LANGUAGE INFINITIES

G F R Ellis and W R Stoeger

Mathematics Department, University of Cape Town.

The ability to use spoken and written language is one of the most significant attributes that separates humans from the rest of the animal kingdom;[1] consequently, it is inevitably the subject of major academic disputes. In particular, there is a disjunction between those who propose a biologically based embodied-language approach, see for example Feldman's book *From Molecule to Metaphor,* [2] and those utilising a much more abstract formal language theory, see for example a recent survey by Komarova and Nowak,[3] hereafter KN.

While the latter approach throws interesting light on some aspects of language (reflected for example in various papers published in *Nature*[4]), one may query whether its formal nature, centering on a proposed algorithmic implementation of a Universal Grammar, adequately reflects the biological underpinnings of the workings of the human mind. We here illustrate this unease by focusing on a strong claim made by KN: namely that there are an infinite number of languages containing an infinite number of sentences (KN, p.320).[5] This is not just proposed as a statement about formally defined languages; it is explicitly claimed that natural languages are infinite (KN, p.323). This is based on the idea that sentences can have an unbounded length: you can just keep adding subsidiary clauses *ad infinitum* to obtain arbitrarily long sentences. But such a ``sentence" is not really a complete thought -- until it is complete. However, it never will be, because that is what infinity means – something that is never completed! Thus, from some common

---

[1] Terrence Deacon (1997): *The Symbolic Species: The co-evolution of language and the human brain* (London: Penguin Books); Derek Bickerton (2001): *Language and Human Behaviour* (Seattle: University of Washington Press).

[2] Jerome A Feldman (2008): *From Molecule to Metaphor: A Neural Theory of Language* (Cambridge, Mass: MIT Press); http://www.icsi.berkeley.edu/NTL/.

[3] N L Kamorova and M A Nowak (2005): "Language, Learning, and Evolution", in *Language Evolution*, Ed. M H Christensen and S Kirby (Oxford: Oxford University Press), pp. 317-337.

[4] M A Nowak, J B Plotkin, and V A A Jansen, *Nature* **404**: 495 (2000); M A Nowak, N L Kamarova, and P Niyogi, *Nature* **417**: 611 (2002).

[5] There is in fact a whole literature on this topic, which one can access by searching for "language is infinite" on Google; however it is convenient to peg the present discussion on the statements in KN.

points of view, it is not really a sentence to begin with. In any case this viewpoint has nothing to do with real-world biologically usable languages, for the simple reason that the human brain has finite storage capacity and only survives for a finite length of time, and so can only carry out a finite number of operations in its lifetime.[6] Hence there is a finite limit to the length of a possible sentence in a biologically realistic language context; while sentences actually usable for communication purposes, the *raison d'etre* of language, are very much shorter

One can calculate an absolute limit on what a human being can possibly read in their lifetime by estimating how many words can be read by a machine doing so continuously for 24 hours a day every day for say 120 years at a rate of say 10 words a second (giving 37,843,200,000 words). This is obviously not infinite. No real person can exceed this limit in their lifetime (inter alia because they have to sleep). But in a realistic approach, one recognizes that if a sentence has meaning, the reader must remember the beginning by the time they reach the end.[7] Given the famous features of short term memory,[8] one can estimate that a maximum length readily understandable sentence has 7 subclauses each with a maximum of 7 words, so 49 words. As sentences become longer than this, they become increasingly un-understandable as a function of the sentence length, and hence unusable for communication purposes; one is probably safe in saying that a sentence more than an order of magnitude larger than this ($10 \times 49 = 490$ words) is of no use for communication purposes and will not occur in any real language except as a possible flowery demonstration of verbal dexterity rather than being intended for genuine communication.[9] Thus the number of practically usable sentences in any real

---

[6] The same kind of restriction applies to any real digital computer; the hypothetical infinite tape of a Turing machine is equally unrealistic. In a sense this embodies a form of the halting problem.

[7] Typical advice to aspiring writers is given in http://www.technical-writing-course.com/essay-writing-length.html: "Make sure you do not allow long sentences — over 40 words — to creep into your writing style. … Faced with long sentences, readers often give up halfway through or forget the start of the sentence by the time they reach the end. The more words in a sentence, the harder it becomes to understand."

[8] Miller, G. A. (1956). "The magical number seven, plus or minus two: Some limits on our capacity for processing information". *Psychological Review*, **63**, 81-97.

[9] For examples of such excesses, see http://en.wikipedia.org/wiki/Longest_English_sentence. In contrast, the present (somewhat lengthy) sentence has 76 words (just compatible with in the more optimistic assessment of 9 phrases of 9 words giving 81words), and the current paragraph has 429 words. The sentence should have been split into two between "purposes" and "one is".

language is finite and they can in principle all be listed, contrary to what is envisaged by KN ("*Natural languages are infinite: it is not possible to imagine a finite list that contains all English sentences*", KN, p.323). This alleged infinity is simply not a realistic description of genuine languages as used for communication by the human species. It is true that sentences will appear in contexts where long range dependencies occur, so meaning is not confined to single sentences. But then the maximum number of sentences that can be contained in a conversation or a book is also finite, for a single author perhaps the equivalent of 30 volumes of 1500 pages each might be a upper limit for their coherent presentation of theories and thoughts over their entire life, so that the maximum number of combinatorial sentences in such presentations is also finite. The possible set of human thoughts expressible in written or spoken language is large but finite. As many have remarked, infinity is not just a very large number: it is a magnitude that is never attained, is always beyond reach. To be real, something has to be completed or have definite boundaries -- but the essence of mathematical infinity is that it's never completed; consequently what is complete in the real world is always finite.

This is of course a rather abstract argument, but in our view this strongly stated use of the concept of infinity in formal language theories definitively demonstrates their lack of realism as compared with embodied-language theories. It indicates that the underlying theory does not take seriously the fundamental purpose and context of language use (communication between living people).[10] Indeed, the great mathematician David Hilbert remarked *"the infinite is nowhere to be found in reality, no matter what experiences, observations, and knowledge are appealed to."* [11] We believe this stricture should be taken seriously, and used to separate fanciful from realistic theories, be they theories of language or of physics.[12]

---

[10] The fundamental definition of a sentence KN employ really does not represent any sentence in any language -- it has neither syntax nor meaning, nor does it need to be complete.

[11] Hilbert D (1964). "On the Infinite". In *Philosophy of Mathematics,* ed. P Benacerraf and H Putnam (Englewood Cliff, N. J.: Prentice Hall), 134.

[12] One could perhaps propose that the theories referred to here don't *really* mean infinity, they really just refer to a very large number. But that would contradict the very purpose of introducing such abstract theories: they are supposed to introduce a precision of thought that was previously lacking.