DOI: http://dx.doi.org/10.31513/linguistica.2017.v13n2a14031

DISCRETE INFINITY AND THE SYNTAX-SEMANTICS INTERFACE

Uli Sauerland[1] and Pooja Paul[2]

Discrete infinity was identified as a central feature of human language by Humboldt who famously spoke of making infinite use of finite means. Later Chomsky refocused attention on this property starting with Chomsky (1957). In a number of works since, Chomsky has repeatedly stressed the centrality of infinity for understanding language. For example, Chomsky (2007) writes that "An I-language is a computational system that generates infinitely many internal expressions". Chomsky also noted that the property of discrete infinity is shared by the natural numbers and language. This connection has also caught the interest of others in cognitive science (e.g. Dehaene 1999, Dehaene et al. 1999). In this squib, we want to discuss concrete reductions of discrete infinity of the natural number. Specifically, we want to investigate the extent to which this connection is compatible with current views of the syntax-semantics interface. We argue that merge alone is not enough to derive infinity, but a minimal lexicon is necessary, as Chomsky (2007) has noted in passing. We furthermore show that Chomsky's assertion that a single lexical item is sufficient to generate the natural numbers depends on two assumptions -- an untyped lambda calculus, and a specific interpretation of the syntactic Merge operation.

In mathematics, the Peano axioms (or Peano-Dedekind axioms) represent one characterization of the natural numbers to a finite set of axioms. This formulation does not uniquely characterize the natural numbers, but also allows so-called "non-standard models" that satisfy Peano's axioms. Von Neumann's set-theoretic construction of the natural numbers in (1) conceives of 0 as the empty set, and then constructs for any number n its successor s(n) by union of n with the singleton set n. This construction leverages the finite means of set theory to construct an infinity.

(1) standard set-theoretic construction of the the natural numbers

$$0 = \{\}$$

$$s(n) = n \cup \{n\}$$

1 E-mail: uli@alum.mit.edu
2 E-mail: puja149@gmail.com

For the construction of natural numbers from language, it seems initially plausible that the internal structure of number words should be our guide. As Ionin and Matushansky (2006) argue, complex number words such as “two hundred and fifty three” can be analyzed as syntactically and semantically compositional. On their view, there is a finite set of simplex cardinals, which may vary across languages -- e.g. in “one”, to “nine”, but also “eleven”, “twelve”, and possibly “twenty”, “thirty” and other multiple of 10. Among the simplex cardinals of English, some powers of 10 can acts as multiples -- “hundred”, “thousand”, “million”, “billion”, “trillion”, “quadrillion” to “nonillion”. Addition and multiplication are also part of Ionin and Matushansky’s system. The view I suggest goes beyond what Ionin and Matushansky propose, so the criticism I then make doesn’t impinge on their theory at all. Namely, I want to consider the idea that the concept of numbers is derived only through the internal structure of complex number words. This would entail that without the syntax for complex number terms in the language only the numbers denoted by simplex terms would be accessible to mathematical thought.

However, this view is not easy to sustain, because it would predict that for any speaker the number line should have an upper limit. Ionin and Matushansky stress that conventions constrain which number labels are actually used -- e.g. while “four hundred” is the valid label for ‘400’ in English, “twenty twenty” or “forty ten” are not. The conventionalization limits the productivity of the number system as afforded by the multiplication and addition operations. Crucially, this predicts an upper bound on the numbers for which a linguistic label can be constructed. In English, the sequence of powers of 1000 (“million”, “billion”, “trillion”, “quadrillion”, …) that are simplex expressions has an upper bound in most speakers’ knowledge. This would seem to predict that the numbers end at some point, perhaps at 999 nonillion 999 octillion 999 septillion 999 sextillion .... But as far as we know, an English speaker wouldn’t conclude that the numbers themselves end. Instead speakers are aware of the fact that the English language, or their knowledge of it, merely runs out of words to denote numbers at some point.

This entails that the reduction of number to language must rely more directly on the internal operations of language, in particular Merge. Chomsky (2007) suggests that “If the lexicon is reduced to a single element, then Merge can yield arithmetic in various ways.” Chomsky himself doesn’t elaborate on this interesting remark. First consider Chomsky’s implicit claim that in the absence of a lexicon, Merge cannot yield discrete infinity. If merge was simply set formation, this would not be the case. A construction analogous to that of Peano above would suffice: since the lexicon is empty, merge could at the base step only apply to an empty input, yielding the empty set as its output. But then the empty set could be the input for Merge yielding {{}}, and furthermore a discrete infinity of representations. But Chomsky has repeatedly argued that the “simplest case” of Merge ought to be restricted to no fewer than two inputs (Chomsky, 2007). Chomsky makes the natural assumption that this restriction of Merge blocks the application of Merge to an empty set.

While this rules out the von Neumann construction of the natural numbers for language, Chomsky suggests in the same paper that a single element in the lexicon suffices to yield the natural numbers. Thus, Merge would need to apply in the simplest case to just a single lexical element - still an apparent contradiction with Merge only applying to two elements. But, the case of a single element lexicon seems different from that of an empty lexicon. We think there are two possible ways to understand Chomsky's remark, both of which ultimately amount to similar results. On the one hand, one lexical item may yield multiple occurrences, X_1 and X_2, when it is copied from the lexicon to the syntactic system. Then X_1 and X_2 can be merged, yielding {X_1, X_2}. This system works, but is perhaps a bit worrying in its reliance on not only the lexicon, but furthermore its dependence on the operation that creates different, possibly infinitely many "occurrences" from the same element of the lexicon. In more recent work, Starke (2001), Citko (2005), Fox & Johnson (2015) and others have sought in different ways to reduce the scope of occurrence creation. An alternative to keep in mind for the natural numbers, is an application of Merge that Guimarães (2000) has called "self-merge". Assume that X is the single unique element of the lexicon, then self-merge of X to X would produce the set {X, X}. Formally {X, X} is not distinct from {X}, but as we discuss below, we assume that the language system is built to regard any node constructed via merge as having two daughters. This predicts the representation of the result of self-merge to be {X, X}. Once {X, X} is constructed in this way, subsequent merge can now construct a discrete infinity by merging {X, {X, X}} and so on. Whether we distinguish multiple occurrences or not, the system then generates from a single lexical item, a discrete infinity of representations.

Let us now consider how Chomsky's reduction of the numbers to language could be mapped at the syntax-semantics interface. For concreteness, we assume following Heim & Kratzer (1998) that only a restricted set of composition principles may apply at the syntax semantics interface. In particular, the two possible composition rules are Function Application and Predicate Modification, as stated in (ii), because all other composition rules that Heim & Kratzer discuss don't apply: the lambda abstraction rule applies only if one of the nodes is a variable abstractor, which X isn't, and all other interpretation rules apply only to primitive nodes (i.e. not the results of merge). We think though that ultimately our results are independent of this assumption.

(i) The interpretation of a node C is given either by:
 a. functional application, namely, if {A, B} = C, then [[C]] = [[A]]([[B]]), or
 b. predicate modification, namely, if {A, B} = C, then [[C]] = λ x . [[A]](x) and [[B]](x)

The interesting consequence of Chomsky's proposal is that the interpretation of the single lexical item X must combine with itself. Neither of the two interpretation rules in (i) gives the right result: self-combination of [[X]] with [[X]] by Predicate Modification (ib) is inadequate, since [[X X]] reduces trivially to [[X]] by the idempotency of conjunction. Function Application [[X]]([[X]]) is necessarily undefined in the typed lambda calculus that Heim & Kratzer use because, if [[X]] is of type α, it cannot take an argument of type α. In some of the theoretical literature, the typed calculus is extended to systems with polymorphic types. It is possible that such polymorphic-type system could overcome the limitation noted in the text. A less rigid system, namely property theory, that allows self-application, has been discussed for natural language semantics in earnest only by Chierchia &

Turner (1988) to account for cases of apparent self-application such as (ii).

(ii) a. Nicety is nice.
b. Goodness is good.

However, it is difficult to see how such an account should be extended to the natural numbers, and furthermore this account of cases like (i) has not been accepted widely in the field. Furthermore, the intuitions about the interpretation of such sentences are not very sharp -- it is quite conceivable that the speakers don't conclude from (ii-a) that the property of "nicety" itself is nice, but only that it's nice if people are nice. Suppose a linguist struggles with the vague concept of "nicety", but likes concepts such as "round" and "square" -- the latter she calls "nice" concepts, while she regards "nicety" as horrible. At the same time, she finds nice people very nice. It is still conceivable that such a person might truthfully utter (i-a). This obviously is not a knock-down argument against this approach, since it may just be a case of domain restriction of the quantifier, and as such, still compatible with Chomsky's proposal. In any event, we pursue a different type of proposal for the natural numbers.

Concretely, we think the Church numerals provide a natural fit for the construction of the natural number system, given the prominence of lambda calculus in semantics. We provide the definition of the Church numbers 0 and 1, as well as the successor function in (iii). As can be seen, their construction relies on the lambda calculus.

(iii) a. $[[0]] = \lambda f. \lambda x . x$
b. $[[1]] = \lambda f . \lambda x . f(x)$
c. $\text{successor}(n) = \lambda n \lambda f \lambda x \ . f (n (f) (x))$

With Church numbers, the number of bound occurrences of *f* that apply recursively to the variable *x* corresponds to the number value of a term. The numbers 2 and 3 are given in (iv).

(iv) $[[2]] = [[\text{successor}(1)]] = \lambda f . \lambda x . f(f(x))$
$[[3]] = [[\text{successor}(2)]] = \lambda f . \lambda x . f(f(f(x)))$

As such the Church numbers still require at least two basic lexical symbols, for example "successor" and "1", to derive the natural number series. But, by making the composition more flexible, it is possible to arrive at a more flexible system. One possibility is to allow addition as a composition rule. Addition is standardly defined as in (v).

(v) $[[\text{addition}]] = \lambda n \lambda m \lambda f \lambda x . m(f) (n(f) (x))$

With that we can state a new composition rule for binary branching nodes, with sub-constituent meanings n and m as $[[\text{addition}]](n)(m)$. This leads us to a system where $[[1\ 1]] = [[2]]$, and for any complex merged structure S of 1's, the semantic value of S corresponds to the number of terminal 1-nodes in the tree. One interesting consequence of this proposal is that numbers greater than or equal to four don't have one unique representation, but rather can be represented in two or more different ways. For three and smaller numbers, this isn't the case, assuming that the merged tree structures aren't linearize. Consider the one possible representation of "three" and two distinct representations of "four" in (vi)

(vi) a. "three" = [[1 1] 1]
b. "four" = [[1 1] [1 1]] ,   [[[1 1] 1] 1]

The ambiguity of "four" and the higher numbers on this implementation is an interesting, and we believe novel, observation. It could predict that identity statements such as "four equals four" could be judged false. This clearly would be a disadvantage of this implementation of the number system, but we do not think it a necessary consequence. Instead of requiring "equals" to demand identity of structural representation, it is more faithful to our view to only require identity of interpretation. In this case, "four equals four" is a necessary truth since both representations are interpreted as "λ f λ x . f (f (f (f (x))))". Then an attractive prediction of this view of numbers is that it leads us to expect that manipulating numbers up to three ought to be cognitively easier than it is for numbers greater than three.

A second, alternative implementation of the numbers would not adopt a new composition rule, but instead propose a lexical ambiguity of the basic number, which we'll call "ONE" for this account. The proposal is as in (v): "ONE" can either represent "1" or the successor function, and the correct interpretation is determined by the structural context.

(v) [[ONE]] = { λ f . λ x . f(x), λ n λ f λ x . f (n (f) (x)) }

We assume that "two" is represented as the result of merging ONE with itself, as in (vi); however, the two occurrences of "ONE" must be interpreted differently: one of them as 1, the other as the successor function. The two occurrences have the same status in every sense, so we cannot say which one receives which of the two interpretations. However, it is necessarily the case that they receive different interpretations. The possibility of different interpretations in a case of self-merge are to our mind, a bit surprising, but we do not know of any other evidence regarding the interpretation of self-merged structure.

(vi) "two": [[ONE ONE]] = successor (1) = lambda f . lambda x . f(f(x))

The proposal in (v), in contrast to the addition-rule, does not predict an ambiguity for any numerals. For instance, "four" must be represented as the second structure in (vii), since the first one is uninterpretable.

(vii) [[ONE ONE] [ONE ONE]]: uninterpretable
"four" = [ONE [ONE [ONE ONE]]]

But one may ask here whether the proposal of a two-way lexical ambiguity isn't really simply a way of packing two lexical items into one morpheme, and therefore departing from Chomsky's idea of using only a single lexical item to generate the full set of natural numbers. The proposal then more accurately should be represented as assuming two lexical items: 1 and the successor function.

One final possibility to consider is to view the ambiguity of (v) as the result of a general type-ambiguity similar to the type ambiguity proposed for connectives by Rooth & Partee (1982) among others. This though would lead us to expect a further ambiguity of ONE in at least three ways, as in (viii).

(viii) $[[\mathrm{ONE}]] = \{ \lambda f . \lambda x . f(x),$
$\lambda n \lambda f \lambda x . f(n(f)(x)),$
$\lambda m \lambda n \lambda f \lambda x . m(f)(n(f)(x)) \}$

The third lexical entry for ONE in (viii) represents the function composition of addition and 1, and therefore can combined with two numerals n and m resulting in a representation of n + m + 1.This proposal predicts again that "four" and higher numerals are ambiguous: If we represent the three interpretations listed for ONE in (viii) as ONE_1, ONE_2, and ONE_3 in that order, we see that the type resolution represented in (ix) renders the structure [[ONE ONE] [ONE ONE]] interpretable.

(ix) [[ONE ONE] [ONE ONE]] = [[ONE_3 ONE_1] [ONE_2 ONE_1]]

We conclude that Chomsky's proposal to represent the natural numbers in language by the use of a single lexical item has an interesting consequence: Numbers greater than "three" must have ambiguous representations. This is not the case if we represent numbers with two lexical items, 1 and the successor function. The difference between the proposals relates to the fact that starting with a single lexical item requires us to assume that both syntactic and semantic composition must be more flexible: self-merge must be possible and a new way of composition must be assumed. The proposal assuming two lexical items can do without such flexibility. The two proposals can thus be distinguished by either investigating further the independent evidence that has be adduced in favor of such compositional flexibility. But a second possibility is to directly target the prediction that the "unambiguous" numbers up to three should be easier to process as compared to the "ambiguous" numbers four and greater.

ACKNOWLEDGMENT

We thank Noam Chomsky and Pierre Pica for personal discussion that lead to some of the ideas written up here.

REFERENCES

Chierchia, Gennaro & Raymond Turner. 1988. Semantics and property theory. Linguistics and Philosophy 11. 261–302.

Chomsky, Noam. 1957. Syntactic Structures.

Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), Interfaces + recursion = grammar? Chomsky's minimalism and the view from syntax- semantics, vol. 89, 1–30. Berlin, Germany: Mouton de Gruyter.

Citko, Barbara. 2005. On the nature of merge: External merge, internal merge, and parallel merge. Linguistic Inquiry 36(4). 475–496.

Dehaene, Stanislas. 1999. The number sense: How the mind creates mathematics. Oxford, UK: Oxford University Press.

Dehaene, S., E. Spelke, P. Pinel, R. Stanescu & S. Tsivkin. 1999. Sources of mathematical thinking: Behavioral and brain-imaging evidence. Science 284(5416). 970–974.

Fox, Danny & Kyle Johnson. 2016. QR is restrictor sharing. In Kyeong-min Kim & other (eds.), Proceedings of WCCFL 33, 1–16.

Guimarães, Maximiliano. 2000. In defense of vacuous projections in bare phrase structure. University of Maryland Working Papers in Linguistics 9. 90–115

Heim, Irene & Angelika Kratzer. 1998. Semantics in Generative Grammar.

Ionin, Tania & Ora Matushansky. 2006. The composition of complex cardinals. Journal of Semantics 23(4). 315–360.

Rooth, Mats & Barbara Partee. 1982. Conjunction, type ambiguity, and wide scope. In Daniel P. Flickinger, Marlys Macken & Nancy Wiegand (eds.), Proceedings of the first West Coast conference on formal linguistics, 353–362. Stanford, CA: Stanford University, Department of Linguistics.

Starke, Michal. 2001. Move reduces to merge: A theory of locality. University of Geneva dissertation