Mira Ariel Tel Aviv University

Numerous puzzles

Dehaene et al (1999) and Menary (2015) propose that it is the precise number *words* that enable the shift from a continuous and approximate arithmetic intuition to a precise and digital system. Pagel and Meade (2017) note the extremely low rate of renewal for numerals, as compared with most words (e.g., *nice* originally 'not knowing=ignorant', replaced by *ignorant*). Such renewals happen at an average rate of 2000-4000 language years. But numerals are only renewed "every 10,000, 20,000 or even more years" (ibid). Pagel and Meade speculate that this stability is due to the fact that number words are unambiguous. But why should numerals remain (i) precise and (ii) unambiguous, when most lexical items, originally precise ones included, often end up (i) approximate and (ii) polysemous? And why are round numbers sometimes approximate nonetheless?

I offer a distinction between sparse and dense lexical fields as explanation for these puzzles. Standard lexical items have prototype category structure and form part of *sparse* lexical fields; nonround numerals do not have prototype category structure, and belong to *dense* lexical fields; and round numerals participate in both systems. It is the differential discourse usage of sparse and dense lexical items that accounts for the stability of the precise meaning of nonround numerals on the one hand, and the variable (approximate) interpretations of sparse-system items on the other hand.

A lexical field is a set of lexemes denoting similar, but distinct and competing meanings. Most lexemes belong to sparse lexical fields, where competing items differ from each other on more than one parameter. Hence, there's a conceptual "no man's land" between their meanings (e.g., between *nice* and *attractive*). Typical interpretations adjust the lexical meaning by encroaching on the item's neighboring lexemes (e.g., 'nice, close to attractive'):

1. I've got a big date, Francis… Do I look **nice**?? (*LA Times,* June 3, 2000).

Interestingly, the same is true for precisely defined lexemes. For example, only a small minority of *straight* and *middle* tokens (13.8%, 8.3% respectively, in the Santa Barbara Corpus of spoken American English) were interpreted as 'exactly straight/middle'. Moreover, since sparse-system lexemes are routinely adjusted in context they need not be overtly marked in order to receive an adjusted (approximate) interpretation (0.7%-5.6% such markings for *nice*, *blue*, *green*, and *red*, 0% for *middle*, 0.8% for *straight*).

The linguistic number system is quite unique. It offers an extremely dense partition of the relevant lexical field. There exist an unusually high number of distinct number expressions in the immediate semantic neighborhood (e.g., *6, 7* for *5*), which are only *minimally* different from each other (on a *single*, +1 parameter). Typically, there is no need to adjust the meaning ('N') of number expressions because there's enough of a variety of them: Within the common, +1 granularity functional in natural discourse, there is no conceptual 'no man's land' to encroach on (I estimate that 0.24% of numeral expressions include decimals). I propose that the density of the system keeps each numeral *distinct* and *literally faithful* (i.e., interpreted as 'exactly N', blocking encroachment upon neighboring number expressions). Indeed, unlike *straight* and *middle*, an examination of 2 SBC conversations revealed that 84.5% of the bare numerals were interpreted as 'exactly' (Ariel, 2002).

Now, the preciseness of the numeral system, seemingly perfect for referential purposes, may be lacking with respect to speakers' abilities and goals. Since numerals do not have prototype category structure with fuzzied boundaries, overt marking is necessary in order to encroach on the neighbors. Indeed, overt adaptors (e.g., *about*) are much more frequent for numerals: 15.5%-32.4% (average, 24.7%) for *two, seven, ten, seventeen* and *thirty* versus sparse-system lexemes (average 2.35% for the above).

Still, although typical number expressions cannot (easily) expand, round numerals are rather easily interpreted as 'approximately N'. I claim that numerals can be so contextually adjusted provided they have a double membership, participating in a sparse lexical system in addition. The question is, then, why can some, but not other numerals, escape the dense system? I propose that while there is no conceptual space surrounding numerals, the aggregate of ‘exactly N’ items can consolidate into a single unified conceptual whole carrying an added meaning. Instead of only conveying an aggregate of N items the numeral can be interpreted as a (i) single conceptual and discourse unit, (ii) comprising of N items. As such, it is no longer only minimally different from its numeral neighbors. Its meaning (also) functions as a standard, prototype category, with fuzzied boundaries, hence no longer restricted to its literal cardinality (enabling e.g., an approximate value interpretation). Interestingly, roundness and the added value of a collective meaning are not necessarily one and the same. Although biblical *seven* is not a round number, it is (also) a "magic" number, and as such, can be used approximately (and see Krifka, 2007 for a distinction between 40 and 45 in reference to time, since the latter stands for a salient time category).

If I am correct, we should find minimal pairs, where the difference between the lexemes pertains to how dense/sparse the system is that the expressions participate in. Indeed, the denser *zero* was interpreted as 'exactly zero' 89.3% of the time, but *nothing* was so interpreted only in 35.7%. *Fifty percent* was interpreted as 'exactly half' in 45.1% of the cases, but *half* was so interpreted in only 8%. In addition, when approximate, *50%* was overtly marked as such 3.3 times more than *half*.

Semantic change (and potential renewal) result from synchronic variation. Variation requires the fuzzied borders of prototype categories within a sparse lexical system. This is why *half* could become polysemous, and doubles as 'part' (*half-baked*). Unlike *two*, *a couple* ('two+added value') is now predominantly interpreted as 'several'. *Hundred* and *thousand* now mean '100' and '1000' respectively, but they originally meant '120' and 1200'. Switching to a base-ten system, speakers preserved the conceptual unit meaning of the two lexemes, rather than their precise numerical value.

In sum, an originally precise meaning cannot guarantee the numerals' precise and monosemous use, nor their diachronic stability. I argue that (i) their membership in a dense lexical system blocks prototype category structure (with fuzzied boundaries); (ii) Their rigid-boundary structure blocks synchronic variation, and (iii) the absence of variation blocks semantic change, thus preserving their precise meaning.

References

Ariel, Mira. 2002. Privileged interactional interpretations. *Journal of Pragmatics* 34:1003–1044.

Dehaene, Stanislas, Elizabeth Spelke, P. Pine, R. Stanescu and N. Tsivkin. 1999. Sources of Mathematical Thinking: Behavioral and Brain-Imaging Evidence. *Science* 284:970-974.

Krifka, Manfred. 2007. Approximate interpretation of number words: A case for strategic communication. In Gerlof Bouma, Irene Krämer and Joost Zwarts, eds., *Cognitive foundations of interpretation*. Amsterdam: Koninklijke Nederlandse Akademie van Wetenschapen, 111-126.

Menary, Richard. 2015. Mathematical Cognition - A Case of Enculturation. In Thomas Metzinger and Jennifer M. Windt, eds., *Open MIND: 25(T)*. Frankfurt am Main.

Pagel, Mark and Andrew Meade. 2017. The deep history of the number words. *Philosophical transactions of the Royal Society of London B 373: 20160517*.