Numerals in the World's Languages: An Update on Status and Interpretation

Harald Hammarstrm
harald2@cs.chalmers.se

March 12, 2008

Numerals are

- **spoken**
- **normed expressions** that are used to denote the
- **exact number** of objects for an
- **open class of objects** in an
- **open class of social situations** with
- **the whole speech community** in question

Notes on History and Scope of Research

- First major work on numerals in various languages:

  Lorenzo de Hervás y Panduro 1785 Aritmetica di quasi tutte le nazioni conosciute, Cesena.

- Since then (solely on numerals):
  - 127 PhD:s/Monographs
  - 634 Articles
- 10 000+ first-hand sources (grammars, vocabularies) thought to represent 4000-5000 languages.
- I owe all I know to fieldworkers and libraries.

Small Numeral Systems: Definition

A numeral system is *small* iff

1. Monomorphemic numerals exist only up to 2 or 3 AND
2. Higher quantities are expressed orally only inexactly or up to ca 10 with additions of 1, 2 and 3 (possibly including ad hoc use of ’hand’ for 5).

Note:

- A numeral system with normed usage of hand + feet for 5, 10, 20 will never be called a “small” numeral system.
- People with small numeral systems may use one or more of the following strategies to cope with higher quantities:
  1. Do without exact counting above 3 and live happily anyway
  2. Use hands, fingers and feet for tallying in an ad-hoc way
  3. Use hands, fingers and feet in a normed way (but without corresponding oral)
  4. Use numerals from another language
  5. Keep track of things by tallying or individuizing

Whence Small Numeral Systems?

- A limit at 2-3 coincides perfectly with the cognitively well-established **subitizing limit** =

  the number of objects one immediately sees how many they are, without grouping or counting.

- There are no languages with “rather small” numeral systems, say, numerals up to, e.g., 6 or 10.

- An independent variable predicts the presence/absense of small numeral systems!
  - All small numeral systems belong to Hunter-Gatherer languages
  - (BUT not all Hunter-Gatherer languages have small numeral systems)

Small Numeral Systems and Hunter-Gatherers

- A language is a Hunter-Gatherer (HG) language iff

  its speakers subsist more than 50% on hunted/gathered food

- Prediction Amendment 1: If a language
  - with known HG-status ethnographically
  - borrows numerals 3+ or 4+

=> then it had an original small numeral system

- Prediction Amendment 2: If a language *family*
  - can reconstruct numerals up to exactly 2 or exactly 3 (inclusive)

=> then it had an original small numeral system

Worldwide HG Survey

Australia: Almost universally small

Papuan Mainland: Almost universally small

SE Asia: Small with Malay Peninsula HG:s, Minor Mlabri and Andamanese. Negritos have ambiguous status.

S Asia: Not small/data lacking for Raute, Kusunda and Indian subcontinent HG

Siberia: Not small even in earliest attestations

N America: Most not small even in earliest attestations, but many have small systems with amendment 2

S America: A vast majority of HG languages have small systems

Africa: A vast majority of HG languages have small systems with amendment 2. Pygmies have ambiguous status.

Worldwide HG Survey: Numbers

	Small	Small-Amendment	Non-Small
HG	85	35	76
NON-HG	0	-	125

Non-Small Systems: Canonical Description

For every non-small system, there is a set of bases:

A small set of numbers b_i [= *b*ases]

Such that for every number $1 \leq x$ up to the limit:

$$x = a_n b_n + a_{n-1} b_{n-1} + \ldots + a_1 b_1 + a_0 \quad (1)$$

Where:

- $a_i < b_i$ for all i.
- $b_i > b_j$ if $i > j$.

Non-Small Systems: Generalizations

$$x = a_n b_n + a_{n-1} b_{n-1} + \ldots + a_1 b_1 + a_0 \tag{2}$$

- All morphemes in a $a_i b_i$ unit must be adjacent to each other in the spoken expression.
- All morphemes in a $a_i b_i$ unit is omitted if $a_i = 0$
- Pluses are expressed overtly or covertly and often different markings are used for different plus:es between different units
- Halves, quarters, zeroes, minuses, towards:es occur marginally
- The order between the additive units may vary (see later)
- All b_i are always overtly expressed, i.e., with a segmental morphemes (see later)

An Excursion: Body-Tally Systems

Count from the fingers of one hand continuing with some points along the lower and upper arm, reaching one or more points of the head, then ending with the corresponding body-parts on the opposite arm and finally hand. A number n is then denoted by nth the body-part-term in the sequence, e.g 'nose' or 'elbow on the other side'.

Body-Tally Systems: An Example

Number	Body part	Word for number	Number	Body part	Word for number
1	small finger	*mindi*	15	other eye	*ambi leetsia*
2	ring finger	*lapo*	16	other ear	*ambi aletsia*
3	middle finger	*tepo*	17	other neck	*ambi matsia*
4	index or forefinger	*tukumindi*	18	other shoulder	*napi pa yatsia*
5	thumb	*yau*	19	below other shoulder	*napi pilitsia*
6	wrist	*wataka*	20	above other elbow	*napi kitupatsia*
7	below elbow	*yanatsia*	21	below other elbow	*napi yanatsia*
8	above elbow	*kitupatsia*	22	other wrist	*napi wataka*
9	below shoulder	*pilitsia*	23	other thumb	*napi yau*
10	shoulder	*payatsia*	24	other index or forefinger	*napi tu kumindi*
11	neck	*matsia*	25	other middle finger	*napi tepone*
12	ear	*aletsia*	26	other fourth or ring finger	*napi lapone*
13	eye	*leetsia*	27	other small finger	*napi mindi*
14	nose	*ingatsia*			

Body-Tally Systems: Properties

- These systems typically have an **odd** number as the length of one round, in particular they **never** have 20 as the length of a round (“base”)
- Ambiguity as to whether one can felicitously count more than one round in abstracto
- Usually not used in an open class of social situations
- Usually there is a small numeral system as well, and the body-tally system have a marked morphosyntax
- Maximal occurrence constrast:
  - 5-10-20 system with the (only) body-related terms hand/foot/man occur **all over the world**
  - Body-tally systems occur **only** in Papua, vertically from the Sepik to the Fly (plus one genuine isolated case in southeast Australia)

Which Numbers Appears As Bases?

- For numbers less than a hundred, 5-10, 10-20, and 5-10-20 systems prevail across the world
- There are also bona fide rare instances of bases 4, 6, 8 and 12

Essentially, the correct solution was delivered by Aristotle 2 300 years ago!

Etymology as Evidence for Iconic Origins

The following are impressionistic estimates by HH:

- Atoms:

n	# roots	% etymology known	Dominant etym when known
5	Thousands	30%	99% ’hand’
10	Thousands	0%	-
20	Thousands	15%	99% ’one [full] man’
100	Hundreds	20%	’heap’/’bundle’/’measure’ and n

- Composites:

n	# cases	Dominant etymology when known
10	Thousands	99% ’[both/two] hands’
20	Thousands	99% ’two hands and two feet’/’two ten’
100	Less than a hundred	60% ’ten ten’/’X-ten’ and 25% ’five twenty’

Why No Place-Value System?

- Place-value systems assign base values to *places* (rather than overt morphemes).
- Place-value systems use overt zero-morphemes to signal empty places

$$
\begin{array}{lllll}
1 & 0 & 2 & 3 & = \text{'one thousand twenty-three'} \\
| & | & | & -> & 10^0 \cdot 3 = 3 \\
| & | & - & -> & 10^1 \cdot 2 = 20 \\
| & - & - & -> & 10^2 \cdot 0 = 0 \\
- & - & - & -> & 10^3 \cdot 1 = 1000
\end{array}
$$

- No language has a place value system! [Except marginal tendencies]
- This a clear counterexample to theories of language performance/economy! E.g. Hawkins Performance Grammar Correspondence Hypothesis predicts that 'one thousand twenty-three' would be:

  three two zero one

Why Homogeneity?

- Builiding blocks (multiplication, addition, base-numbers) of numerals are homogenous across languages
- **Also** the linguistics expression of them is homogenous. Contrast:
  - Clauses with subjects, object, topics, tenses, instruments show incredible hetergeneity across languages. Why does no language do:

    ’As for five, the hundred thousands the million with the four ontop of ten’ = 1 001 514
  - Other small/closed domains show heterogeneity across languages, e.g., spatial expressions, kinship systems
- Frequency for Homogeneity?

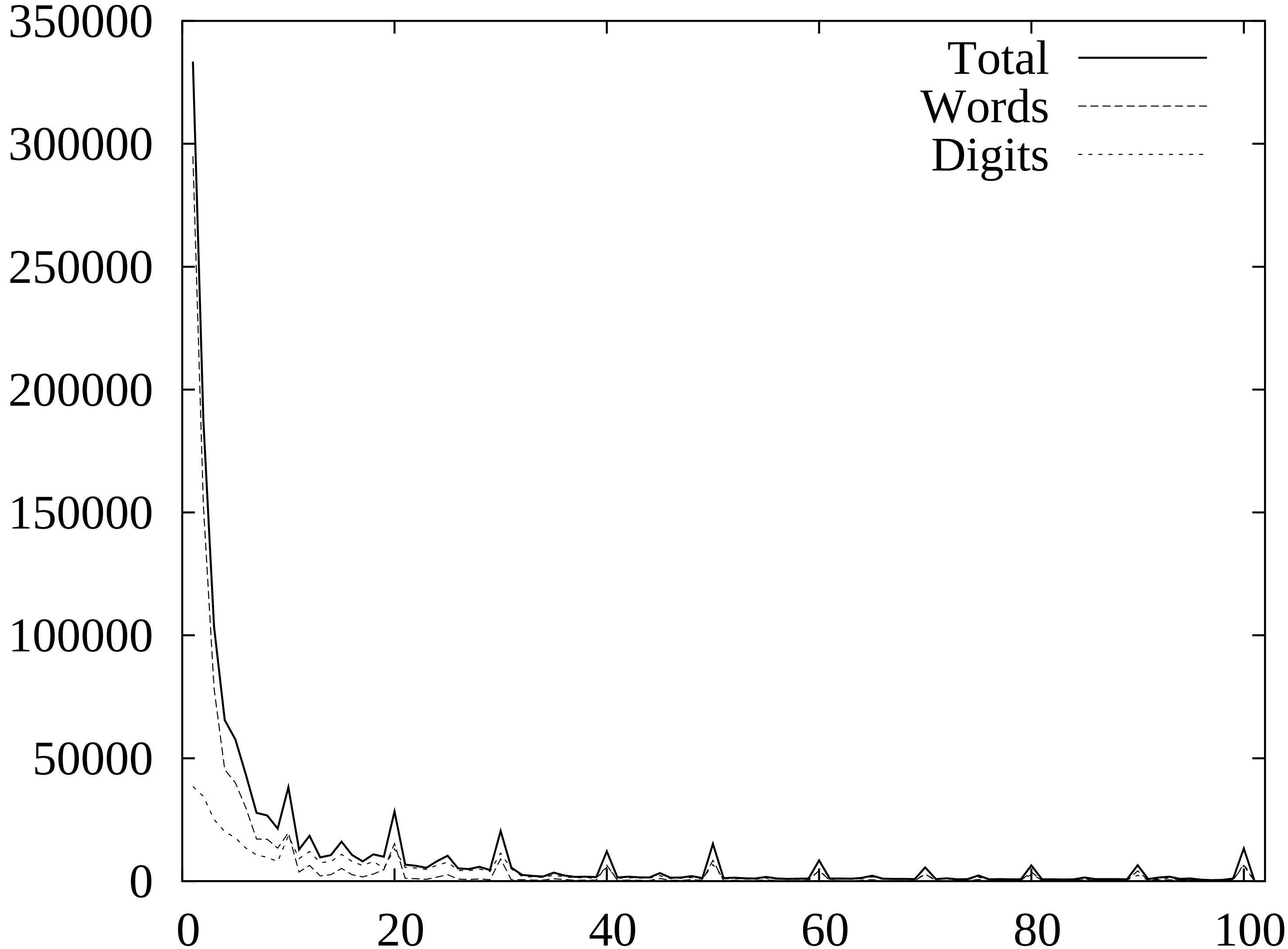

Figure 1: A frequency curve for numerals in the British National Corpus.

Why Ordering Restrictions?

- Below 100, one commonly finds both unit + ten and ten + unit order.

  ’twenty-one’ [ten+unit] vs. ’einundzwanzig’ [unit+ten]

- But above 100, the order is consistently bigger first, i.e. hundred + ten/unit rather than ten/unit + hundred, in almost all languages.
- There are so many independent cases that it cannot be accounted for by borrowing/diffusion.
- This is a semantic generalization (bigger-first):

  $=>$ Phrase structure grammars are not sufficient to account for legal numeral expressions.

The End

Thank You For Listening