

Opinion piece



Cite this article: Calude AS. 2021 The history of number words in the world's languages—what have we learnt so far? *Phil. Trans. R. Soc. B* **376**: 20200206.
<https://doi.org/10.1098/rstb.2020.0206>

Accepted: 24 September 2020

One contribution of 17 to a theme issue
'Reconstructing prehistoric languages'.

Subject Areas:
cognition

Keywords:
number words, numerals, cognition, frequency
of use, variation, linguistic typology

Author for correspondence:
Andreea S. Calude
e-mail: andreea.calude@waikato.ac.nz

The history of number words in the world's languages—what have we learnt so far?

Andreea S. Calude

School of General and Applied Linguistics, University of Waikato, Knighton Road, Hilcrest, 3216, Hamilton, New Zealand

ASC, 0000-0002-4190-5637

For over 100 years, researchers from various disciplines have been enthralled and occupied by the study of number words. This article discusses implications for the study of deep history and human evolution that arise from this body of work. Phylogenetic modelling shows that low-limit number words are preserved across thousands of years, a pattern consistently observed in several language families. Cross-linguistic frequencies of use and experimental studies also point to widespread homogeneity in the use of number words. Yet linguistic typology and field documentation reports caution against positing a privileged linguistic category for number words, showing a wealth of variation in how number words are encoded across the world. In contrast with low-limit numbers, the higher numbers are characterized by a rapid and morphologically consistent pattern of expansion, and behave like grammatical phrasal units, following language-internal rules. Taken together, the evidence suggests that numbers are at the cross-roads of language history. For languages that do have productive and consistent number systems, numerals *one* to *five* are among the most reliable available linguistic fossils of deep history, defying change yet still bearing the marks of the past, while higher numbers emerge as innovative tools looking to the future, derived using language-internal patterns and created to meet the needs of modern speakers.

This article is part of the theme issue 'Reconstructing prehistoric languages'.

'Numberland is a remarkable place. I would recommend a visit.'

[1, p. 11]

1. Why numbers?

As cultural objects go, numbers occupy a niche of their own; some can be surreptitious and veiled under a cloak of superstition, for example, the association of number 13 with bad luck in American culture, the link between number 4 and death in Chinese culture [2], the avoidance of counting people by numbers in Jewish orthodoxy on account of its bad omen [1]. Ongoing research of number words across languages paints a landscape of contrasts ([3–5] *inter alia*). On the one hand, we observe extreme variation in the different types of systems that various cultures use (or not!) to refer to quantities [6], and on the other hand, there is remarkable agreement and stability across speakers and time with regard to systems of counting once these are established, particularly for low-limit numbers [7]. So, what can number words tell us about our linguistic prehistory and human evolution?

2. The world of numbers—variation at every step, especially in restricted number systems

For more languages than previously assumed, either number words are virtually non-existent, or else they do not designate exact quantities [8, p. 414].

The most cited examples of restricted number systems come from two Amazonian languages: Pirahã, which lacks numbers beyond 1 and 2 and whose number words are not used consistently [9,10], and Mundurukú, which has number words for 1 to 4, but which only uses the first three consistently [11, p. 8]. Somewhat arbitrarily but not unusually, Mundurukú uses *fat*, *arms* and *parents* to represent *two*, *three* and *four*, respectively [12]. Number systems based on body parts are not uncommon and can also be found in New Guinea, as illustrated from the Oksapmin and Yupno languages, among others [13,14].

The Amazon is by no means the only place with innumerate or restricted number systems; others include Australia ([15], cited in [16, p. 260], [17]), and Papua New Guinea [6]. Low-limit numbers typically comprise unanalysable atoms that originate from different means of subsistence, such as hunting, farming, fishing, or from physical objects observed in the environment. For instance, some dialects of Hup, spoken in the Vaupés River use variants of *kəwəʔg-ʔap* meaning ‘eye-quantity’ for *two* (presumably because eyes come as a pair; this also holds for an unrelated language, Karitiana (D. L. Everett 2021, personal communication) and variants of *məʔ-wig-ʔap* meaning ‘rubber-tree seed quantity’ for *three* [16, p. 267]. Systems such as the Hup counting system illustrate how languages change and rework their representations of number. Studying these languages allows us to ‘read the history of a [number] system, just like the history of an old building, from the contrasting style of its pieces, from the foundations up’ [18, p. 83]. Yet caution needs to be exercised in this ‘reading’ because foundations can occasionally be renovated beyond recognition. In some languages spoken in New Guinea, decimal number systems used for higher numbers were eventually replaced by quinary or body-counting systems [19, p. 215]. In other words, languages can go in either direction, changing from simpler to more sophisticated systems but also from more sophisticated to simpler ones.

Number systems of the world’s languages do not only show a contrast between highly restricted and almost infinitely productive systems. Typological inquiries reveal numerous peculiarities and a tangled mix of coexisting number-forming patterns for varying quantities as numbers get larger. Thus, variation in number word patterns is not merely a foible of languages with few number words; variation can be found at every step.

Some languages have multiple counting systems that coexist side by side, depending on what is being counted. In the Austronesian language of Takuu, different number words are used to count humans (4 = *takahaa*), fish (4 = *haa*), canoes (4 = *tauwakahaa*), length of rope or wood (4 = *lohahaaa*), money (4 = *ha*) and coconuts or stones (4 = *ruaoa*) [20]. In fact, the system is even more complex, because money terms are counted in units of 10 cents, and coconuts and stones are counted in units of two up to quantities of 10, beyond which individual items can be counted (say, 11 coconuts). These counting words combine the numeral proper with a numeral classifier, and Takuu is by no means alone in using such a system (see [21] for further examples). Seen with European eyes, wherein higher-number words can be used universally to count any object, be it fish, humans or coconuts, it might be tempting to regard such object specificity found in languages like Takuu as inefficient and primitive, betraying a less sophisticated understanding of number concepts. However, this evaluation misses the important

advantage that classifier systems provide (see [19,22] and a discussion of the early accounts of Māori counting system in [23]). Numeral classifier systems are devised in order to ‘count’ the entities that speakers need counted, according to the cultural niche of their specific environment. A trade-off is made in favour of shorter number words, which facilitate arithmetic operations of those particular objects that are counted frequently, with the burden being shifted to remembering a higher number of distinct (number) words.

Then there are also languages that employ multiple strategies for expressing different quantities. Haruai (New Guinea) has three different coexisting counting systems: the old indigenous system, a body part system that draws on Kobon (a geographically close language) and the Tok Pisin system [24]. The Hup language Dāw uses unrelated atoms for numbers 1–5 (though unusual for their semantic transparency, as illustrated earlier) and switches to a tally system based on body parts for numbers 5–20, and to a third system for higher numbers, borrowed from Portuguese [16, pp. 270–271]. In the Indo-European languages Czech, Faroese and Bokmål, two alternative word orders are available for numbers higher than 20, such that both 20 + 3 and 3 + 20 are possible, as was also the case in Latin [2].

Restricted number systems, numeral classifiers and systems with multiple strategies for forming number words corroborate the view that a language is a tool embedded within a cultural context and an environmental niche, constantly evolving to better serve the needs of the speakers using it [3].

3. Our numerical past—where number systems are predictable and consistent, low-number words are old

One pattern that appears to hold mostly unchallenged (dialects of Hup notwithstanding) is the representation of low-limit number words by means of unanalysable atoms. For languages that do have them, the history of such low-limit number words is astonishingly deep. Indo-European, Bantu, Austronesian and Pama-Nyungan languages have all been shown to preserve cognate forms of low-number words that have stubbornly lingered around for tens of thousands of years of linguistic evolution [7,17].

Lexical replacement rates vary enormously among words and among languages. In words that linguists believe to be the least rapidly changing within a given language, namely words that designate basic vocabulary terms, like *foot*, *green*, *man*, *dirty*, *husband*, *wife*, *mother*, and including numbers *one* through to *five*—a collection of words termed the Swadesh List (named after Morris Swadesh, who formulated various such lists)—rates of lexical replacement can still vary between word-forms as much as 100-fold [7, p. 8]. But number words stand out as being among the most conservatively preserved word-forms even in such basic vocabulary lists [7]. Remarkably, in the Indo-European language family, a *single* cognate set can be traced throughout its entire history, indicating astonishing agreement across speakers and time [7]. Put another way, speakers of Indo-European languages have preserved ancestral forms for low-limit numbers with extreme fidelity over thousands of years of language change.

Low-limit numbers are acquired early in life, reminiscent of Ernst Haeckel’s nineteenth century catch phrase *ontogeny recapitulates phylogeny* (the acquisition process mirrors the

evolutionary development process). Number space mapping experiments with US-American and Mundurukú adults and children show that despite linguistic differences in their available lexical resources for expressing numbers (Mundurukú only has numbers equivalent to *one* through to *four*, see [25]), all participants can place smaller numbers in a number space more or less accurately (when being asked to place dots of varying quantities onto a computer screen after being presented with a set-up of one dot to the left and, say, ten dots to the right). However, both the US-American and Mundurukú groups become less accurate as the numbers get larger [26, p. 1217]. These findings support Weber's Law, whereby 'increasingly larger quantities are represented with proportionally greater imprecision, compatible with a logarithmic internal representation with fixed noise' [26]. Where the English-speaking and Mundurukú participants differ is the quantity of dots at which logarithmic representations kick in: earlier for the Mundurukú participants (the effect was observed for both children and adults). Interestingly, the two sets of participants perform more similarly to each other when the Mundurukú use Portuguese number words, exposing the importance of the context of use in the construal of number words. Language is by no means a linguistic prison, but it does tend to send us down the well-beaten path.

Corpus frequencies of number-word use also point to high rates of agreement between Indo-European languages for numbers 1–10, as well as for higher decimal numbers (10, 20, 30 and so on) and 100s and 1000s [27,28]. Smaller numbers are used most frequently, with a linear decrease in frequency observed for higher numbers, though there are small peaks among the higher numbers, at 10, 12, 15, 20, 50 and 100 [27]. That said, the high frequency of use of low-limit numbers is still not sufficiently high to explain the deep history observed for these forms [7].

In a recent study, Pagel *et al.* [29] modelled the frequency distributions of various answers extracted in priming experiments of American English, included in the LAMSAS [30] and LAGS [31] datasets. The results confirm that number words are under high amounts of positive frequency-dependent selection, having extremely few variants and incredible rates of agreement among speakers. While, for some concepts, American English speakers provided a variety of different answers, for example, in some cases, answers included as many as ten different near-synonyms (*parlour*, *living-room*, *sitting-room*, *setting-room*, *front room*, *drawing room*, *hall*, *library*, *den*, *big room* and others), for number words, answers tended to concentrate on one highly dominating variant. As an evolutionary mechanism, the positive frequency-dependent selection is neither concerned with inherent linguistic properties (of the sort, short words are preferred), nor explained in terms of matching frequencies of use (of the type, use the same proportional frequencies of word-form X as others do). Instead, positive frequency dependence selection denotes a bias that makes speakers disproportionately likely to use a word (say the number *two*) that most other speakers use. This bias grants an increasing advantage to those forms that are used productively and provides a mechanism to explain how a shared vocabulary can spontaneously self-organize and then be maintained for centuries or even millennia, despite new words continually entering the lexicon. The fact that frequency distribution curves of number words are best captured by positive frequency dependence selection suggests that there is more going on than just

adapted linguistic form and social pressures to conform to speaking norms.

4. Number words—not a privileged category

All the evidence thus shows that low-limit numbers in languages that have productive higher numbers behave in a stable, uniform manner across large time scales and varied speaker populations. The question is, why are these low-limit numbers so resistant to change? A highly plausible hypothesis comes from the lack of variation in the system [7]. Owing to their concrete and specific meanings [32], there is less room for near-synonyms to develop and even when they do, these remain context-restricted and low in frequency (compare *twelve* with *dozen*), leading to fixation. This is precisely what was observed of the LAMSAS and LAGS American English data [31]. The findings also support a more general law of semantic change, the Law of Innovation, proposed by Hamilton *et al.* [33], which contends that polysemous words tend to change their meanings faster, showing Social Conformist Bias effects in language change. Yet, it is still unclear what keeps the variation among number words so low; why do we entertain various words for *parlour* but only one for *three*?

Could it be that number words constitute a privileged category unlike all other linguistic categories? Given that children learn number words sooner than we might predict based on the general frequency of use of number words in a given language, one suggestion could be that our brains may be innately predisposed to number words and 'number cognition' [7, p. 5].

However, taking typological variation seriously, especially in non-industrialized societies, the evidence is stacking against the idea that numbers and numerosity may be conceptualized in a one-dimensional line [14,34] and against number cognition as an innate capacity [8]. Núñez allows a (weaker) predisposition towards *quantical* cognition [8]. His notion of *quantical* capacity refers to 'biologically endowed abilities for perceiving and discriminating quantities' [8, p. 421], which *can* then lead to an explicit encoding of (more or less) exact numerical quantities, should the cultural and environmental context encourage it. Crucially, while such abilities *can* lead to this explicit coding, they *need not* do so. Following on from Núñez, Everett [35] checks whether languages privilege quantical concepts, for example, singular versus plural distinctions are highly common cross-linguistically, but dual and especially trial and paucal systems, though possible, are extremely rare. Backing Núñez, Everett arrives at the conclusion that number concepts are not privileged, in other words, that 'there is no neat relationship between some blueprint of our cerebral architecture and the edifice of numerical language constructed in a given culture' [35, p. 13].

Further support for this conclusion comes from a clinical study which shows that it is possible for aphasia patients to lose their ability to use number words without losing other parts of the language system [36]. Typology also shows that it is similarly possible for a language to lose number words once acquired, for instance the ancestral number 4 seems to have been lost in some languages of the Australian Pama-Nyungan family [37, pp. 5–6].

One characteristic that sets the Pama-Nyungan languages apart from the Indo-European ones with regard to number

systems is the fact that the former language family has primarily languages with a restricted number system. So it appears that both higher numbers behave differently from low numbers, and languages with productive and consistent higher numbers tend to behave differently from those with restricted systems. As we progress from smaller quantities towards larger quantities, the morphological patterns observed become more regular and transparent, with an elegant pattern capturing this development: ‘the degree of morphological fusion varies inversely with the size of the numerical value’ [38, p. 281].

Neatly, the regularities spill beyond number-formation processes themselves and into the wider language system. For most Indo-European languages, building higher numbers and extending number words to denote larger quantities involves following (existing) language-internal rules. Thus, excluding the lower numbers (1–9), which are encoded by unanalysable atoms, and the running numbers (11–19), given their considerable variation, for the remaining higher numbers, stable morphological patterns can be detected where the base (10s, 20s, 30s, 40s) is combined with the atom (3 in 23, 5 in 55) in a head-dependent manner, in accordance with patterns involving nominal and clausal word orders [2]. That is, morphological word-formation process ties in with syntactic constraints. These patterns suggest that number words are at the cross-roads between lexical content and grammar (as also remarked by [5]). While exciting for the understanding of number words themselves, the findings render higher numbers unsuitable avenues for examining deep history.

5. Conclusion

The history of ideas and insights gained about number words is too long to be aptly and faithfully captured in a few pages.

References

1. Bellos A. 2010 *Alex's adventures in numberland*. Oxford, UK: Bloomsbury.
2. Calude A, Verkerk A. 2016 The typology and diachrony of higher numerals in Indo-European – a phylogenetic comparative study. *J. Lang. Evol.* **1**, 91–108. (doi:10.1093/jole/lzw003)
3. Beller S, Bender A, Chrismalis S, Jordan F, Overmann K, Saxe GB, Schlimm D. 2018 The cultural challenge in mathematical cognition. *J. Num. Cogn.* **4**, 448–463. (doi:10.5964/jnc.v4i2.137)
4. Dehaene S. 1997 *Number sense: How the mind creates mathematics*. Oxford, UK: Oxford University Press.
5. Epps P, Bower C, Hansen C, Hill J, Zent J. 2012 On numeral complexity in hunter-gatherer languages. *Ling. Typol.* **16**, 41–109. (doi:10.1515/lity-2012-0002)
6. Hammarström H. 2010 Rarities in numeral systems. *Rethink. Universals* **45**, 11–53. (doi:10.1515/9783110220933.11)
7. Pagel M, Meade A. 2017 The deep history of number words. *Phil. Trans. R. Soc. B* **373**, 20160517. (doi:10.1098/rstb.2016.0517)
8. Núñez R. 2017 Is there really an evolved capacity for number? *Trends Cogn. Sci.* **21**, 409–424. (doi:10.1016/j.tics.2017.03.005)
9. Gordon P. 2004 Numerical cognition without words: evidence from Amazonia. *Science* **306**, 496–499. (doi:10.1126/science.1094492)
10. Everett DL. 2005 Cultural constraints on grammar and cognition in Pirahã. Another look at the design features of human language. *Curr. Anthropol.* **46**, 89–130. (doi:10.1086/431525)
11. Gelman R, Butterworth B. 2005 Number and language: how are they related? *Trends Cogn. Sci.* **9**, 6–10. (doi:10.1016/j.tics.2004.11.004)
12. Rooryck J, Saw J, Tonda, A, Pica P. 2017 Mundurucu number words as a window on short-term memory. See <https://halshs.archives-ouvertes.fr/halshs-01497577>.
13. Wassmann J, Dasen P. 1994 Yupno number system and counting. *J. Cross-Cult. Psychol.* **25**, 78–94. (doi:10.1177/0022022194251005)
14. Núñez R, Cooperrider K, Wassmann J. 2012 Number concepts without number lines in an indigenous group of Papua New Guinea. *PLoS ONE* **7**, e35662. (doi:10.1371/journal.pone.0035662)
15. Hale K. 1975 Gaps in culture and grammar. In *Linguistics and anthropology: in honour of C. F. Voegelin* (eds MD Kinkade, KL Hale, O Werner), pp. 285–315. Lisse, The Netherlands: Peter de Ridder Press.
16. Epps P. 2006 Growing a numeral system. *Diachronica* **23**, 259–288. (doi:10.1075/dia.23.2.03epp)
17. Bower C, Zent J. 2012 Numeral systems in Australian languages. *Anthropol. Ling.* **54**, 130–166. (doi:10.1353/anl.2012.0008)
18. Hurford J. 1987 *Language and number*. Oxford, UK: Basil Blackwell.
19. Beller S, Bender A. 2008 The limits of counting: numerical cognition between evolution and culture. *Science* **319**, 213–215. (doi:10.1126/science.1148345)
20. Moyle R. 2011 *Takuu grammar and dictionary: a Polynesian language of the South Pacific*. Canberra, Australia: Pacific Linguistics.
21. Bender A, Beller S. 2006 Numeral classifiers and counting systems in Polynesian and Micronesian

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. This study was supported by the New Zealand Royal Society Catalyst Seeding Fund.

Acknowledgements. I thank Mark Pagel for reading drafts of this review, Caleb D. Everett for his example from Karitiana, Søren Wichmann for his generous support, and the two anonymous reviewers as well as the editors of the special issue, Antonio Benítez-Burraco and Ljiljana Progovac, for their suggestions and comments. I am grateful to the NZ Royal Society Catalyst Seeding Fund for their generous funding. Any remaining errors are, of course, my own.

- languages: common roots and cultural adaptations. *Oceanic Ling.* **45**, 380–403. (doi:10.1353/ol.2007.0000)
22. Bender A, Beller S. 2014 Mangarevan invention of binary steps for easier calculation. *Proc. Natl Acad. Sci. USA* **111**, 1322–1327. (doi:10.1073/pnas.1309160110)
 23. Overmann KA. 2020 The curious idea that Māori once counted by elevens, and the insights it still holds for cross-cultural numerical research. *J. Polynesian Soc.* **129**, 59–84. (doi:10.15286/jps.129.1.59-84)
 24. Comrie B. 1999 Haruai numerals and their implications for the history and typology of numeral systems. In *Numeral types and changes worldwide* (ed. J. Gvozdanović), pp. 81–94. Amsterdam, The Netherlands: Mouton de Gruyter.
 25. Pica P, Lemer C, Izard V, Dehaene S. 2004 Exact and approximate arithmetic in an Amazonian indigene group. *Science* **306**, 499–503. (doi:10.1126/science.1102085)
 26. Dehaene S, Izard V, Spelke E, Pica P. 2008 Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* **320**, 1217–1220. (doi:10.1126/science.1156540)
 27. Dehaene S, Mehler J. 1992 Cross-linguistic regularities in the frequency of number words. *Cognition* **43**, 1–29. (doi:10.1016/0010-0277(92)90030-L)
 28. Calude A, Pagel M. 2011 How do we use language? Shared patterns in the frequency of word-use across seventeen world languages. *Phil. Trans. R. Soc. B* **366**, 1101–1107. (doi:10.1098/rstb.2010.0315)
 29. Pagel M, Beaumont M, Meade A, Verkerk A, Calude A. 2019 Dominant words rise to the top by positive frequency-dependent selection. *Proc. Natl Acad. Sci.* **116**, 7397–7402. (doi:10.1073/pnas.1816994116)
 30. Kretzschmar WA. 1993 *Handbook of the linguistic atlas of the Middle and South Atlantic*. Chicago, IL: University of Chicago Press.
 31. Pederson L, McDaniel SL, Bailey G, Basset M. 1986 *Handbook for the linguistic atlas of the Gulf States*. Athens, GA: University of Georgia Press.
 32. Brysbaert M, Warriner AB, Kuperman V. 2014 Concreteness ratings for 40 thousand generally known English word lemmas. *Behav. Res. Methods* **46**, 904–911. (doi:10.3758/s13428-013-0403-5)
 33. Hamilton WL, Leskovec J, Jurafsky D. 2016 Diachronic word embeddings reveal statistical laws of semantic change. In *Proc. 54th Ann. Meeting Assoc. Comput. Ling., Berlin, Germany, 7–12 August 2016*, vol. 1, pp. 1489–1501. Stroudsburg, PA: Association for Computational Linguistics. (doi:10.18653/v1/P16-1141)
 34. Cooperrider K, Marghetis T, Núñez R. 2017 Where does the ordered line come from? Evidence from a culture of Papua New Guinea. *Psychol. Sci.* **28**, 599–608. (doi:10.1177/0956797617691548)
 35. Everett C. 2019 Is native quantitative thought concretized in linguistically privileged ways? A look at the global picture. *Cogn. Neuropsychol.* **37**, 340–354. (doi:10.1080/02643294.2019.1668368)
 36. Domahs F, Bartha L, Lochy A, Benke T, Delazer M. 2006 Number words are special: evidence from a case of primary progressive aphasia. *J. Neuroling.* **19**, 1–37. (doi:10.1016/j.jneuroling.2005.07.001)
 37. Zhou K, Bower C. 2015 Quantifying uncertainty in the phylogenetics of Australian numeral systems. *Proc. R. Soc. B* **282**, 20151278. (doi:10.1098/rspb.2015.1278)
 38. Greenberg J. 1978 Generalisation about numeral systems. In *Universals of human language: word structure* (eds J. Greenberg, C. Ferguson, E. Moravcsik), pp. 249–295. Stanford, CA: Stanford University Press.
 39. Núñez R. 2011 No innate number line in the human brain. *J. Cross-Cult. Psychol.* **42**, 651–668. (doi:10.1177/0022022111406097)