Cardinal Numerals Revisited in GF

Harald Hammarström & Aarne Ranta
Dept. of Computing Science, Chalmers University, 41296 Göteborg, Sweden

We have implemented the grammars for forming cardinal numerals in a wide array of languages in the general CL tool Grammatical Framework (GF)[1]. GF is based on the notion of a shared abstract syntax which is linearized into different languages according to each language's specific grammar. The format GF mandates the grammars to be written in enables both parsing and generation, so that GF can also perform translation via the abstract syntax.

In our sample of roughly 60 languages from all over the world we survey what kind of parameters govern different formations. Then we show how these can be fully represented and implemented. The basics about generative grammars for numerals has of course been known since the late 60s[2]. The novelty in our case it that we show that it is indeed feasible to unite very diverse numeral systems under one single abstract syntax.

The abstract syntax is designed with two goals in mind: generality and prototypicity. It should be general enough to be able to, without resorting to blunt listing, accommodate the varieties of generality we actually find in natural languages. Secondly, we want to have very concise implementations for most languages in that as many features should be 'built-in' as possible without disturbing generality. This amounts to the same thing as saying that there is such a thing as a prototypical numeral system where many languages cluster around the prototype and the more deviation from it the scarcer the languages.

We present empirical evidence that the trade-off between generality and prototypicity lies (not surprisingly) around a zeroless multiplicative-additive base 10-system, allowing for base 20-variation under 100 and that 11–19 aren't to be formed like 20–99 but we have failed to find non-trivial prototypicity on e.g. agreement features, classifier syntax or linking elements. Obviously, many details merit deeper discussion, such as the question of a one and the same abstract syntax for numeral systems of different bases[3]. We can also note some interesting cases for the numeral grammar engineer; some well-known e.g the irregularity of Hindi numerals < 100 or the absence of an atom for 1000 for Ge'ez, as well as some lesser known e.g. Misantla Totonac where a non-stabilized indigenous numeral system (base 10–20) meets the borrowed Spanish and gives rise to a large number of variants.

At the same time as defining how numerals are formed we had the chance to gather some typological data. Surely, anthropologists and typologists[4] as well as historically interested mathematicians[5] have long pointed out the generalities and universal constraints on numerals. But it still interesting to add more empirical data and to do counts the way typologists do i.e. sample randomly and beware of areal and genetic bias. Our sampling resembles that of Bell[6] but due to limitations on time we have only been able to discount for genetic inheritance. (So our results are only really interesting under an assumption that the other factors aren't significantly influential). For instance, we find that with overwhelmingly greater than chance frequently, order of larger additive units precede smaller if over 100, but below 100 both unit + ten, ten + unit (or five, twenty + unit etc) are common. We can also confirm that NumN correlates with multiplier multiplicand. As shown by Dryer[7], NNum is only really common in Africa and really uncommon outside Africa. Our database also includes data on irregularity, incidence on participants in complex forms (taking into account etymology where available), linking morphemes and more.

The work results in a numeral applet freely accessible from the web, in which users can generate and translate numerals in the interval of 1–999,999 in more than 50 languages.

[1] Aarne Ranta, **Grammartical Framework: A Type-Theoretical Grammar Formalism** To appear in the *Journal of Functional Programming.*

[2] Hugo Brandt Corstius (ed.), **Grammars for Number Names** D. Reidel Publishing, 1968 and James R. Hurford, **The Linguistic Theory of Numerals**, Cambridge University Press, 1975

[3] Much new data on bases is presented in Glendon Lean by way of Kay Owens, **The Work of Glendon Lean on the Counting Systems of Papua New Guinea and Oceania**, *Mathematics Education Research Journal* 13:1(2001):47-71

[4] Notably W. C. Eels, **Number Systems of the North American Indians**, *The American Mathematical Monthly* XX:10(1913):293-299, Joseph H. Greenberg, **Generalizations About Numeral Systems** in Universals of Language Vol. 3, Stanford University Press, 1978: 250-295, Frans Plank and Elena Filimonova, **The Universals Archive** http://ling.uni-konstanz.de/pages/proj/sprachbau.htm, accessed 20030911

[5] The most thorough source from this angle is Georges Ifrah, **The Universal History of Numbers**, John Wiley & Sons, 2000 (Translation from the French 1982 original)

[6] R. D. Perkins, **Statistical Techniques for Determining Sample Size**, *Studies in Language* 13:2(1989):293-315

[7] Matthew S. Dryer, **Large Linguistic Areas and Language Sampling**, *Studies in Language* 13:2(1989):257-292