10^{15}$ |
| exa- | $10^{18}$ |
| zetta- | $10^{21}$ |
| yotta- | $10^{24}$ |

[English / International]

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

(44) cxr-as otxm-oc-da-cxra-meti
nine-hundred four-twenty-and-nine-teen
999 [(9 x 100) + (4 x 20) + (10 + 9)]
Bases: (10,) 20, 100

[Georgian; Hewitt 1995: 54]

(45) kàmpwòò ná ḥkwuu sícycéřé ’ná bée-tàärnɛ́ ná kɛ́ ’ná báář-icyčɛ́ɛ́řɛ́
fourhundred and 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

[Supyire Senoufo; Carlson 1994: 169]

(46) Note that the next higher base (nhb) is nearly always equal to or less than the square of the current base (cb):

$$\text{nhb} \leq \text{cb}^2$$
But note the older Germanic “long hundred”:

(47) 10  tíu
30  þrír tígar
100  tíu tígar
110  ellifu tígar
120  hundræð (“long hundred”)
240  tvæu hundræð
1200  þúsund (“long thousand”)

[Old Norse; Gordon 1957: 292-293]

(48) A higher multiplicative base is nearly always a product of the lowest multiplicative base. An exception is provided by Coahuiltecan, with multiplicative bases 3 (for numerals 12-19) and 20; surprisingly, the lower base 3 is itself complex (2 + 1).

(49) 12  4 x 3
13  4 x 3 +1
14  4 x 3 + 2
15  5 x 3
18  6 x 3 (?)
20  20
30  20 + 10
40  2 x 20

(50) puwânt'än axti-k-pil' ko pil'
four two-and-one and one
13 [i.e. \(4 \times (2 + 1)\) + 1]

[Coahuiltecan; Swanton 1940: 48]

Alternating bases

(51) 10  désat  10
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

[Resia Slovenian; Steenwijk 1992: 125]

F. Other arithmetic processes

Subtraction
(52) un-de-viginti
   one-from-twenty
   19 [20 – 1]

[Latin]

Division (actually: multiplication by fraction)

(53) hanner cant
    half hundred
    50 [$\frac{1}{2} \times 100$]

[Welsh; King 1993: 113]

Subtraction and addition

(54) éks bônsaŋ ki?
    twenty without hundred
    80 [100 – 20]

(55) ínam ókam éks bônsaŋ ki?
    two left over twenty without hundred
    82 [(100 – 20) + 2; NB: not 100 – (20 + 2)]

[Ket; Georg 2007: 179-181]

(56) Successive approximation, cf. time expressions in some languages

    drie (uur)
    three hour
    03:00

(57) half drie
    half three
    02:30

(58) vijf over half drie
    five after half three
    02:35

[Dutch]

Overcounting

(59) halv-tred-sinds-tyve
half-third-times-twenty
50 [half of the third, times twenty]

[Danish; Allan et al. 1995: 127; now usually: halvtreds]

(60) paūne tini šata
less.quarter three hundred
275 [less a quarter of the third hundred]

[Oriya; Karpuškin 1964: 38]

Pairing

(61) 1 seénu, wepul
2 woóí
3 báhi
4 naíki
5 mámni
6 búsani
7 wóo-búsani two-six (i.e. ‘second six’?)
8 wóh-naíki two-four (i.e. 2 x 4) ←
9 bátani
10 wóh-mámni two-five (i.e. 2 x 5) ←

[Yaqui; Dedrick & Casad 1999: 229]

(62) 1 hito 2 huta
3 mi 6 mu
4 yo 8 ya

[Japanese] 

Non-arithmetic structures

(63) 10⁸ arbudá-
10⁹ mahārbuda- (maha- ‘big’)

[Sanskrit; Whitney 1889: 177]

(64) 10³ mille (PL mila)
10⁶ milione (one AUGMENTATIVE)

[Italian]

G. Ordering of constituents
From larger to smaller

(65) sān-bāi  wǔ- shí  sì  
three-hundred  five-ten  four  
354 [i.e. 300 + 50 + 4]

[Mandarin Chinese]

From smaller to larger

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

[Malagasy (Standard); Rajaonarimanana 2001: 67]

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

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

(68) 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)]

[German]

(69) order smaller-larger (contrasting with higher larger-smaller)  
through 12 (ðó-ðeka, cf. 13 ðeka-trís)  Modern Greek  
through 15 (quin-ce, cf. 16 diec-i-séis)  Spanish  
through 16 (se-dici, cf. 17 dici-as-sette)  Italian  
through 19 (nine-teen)  English  
through 99 (neun-und-neun-zig)  German  

[For Modern Greek: Holton et al. 1997: 103-104]

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

(70) limam-polo  roe  amby,  amby  telo-njato  
five-ten  two  plus  plus  three-hundred  
‘352 [i.e. 50 + 2 + 300]’

[Malagasy (Nosy Be); Dahl 1968: 14]
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. tens 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. Finiteness and gaps (Comrie 2020)

“Every language has a numeral system of finite scope.”
(Generalization 1, Greenberg 1978)

<table>
<thead>
<tr>
<th>Short scale</th>
<th>Long scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>without intermediaries</td>
<td>with intermediaries</td>
</tr>
<tr>
<td>$10^6$</td>
<td>million</td>
</tr>
<tr>
<td>$10^9$</td>
<td>billion</td>
</tr>
<tr>
<td>$10^{12}$</td>
<td>trillion</td>
</tr>
<tr>
<td>$10^{15}$</td>
<td>quadrillion</td>
</tr>
<tr>
<td>$10^{18}$</td>
<td>quintillion</td>
</tr>
</tbody>
</table>

(74)  
- a. a new single-word term is available each time one multiplies the previously highest single-word term by 1,000,000, up to trillion, i.e. there are no terms *quadrillion* and beyond;
- b. there are no intermediaries, e.g. *milliard*, so $10^9$ can only be expressed as thousand million;
- c. numerals from million up can be combined to express multiplication, e.g. million trillion ($[10^6 \times 10^{18}] = 10^{24}$);
- d. there is a strong preference to implement the “Packing Strategy” (Hurford 1975: 67), i.e., informally speaking, to avoid combining terms with lower numerical value when there is an available appropriate term with higher numerical value; thus, although million million is possible, billion is preferred;
- e. it is not possible for a higher-value term to precede a lower-value term, e.g. for $10^{30}$ one can say billion trillion but not *trillion billion.*

Traditional British long-scale system (my version)
(76) Every number \( n (0 < n < L) \) can be expressed as part of the numerical system in any language.

[Generalization 2, Greenberg 1978]

(77) decillion, vigintillion, centillion
What is the 23rd term in the series?
tresvigintillion [https://en.wikipedia.org/wiki/Names_of_large_numbers]

I. Ambiguity

Parsing ambiguities

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

[English]

(79) un fil a thri ugain o asynod
one thousand and three twenty of asses
   (i) 1060 asses (possible interpretation)
   (ii) 61,000 asses (Numbers 31.34 – intended interpretation)

(80) deuddeg a thri ugain mil o eidionau
twelve and three twenty thousand of cattle
72,000 cattle (Numbers 31.33)

(81) saith mil a phedwar ugain mil
seven thousand and four twenty thousand
87,000 (I Chronicles 7.5)

[Welsh (Biblical); Hurford 1975: 192, 184, 176]

Abbreviation

(82) (nỳ) phan sɔɔŋ
one thousand two
   (i) 1002
   (ii) 1200
   (iii) *1020

(83) (nỳ) phan sɔɔŋ réɔy
one thousand two hundred
1200
(84) (นี่) phan กะ ส๊อง
one thousand with two
1002

[Thai; Smyth 2002: 173]

(85) (a) สาม-บ๊าย หก-สิบ
three-hundred six-ten
(b) สาม-บ๊าย หก
three-hundred six
360

(86) สาม-บ๊าย สาม-บ๊าย หก
three-hundred three-hundred six
306

Note: Obligatory use of ลิ้ง, creating a partial place-value system, avoids ambiguity.

(87) สาม-กิ๊น สาม-บ๊าย หก
three-thousand three-hundred six
3006

(88) สาม-กิ๊น สาม-บ๊าย หก-สิบ
three-thousand three-hundred six-ten
3060

[Mandarin Chinese]

Note: An older usage required ลิ้ง to be repeated as many times as there are zeroes in the “Arabic” numeral representation, so (87) would be สี่-กิ๊น สาม-บ๊าย หก (Chao 1968: 575).

Diachronic merger

(89) thirty, thirteen

[English]

(90) 7 สัต
100 สัต > จานิยสัต (จานิย ‘big’)

[Northern Mansi; Honti 1993: 125]

Semantic change

18
(91) billion
  (i) older, "long scale"  $10^{12}$
  (ii) newer, "short scale" = US English $10^9$
  (long scale: $10^{6n}$, short scale: $10^{3(n+1)}$)

[British English]

(92) 万/萬 亿/億 兆 京
     wàn  yì  zhào  jīng
  (i)  $10^4$  $10^5$  $10^6$  $10^7$
  (ii) $10^4$  $10^8$  $10^{12}$  $10^{16}$
  (iii) $10^4$  $10^8$  $10^{16}$  $10^{24}$
  (iv) $10^4$  $10^8$  $10^{16}$  $10^{32}$

[East Asian]

[Row (ii) is the one in current use for Chinese, Japanese, and Korean, though with occasional variation. Rows (i), (iii), (iv) are presented by Martzloff (1997: 99), based on the Chinese mathematical classic Shushu jiyi.]

Specialized use

(93) bak
     usually 400
     but 360 days (long calendar)

[Mayan]

[Tozzer 1921: 97, and more generally 97-103. The forms cited by Tozzer are more specifically Yucatec Maya; note that Tozzer uses the spelling baq. The Classical Mayan inscriptions and codices do not clearly include large numbers other than in calendrical accounts (Chrisomalis 2010: 292-294).]

(94) kilo-
     1000
     but: kilobyte ‘1024 bytes’

[English/International]

Body part systems

(95) siduŋ ‘shoulder’ = 10, 14, 33, 37, 56, 60

[Kobon; Davies 1981: 206-208]
J. Internal structure and psychological reality

(96) 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 sechsundfünfzig, literally
‘six and fifty’, for 56?

b) When people do arithmetic, do they operate with linguistic representations of
numerals (in which case problems (a) might arise) or do they operate with
abstract quantities or arithmetic notations (in which case they should not)?

c) But note that some arithmetic calculations might be easier in a non-decimal
base, e.g. 68 + 7 might be easier in a vigesimal base, where one does not have
to “carry over” a number; in decimal notation, 68 + 7 = 75; in vigesimal
notation 38 + 7 = 3F (where F = 15)

(97) Writing ‘six and fifty’ for 56 Dutch zesenvijftig
               German sechsundfünfzig

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.

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

(99)  
| 10  | ti  | tiende |
| 20  | tyve| tyvende|
| 30  | tred(i)ve | tred(i)vte |
| 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 | halvfemsinstyvende | cf. femte 5th |
| 100 | (et) hundred(e) | hundrede |

[Danish; Allan et al. 1995: 122]
There can be other discrepancies between formal structure (often historical/etymological) and current analysis for native speakers, cf. [Georgian; Hewitt 1995: 51-54]

Possible competing pedagogical advantages

a) Having a numeral system that corresponds to the arithmetic notation facilitates learning arithmetic.

b) Having a numeral system that does not correspond to the arithmetic notation provides extensive practice in arithmetic and leads to a higher level of arithmetic skill.

References

This list includes references from the body of the handout, plus some other more general works.

Chan, Eugene (compiler). Numeral systems of the world’s languages. https://mpi-lingweb.shh.mpg.de/numeral/.


Whitney, William Dwight. 1889. Sanskrit grammar, including both the classical language and the older dialects of Veda and Brahmaṇa. 2nd ed. Leipzig: Breitkopf & Härtel.


**Index of Languages**

The classification follows Glottolog 4.3 <http://glottolog.org>. The three-letter abbreviations are ISO 639-3 codes; parentheses indicate a partial match.

<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
<th>Type</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adzera</td>
<td>ADZ</td>
<td>Oceanic, Austronesian</td>
<td>Morobe Pr., Papua New Guinea</td>
</tr>
<tr>
<td>Balinese</td>
<td>BAN</td>
<td>Malayo-Sumbawan, Austronesian</td>
<td>Bali, Indonesia</td>
</tr>
<tr>
<td>Berom</td>
<td>BOM</td>
<td>Benue-Congo, Atlantic-Congo</td>
<td>Plateau State, Nigeria</td>
</tr>
<tr>
<td>Chukchi</td>
<td>CKT</td>
<td>Chukotko-Kamchatkan</td>
<td>Chukotka, Russia</td>
</tr>
<tr>
<td>Classical Nahuatl</td>
<td>NCI</td>
<td>Uto-Aztecan</td>
<td>Mexico [extinct]</td>
</tr>
<tr>
<td>Coahuilteco</td>
<td>XCW</td>
<td>isolate</td>
<td>Texas, USA; Mexico [extinct]</td>
</tr>
<tr>
<td>Danish</td>
<td>DAN</td>
<td>Germanic, Indo-European</td>
<td>Denmark</td>
</tr>
<tr>
<td>Dutch</td>
<td>NLD</td>
<td>Germanic, Indo-European</td>
<td>Netherlands; etc.</td>
</tr>
<tr>
<td>Ekari</td>
<td>EKG</td>
<td>Paniai Lakes, Nuclear Trans New Guinea</td>
<td>Papua, Indonesia</td>
</tr>
<tr>
<td>English</td>
<td>ENG</td>
<td>Germanic, Indo-European</td>
<td>England; USA; etc.</td>
</tr>
<tr>
<td>French</td>
<td>FRA</td>
<td>Italic, Indo-European</td>
<td>France; etc.</td>
</tr>
<tr>
<td>Georgian</td>
<td>KAT</td>
<td>Kartvelian</td>
<td>Rep. of Georgia</td>
</tr>
<tr>
<td>German</td>
<td>DEU</td>
<td>Germanic, Indo-European</td>
<td>Germany; etc.</td>
</tr>
<tr>
<td>Haruai</td>
<td>TMD</td>
<td>Piawi</td>
<td>Madang Pr., Papua New Guinea</td>
</tr>
<tr>
<td>Hindi</td>
<td>HIN</td>
<td>Indo-Aryan, Indo-European</td>
<td>North-Central India</td>
</tr>
<tr>
<td>Hixkaryána</td>
<td>HIX</td>
<td>Cariban</td>
<td>Amazonas, Brazil</td>
</tr>
<tr>
<td>Italian</td>
<td>ITA</td>
<td>Italic, Indo-European</td>
<td>Italy</td>
</tr>
<tr>
<td>Japanese</td>
<td>JPN</td>
<td>Japonic</td>
<td>Japan</td>
</tr>
<tr>
<td>----------------</td>
<td>-----</td>
<td>---------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Ket</td>
<td>KET</td>
<td>Yeniseian</td>
<td>W. Siberia, Russia</td>
</tr>
<tr>
<td>Kobon</td>
<td>KPW</td>
<td>Madang, Nuclear Trans New Guinea</td>
<td>Madang Pr., Papua New Guinea</td>
</tr>
<tr>
<td>Komnzo</td>
<td>(TCI)</td>
<td>Morehead-Maro, Yam</td>
<td>Western Pr., Papua New Guinea</td>
</tr>
<tr>
<td>Latin</td>
<td>LAT</td>
<td>Italic, Indo-European</td>
<td>Rome [extinct]</td>
</tr>
<tr>
<td>Malagasy, Nosy Be (variety of Sakalava Malagasy)</td>
<td>(SKG)</td>
<td>Basap-Greater Barito, Austronesian</td>
<td>NW Madagascar</td>
</tr>
<tr>
<td>Malagasy, Standard (= Plateau Malagasy)</td>
<td>PLT</td>
<td>Basap-Greater Barito, Austronesian</td>
<td>Madagascar</td>
</tr>
<tr>
<td>Mandarin Chinese</td>
<td>CMN</td>
<td>Sinitic, Sino-Tibetan</td>
<td>China</td>
</tr>
<tr>
<td>Mangarrayi</td>
<td>MPC</td>
<td>Mangarrayi-Maran</td>
<td>Northern Terr., Australia</td>
</tr>
<tr>
<td>Mayan</td>
<td></td>
<td>[language family]</td>
<td>Mesoamerica</td>
</tr>
<tr>
<td>Modern Greek</td>
<td>ELL</td>
<td>Greek, Indo-European</td>
<td>Greece, Cyprus</td>
</tr>
<tr>
<td>Mundurukú</td>
<td>MYU</td>
<td>Tupian</td>
<td>Amazonas/Mato Grosso/Pará, Brazil</td>
</tr>
<tr>
<td>Nauete</td>
<td>NXA</td>
<td>Central Malayo-Polynesian, Austronesian</td>
<td>East Timor</td>
</tr>
<tr>
<td>Ngiti</td>
<td>NIY</td>
<td>Central Sudanic</td>
<td>Orientale Pr., DR Congo</td>
</tr>
<tr>
<td>Northern Mansi</td>
<td>(MNS)</td>
<td>Mansic, Uralic</td>
<td>W. Siberia, Russia</td>
</tr>
<tr>
<td>Northern Pame</td>
<td>PMQ</td>
<td>Pamean, Otomanguean</td>
<td>San Luis Potosí, Mexico</td>
</tr>
<tr>
<td>Oksapmin</td>
<td>OPM</td>
<td>Asmat-Awyu-Ok, Nuclear Trans New Guinea</td>
<td>Sandaun Pr., Papua New Guinea</td>
</tr>
<tr>
<td>Old Norse</td>
<td>NON</td>
<td>Germanic, Indo-European</td>
<td>Scandinavia; Iceland [extinct]</td>
</tr>
<tr>
<td>Oriya</td>
<td>ORI</td>
<td>Indo-Aryan, Indo-European</td>
<td>Orissa, India</td>
</tr>
<tr>
<td>Pirahã</td>
<td>MYP</td>
<td>isolate</td>
<td>Amazonas, Brazil</td>
</tr>
<tr>
<td>Polabian</td>
<td>POX</td>
<td>Slavic, Indo-European</td>
<td>Germany [extinct]</td>
</tr>
<tr>
<td>Resia Slovenian</td>
<td>(SLV)</td>
<td>Slavic, Indo-European</td>
<td>Udine Pr., Italy</td>
</tr>
<tr>
<td>Russian</td>
<td>RUS</td>
<td>Slavic, Indo-European</td>
<td>Russia</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>SAN</td>
<td>Indo-Aryan, Indo-European</td>
<td>India [extinct]</td>
</tr>
<tr>
<td>Spanish</td>
<td>SPA</td>
<td>Italic, Indo-European</td>
<td>Spain; Latin America</td>
</tr>
<tr>
<td>Supyire Senoufo</td>
<td>SPP</td>
<td>Senufo, Atlantic-Congo</td>
<td>Mali</td>
</tr>
<tr>
<td>Thai</td>
<td>THA</td>
<td>Tai-Kadai</td>
<td>Thailand</td>
</tr>
<tr>
<td>Welsh</td>
<td>CYM</td>
<td>Celtic, Indo-European</td>
<td>Wales, United Kingdom</td>
</tr>
<tr>
<td>Yaqi</td>
<td>YAQ</td>
<td>Uto-Aztecan</td>
<td>Mexico</td>
</tr>
<tr>
<td>Yidiny</td>
<td>YII</td>
<td>Pama-Nyungan</td>
<td>Queensland, Australia</td>
</tr>
<tr>
<td>Yucatec Maya</td>
<td>YUA</td>
<td>Mayan</td>
<td>Yucatán, Mexico</td>
</tr>
</tbody>
</table>