Epistemic Numbers

Jon Ander Mendia – Heinrich-Heine University

INTRODUCTION Epistemic Numbers (ENs) are complex cardinal numerals formed by combining a number with a non-numerical expression, typically an indefinite. Unlike other cardinal numbers, ENs (*i*) denote a range of possible values, and (*ii*) convey an epistemic effect of uncertainty or vagueness about the exact number. For instance, (1) is an acceptable answer to a question like *how many people came?*, and it can denote integers ranging from 21 to 29, not more and not less.

(1) Twenty-some people came $\rightsquigarrow$ $^{\checkmark}[21,29]$; $^{\star}\leq 20$; $^{\star}\geq 30$

GOALS This paper provides two things: an initial description of the main cross-linguistic facts about ENs and a compositional semantic analysis that is flexible enough to account for a variety of languages. I focus on three main cross-linguistic facts: (*i*) some languages allow ENs with a variety of quantifiers, not just indefinites akin to *some* or *wh*-indefinites, (2); (*ii*) in some languages ENs can either precede or follow the numeral, (3); and (*iii*) the interpretation of ENs depends on the numerical base of the language, and cannot combine with just any numeral, (4).

(2) **a.** 20 y muchos euros
20 and many euros
$^{\checkmark}[26,29]$; $^{\star}\lesssim 26$; $^{\star}\geq 30$

b. 20 y pocos euros [Spanish]
20 and few euros
$^{\checkmark}[21,24]$; $^{\star}\leq 20$; $^{\star}\gtrsim 24$

(3) **a.** juu - nan - nin - ka - ga
10 what CL_{PERSON} PART NOM
$^{\checkmark}[11,19]$; $^{\star}\leq 10$; $^{\star}\geq 20$

b. nan - juu - nin - ka - ga [Japanese]
what 10 CL_{PERSON} PART NOM
$^{\checkmark}\{10,20,30\ldots,90\}$; $^{\star}\leq 10$; $^{\star}\geq 90$

(4) **a.** hogei-ta-zak
20-and-"some"
$^{\checkmark}[21,39]$; $^{\star}\leq 20$; $^{\star}\geq 40$

b. *hogei-ta hamar-zak [Basque]
20-and 10-"some"
Intended: $[31,39]$

The meanings of (2) are more restrictive than (1): the possible numbers that (2a) denotes are limited to the set formed by the upper half of (1), that is, any integer from 26 to 29, maybe less, but not more. Similarly, (2b) denotes an integer in the lower half of (1)–some integer from 21 to 24; maybe more, but not less. (3) shows that the mode of combination matters: unlike (3a), (3b) denotes tuples of 10. (4) shows that ENs are sensitive to the internal complex structure of cardinal numbers in the language: (4a) is compatible with a range of values that corresponds to the base of 20 that Basque has in this case, but an attempt to overcome this limitation results in ungrammaticality, (4b). Notice that none of these expressions are simply *approximate* (cf. Krifka 2009, Solt 2017), their meaning is more limited. Moreover, upon hearing any of (1)–(4), a hearer may draw the inference that the speaker does not know what the exact number is, or that she thinks that the exact value is irrelevant. **ANALYSIS** The analysis has three main ingredients: (*i*) a baseline proposal for the syntax/semantics mapping of cardinal numerals across languages; (*ii*) an extension to the indefinite expressions that participate in ENs, where they are analyzed as determiners introducing subset selection functions over a set of numbers; and (*iii*) a pragmatic calculus that derives the uncertainty/vagueness component as a form of primary quantity implicature. **The baseline.** I assume that cardinal numbers (simple and complex) are properties of degrees (5a) (Landman 2004). All numbers have a complex internal structure (cf. Huford 1975): following the positional notation for numbers, I propose that non-additive cardinals are the product of some integer $i \in \{1,2,\ldots,9\}$ and the corresponding numerical BASE B^i, where $\text{B}^i = 10^i$ for any $i \in \mathbb{N}$ (e.g., $6 \times \text{B}^1 = 6 \times 10 = 60$). In the decimal system, the BASE is a power of 10, and so its denotation is a property of degrees too: $\lambda d.[d = 10^i]$. The multiplication is carried out by the head MUL (5c). Different BASE values are freely eligible by MUL; for example, the number 3000 has the structure $[3[\text{MUL B}^3]]$, which is interpreted as $3 \times \text{B}^3 = 3 \times 1000 = 3000$. Complex additive cardinals are assumed to have an underlying coordinate structure (Ionin & Matushanky 2006), headed by an additive head ADD–sometimes spelled-out as the connective *and*– with the semantics of addition (Anderson 2015; (5b)). (Notice that ADD cannot take two cardinals formed by the same BASE as arguments (**veinte y treinta*, "twenty-thirty").) Thus, $[80[\text{ADD } 9]] = 80 + 9 = 89$.

(5) **a**. $[\![60]\!] = \lambda d\,.\,[d = 60]$ **b**. $[\![\text{ADD}]\!] = \lambda D^{i}\,.\,\lambda D^{i+n}\,.\,\lambda d\,.\,\exists d'd''[d = d' + d'' \wedge D^{i}(d') \wedge D^{i+n}(d'')]$

c. $[\![\text{MUL}]\!] = \lambda D'\,.\,\lambda D''\,.\,\lambda d\,.\,\exists d'd''[d = d' \times d'' \wedge D'(d') \wedge D''(d'')]$

d. $[[3[\text{MUL B}^{3}]][\text{ADD}[6[\text{MUL B}^{2}]][\text{ADD}[8[\text{MUL B}^{1}]][\text{ADD}[9[\text{MUL B}^{0}]]]]]]]]$

$\Leftrightarrow$ [3000 [ADD [600 [ADD [80 [ADD [9]]]]]]] $\Leftrightarrow$ 3689

The analysis can be adjusted to different languages by adjusting the bases accordingly to different values. Moreover, idiosyncratic differences (i.e. cross-linguistic, diachronic) can be accounted for by limiting different aspects of the construction. E.g., in the general case, the first argument of ADD must always be smaller (**twenty-and-hundred*, **veinte y cien*), as indicated by the superscript on the property variables in (5b), but the choice seems to be largely conventional and semantically it would not matter to have them either way (c.f. the nursery rhyme *four-and-twenty blackbirds*). **Extension to ENs.** The denotation of the indefinites that partake in ENs (*some*, *-nan-*, et.), $[\![\text{INDF}]\!]$, is modeled after the epistemic indefinite *algún* in Spanish, as defined in (6a). $[\![\text{INDF}]\!]$ denotes a property of degrees, a subset selection function over degrees f that selects some integer from a set of numbers (cf. Alonso-Ovalle & Menéndez-Benito 2013). The size of the set restricting f depends entirely on the BASE that $[\![\text{INDF}]\!]$ combines with: the *minimum* is determined by the BASE itself (B^{i}), whereas the *maximum* is the next power of the base: B^{i+1}. (This constraint could enforced in a number of ways, for instance as a presupposition on the domain of f; cf. φ-features on pronouns.) Combining $[\![\text{INDF}]\!]$ with the BASE via MUL results in a property of degrees that can then be added to some other property of degrees, which, in turn, was also formed by the same process out of a bigger BASE (e.g., $[6[\text{MUL B}^{1}]] = \lambda d\,.\,\exists d'd''[d = d' \times d'' \wedge d' = 10 \wedge d'' = 6]$, which is equal to $6 \times 10 = 60$, as in (6c)). The smallest BASE B^{0} is a special case, the *minimum* is 0, not B^{0}.

(6) **a**. $[\![\text{SOME}]\!] = \lambda d\,.\,d \in f(\{n \in \mathbb{N} : \text{B}^{i} < n < \text{B}^{i+1}\})$

b. $[\![\text{MUL}]\!]([\![\text{B}^{0}]\!])([\![\text{INDF}]\!]) = \lambda d\,.\,\exists d'd''[d = d' \times d'' \wedge d' = 1 \wedge d'' \in f(\{n \in \mathbb{N} : 0 < n < 10\})]$

c. $[\![\text{ADD}]\!]([\![(6b)]\!])([\![20]\!]) = \lambda d\,.\,\exists d'd''[d = d' + d'' \wedge d' \in f(\{n \in \mathbb{N} : 0 < n < 10\}) \wedge d'' = 20]$

Refinement: *muchos/pocos*. ENs with *muchos/pocos* in Spanish further restrict the range of possible values the ENs can denote. In these cases *muchos* and *pocos* are proportional, as they are interpreted relative to the BASE they combine with: in *60 y muchos*, *muchos* ranges from 6 to 9, but in *ciento y muchos* ("hundred and many") it ranges from 51 to 99, and so on. Following common proposals on proportional quantifiers (cf. Partee 1988), I define $[\![\text{MUCHOS}]\!]$ so that it further restricts the *lowest* possible value of f by relying on a fraction of the BASE^{+1}; $[\![\text{POCOS}]\!]$ is defined so that the *highest* possible value of f is now a fraction of BASE^{+1}.

(7) **a**. $[\![\text{MUCHOS}]\!] = \lambda d\,.\,d \in f(\{n \in \mathbb{N} : \frac{\text{B}^{i+1}}{2} < n < \text{B}^{i+1}\})$ **b**. $[\![\text{POCOS}]\!] = \lambda d\,.\,d \in f(\{n \in \mathbb{N} : \text{B}^{i} < n < \frac{\text{B}^{i+1}}{2}\})$

The rest of the derivation proceeds as above: *pocos/muchos* and BASE serve as arguments to MUL, and the resulting property of degrees serves as the first argument of ADD. **Calculating implicatures.** I assume that utterance of a sentence φ by a speaker S commits S to the knowledge of φ ($\text{K}_{S}\varphi$; by Hintikka (1962)'s Epistemic Implication). The fact that f's domain is always restricted to a non-singleton set triggers the question as to why the speaker did not restrict the domain further. As a consequence, Stronger Alternatives (SAs) to the assertion (8b) are negated and this, together with the assertion, entails that for every SA $[n]$, $\neg\text{K}_{S}[n]$. In order to derive ignorance ($\neg\text{K}_{S}[\varphi] \wedge \neg\text{K}_{S}\neg[\varphi]$), SAs must be closed under disjunction: the disjunction of two SAs is also a SA if it asymmetrically entails the assertion. Together, the assertion and the full set of negated SAs entail that every value in the range denoted by the EN must be an epistemic possibility for the speaker, (8c).

(8) **a**. *I spent twenty-some dollars* $\rightsquigarrow \text{K}_{S}[21, 29]$ **b**. $SA([\![(8a)]\!]) = [21], [22], [23], \ldots, [29], [21 \vee 22], \ldots [27 \vee 28 \vee 29] \ldots$

c. $\text{K}_{S}[21, 29] \wedge \neg\text{K}_{S}[21] \wedge \neg\text{K}_{S}\neg[21], \neg\text{K}_{S}[29] \wedge \neg\text{K}_{S}\neg 29]$

CONCLUSION Semantically, the general baseline analysis is flexible enough to account for a wide variety of number systems. The full paper explains how the resulting property of degrees proceeds its life as an argument to an entity-taking predicate. Pragmatically, the implicatures calculated provide *total ignorance*, i.e. ignorance about every value in the range, which is generally a welcome result. The paper discusses a number of exceptions, and argues that they can be handled by trimming the set of SAs in ways that are consistent with the information that is common ground among the participants in the conversation.