LACUS FORUM XXXVII

Communication and Cognition: Multidisciplinary Perspectives

McMaster
University

LACUS Forum 37 Referees

Barbara Bacz
David C. Bennett
Lucas van Buuren
Douglas W. Coleman
Kara VanDam

❧

Patrick Duffley
Connie Eble
Sheila Embleton
John Hewson
ShinJa Hwang

❧

Michael Kliffer
Sydney M. Lamb
Alan K. Melby
Robert Orr
Alex Sévigny

❧

H. Stephen Straight
William J. Sullivan
Patricia Casey Sutcliffe
Bernard Sypniewski
Sarah Tsiang

❧

LACUS FORUM XXXVII

COMMUNICATION AND COGNITION: MULTIDISCIPLINARY PERSPECTIVES

EDITED BY

Michael D. Kliffer
Alex Sévigny
Patricia Casey Sutcliffe
William J. Sullivan
Daniel S. Mailman

XIV

~

Presented at
LACUS 35

~

AN INFINITE NUMBER OF SENTENCES

Bernard Sypniewski

Rowan University–Camden Campus

Abstract. The idea that what is called language can create an infinite number of unique sentences is shown to lead to the possibility of sentences of infinite length, requiring an infinite sentence generator hardwired in the brain or pre-specified in the DNA. An examination of these ideas and concepts shows that they have nothing to do with the operation of human beings in the real world.

Keywords: language, natural language, sentence, infinite, communication

Languages: (none)

MANY LINGUISTS CLAIM THAT LANGUAGE CAN CREATE AN INFINITE SET OF SENTENCES USING A FINITE SET OF ELEMENTS, i.e., symbols (Lyons 1992:7) or words. The claim that language can create an infinite set of sentences (IS claim) is not unique to any linguist but has become widespread among modern linguists, e.g., Moulton (1974:6). Still, this claim is not universally accepted for various reasons (cf. Yngve 1996). Though some linguists have challenged the IS claim; these challenges usually have something to do with the mathematics of infinity. In this paper, we will examine the IS claim, assuming that the IS claim is meant to be taken literally. We will assume that when the term infinite is used by linguists in the IS claim they use the term in its mathematical sense to mean an inexhaustible set of sentences; we will also assume that everyone means the same thing when they use terms like sentence. I make these assumptions only for the purposes of discussion. As a Hard Science Linguistics researcher, I do not adopt any of these postulates as my own, however.

1. THE IS CLAIM. Moulton (1974) explains the claim in considerable detail. He does not simply speak of the claim as linguistic gospel use the word infinite. He says, "Human language … can be used to send and receive an unlimited number of messages; there is simply no end to the number of things we can say, in any language. A communications system with this unlimited capacity is ingenious indeed" (Moulton 1974:6). Even though the word used is *unlimited,* the effect is *infinite*, since both words mean *without end*. The IS claim is used as a justification for the claim that language is rule-based, i.e., has a grammar. It is actually the foundational claim for grammar and not just for Moulton. In short, if the IS claim fails, traditional linguistics needs other support for its claim that language needs a fixed grammar. No such support has been forthcoming.

The IS claim is usually examined in reference to two essentially mathematical ideas: 1) natural language combines linguistic elements into sentences, just as a formal language combines its elements into statements in the formal language; and 2) language has a finite or infinite sentence length, depending on the preference of the theorist. Both of these ideas confuse two things. First, the IS claim tacitly says that

natural languages work like the theoretical language discussed in most linguistic theories and not vice versa. Second, the IS claim confuses strings (as understood by computer science and logicians) with sentences (as understood by linguists). Strings are meaningless (in the semantic sense) groups of elements from some character set and exist only in formal languages. The character set that serves as a source of elements for strings need not be the character set of any known human language. It can be a set of symbols that are meaningless outside a particular system. We are accustomed to such sets, e.g., music notation, Morse code, electrical symbols. Strings only exist in formal languages because we have invented them. Strings do not exist independently in the real world.

The IS claim arises because of serious domain confusion inherent in traditional linguistics (Sypniewski 2007). The IS claim confuses artificial languages with people communicating with each other. The IS claim says that language rather than people creates an infinite number of strings but does not actually say that people operating in the real world can create either an infinite number of sentences or infinitely long sentences. The language that the tradition talks about when it talks about the IS claim is not any natural language but a formal language that results from the abstract grammar that is created by traditional theories. The tradition claims that there is no confusion, because natural languages act just like artificial or formal languages and can therefore do what artificial languages do. This is precisely the opposite of the way that artificial languages are designed to work. Artificial or formal languages are simplified languages, in part because we cannot duplicate the complexities of human communication in an artificial scheme.

1.a. RETURN TO STRINGS. A string is an arbitrary collection of elements from some character set, assembled according to the rules of some formal language. The natural language sentence *the cat sat on the mat* is identical in form with the string *the cat sat on the mat* produced in an artificial language. There is no way that we can tell by inspection whether *the cat sat on the mat* comes from a natural or a formal language without knowing its origins. If we accept that the difference between a string and a sentence is, at least in part, the difference between a formal structure that does not have meaning (by definition) and a linguistic structure that must evoke meaning in order to be useful, we can see that the IS claim is not a claim about analysis but, rather, one about production.

In formal languages, the method for creating new strings is by repeated concatenation of elements, whether those elements are from the formal language's character set or other strings (often referred to as *substrings*). Traditional linguists often point to assembling words out of smaller elements, e.g., phonemes, as an analogous feature of language. It does not matter whether a traditional linguist prefers to see that the smallest elements of a natural language are some things on the level of phonemes or words or something else, concatenating those elements in an essentially arbitrary way produces a huge amount of nonsense which must be discarded. This fact is taken

by traditional linguists to support the claim that every natural language must have a grammar to distinguish the good sentences from the bad.

If we use the concepts developed by Chaitin (1998, 2001) to examine the IS claim, we have to confront the notion that any such system must contain a substantial amount of randomness. Indeed, a system that can create an infinite number of sentences or strings can produce at least one sentence or string that is infinite in length or, to put it another way, that will never end, no matter how much time we give the IS generator. When the IS generator begins, we cannot predict whether the sentence or string that it is in the process of producing will be infinite in length. Obviously, if we wait for such a sentence or string to be completed, we will never be able to parse it and determine whether it is a good or bad sentence, because the sentence or string will never be completed. If we parse the sentence while the IS generator is in the process of generating it, we cannot say that the parse will be correct in the end. If and when we apply grammar rules to this production, we cannot say with any certainty that a grammar rule has been correctly applied, because something that is still to be generated might change the application of the rule. Therefore, we can neither completely parse nor apply a set of grammar rules to a sentence or string that is infinitely long. As a result, there is at least one sentence or string that can be produced by language that cannot be determined to be either grammatical or ungrammatical. There is some similarity between this argument and Gödel's Incompleteness Theorem. No traditional linguistic system can produce a complete grammar of any language which claims to incorporate an IS generator, because such a grammar cannot resolve the issue that is the *raison d'etre* of the grammar in all cases.

Any attempt to modify an IS generator to produce large numbers of strings that satisfy the grammar of the language results in a series of circular arguments. Essentially, if language applies grammar rules before or during the creation of a sentence or string using an IS generator, we must also claim that grammar exists separately from and prior to the IS generator. If that is so, there is no need for an IS generator because language could directly create grammatical sentences or strings from the rules of grammar. But then we would encounter another problem: if language only produces good sentences, there is no reason for grammar. If the traditional belief in the need to distinguish between good sentences and bad is the only reason for the existence of a set of grammar rules, and if language cannot produce bad sentences, what does language need a set of grammar rules for? Traditional linguistics sees a grammar as a sort of filter, but if there is nothing to filter out, why have a filter in the first place?

2. The real world. If we consider our daily routine, we do not find that people produce large numbers of meaningless sentences. In fact, doing so might be a cause for medical intervention unless the nonsense was intentionally created for some artistic or other purpose. The qualification about creating art shows us something else about the IS claim: an IS generator operates without intent, a consequence of producing strings rather than sentences. There is no significant evidence that people create enough meaningless sentences to warrant the evolutionary modification of the

human body in order to deal with such a problem. Yet, grammatical rules and the IS generator are supposed to exist on a genetic level which is subject to evolutionary pressures.

We might be tempted to agree with Moulton's claim that we can speak about anything at all. Is this so? I have randomly selected a book called *The Dreams of Reason – The Computer and the Rise of the Sciences of Complexity* by Heinz Pagels from my bookshelf. I randomly opened it to page 201 and randomly selected this sentence:

> *People like to make the brain-computer comparison,*
> *sometimes for fun and sometimes seriously.*

Could every human being who ever lived or might ever live generate this sentence? Note that when we introduce people into the equation, we must talk about sentences rather than strings. For the vast span of human existence, the answer would have to be no. Most trivially, the sentence is written in modern English, though there are non-trivial reasons. If we ask whether most people could generate such a sentence randomly, the answer must be no once again. Not only is it highly unlikely that this sentence would be randomly generated at all by any one person, it is a vanishingly small possibility that it would be randomly generated by everyone. To generate this sentence intentionally requires a person to know the meaning of the sentence, i.e., that the elements of the sentence exist in some lexicon and that they can be combined in precisely this way. Children below a certain age do not have the required skills or knowledge to do so. Intentionally creating such a sentence requires a cultural background that large numbers of people do not have. Most people have no need to generate such a sentence.

Moulton's claim also tacitly assumes that language never changes and that the world is full of Plato's eternal ideas. We know that language does change over time and that ideas are not eternal. It is as unlikely that a 9th century Anglo-Saxon would use the word *computer* (or whatever it might be in Anglo-Saxon, even if it were conceivable) as it is for your barber or hairdresser to talk about *wergeld*. Pagels' sentence is not isolated, as traditional linguistic examples tend to be. It comes from a book and is preceded by over 200 pages of text. For the IS claim to work, each of the preceding sentences must also have been randomly generated. Furthermore, each word in each sentence must have been randomly selected from some collection of words which was assembled from randomly selected elements from some other collection of sub-word elements. To say that this is possible confuses Heinz Pagels with one of the mythic chimps trying to type Shakespeare; indeed, even Shakespeare is confused with one of those chimps.

Probably no one would say that Pagels' book or any other book was assembled with an IS generator. If we pushed back, through short stories and poetry and jokes and captions to everyday discourse on an elevator and beyond, we would eventually have to confront the fact that there is nothing that we could point to that is the product of an IS generator. For those who would say that the IS generator accurately models

something about actual people, there are two questions: what does it model and why do we have it?

3. THE IS MODEL. We cannot say that the IS claim accurately models people's flexibility in communicating with other people, because the IS claim assumes too much randomness and produces too much nonsense. Having a true IS generator would have been very detrimental to our evolutionary development. Warnings of danger need to be accurate, quickly generated and very brief. If our ancestors had the type of IS generator predicted by the IS claim, a person might try to warn others about the approach of a hungry saber tooth tiger only to have their generator start a nonsensical nine million word string. If early homo sapiens did not have the same sort of IS generator that we have, when did we develop ours and, once again, why did we develop it?

Some might be tempted to say that meaningful sentences can be infinitely long or, at least, arbitrarily long. We now confront what we may call the **Very Problem**. Here is a sentence:

I love you very much.

We can easily expand it: *I love you very, very much*. We can even cram in 4500 *verys* or any number of *verys* we wish. The question now becomes: what additional meaning does the 4500th *very* add that the 4499th *very* does not already supply? This question asks several things, actually. Does each *very* have the same quantum of meaning as every other *very*? Does adding a *very* somehow dilute or add to the quantum of meaning of every other *very* and does retracting a *very* somehow increase the quantum of meaning of its fellows or dilute the overall meaning of the sentence? Does adding a *very* do any more than add tedium to the conversation? Why would anyone want to do this? Even if we grant that we can generate an infinite sentence this way for the sake of argument, we, in turn, need to admit that we cannot do this with every collection of words. For example, if we had the sentence:

That color is very blue

We could not generate a meaningful sentence by infinitely repeating the words *that, color, is,* or *blue,* whether we do so by multiplying words or by randomly scattering them throughout the sentence.

Perhaps, one might say, that we can get around these problems by saying that we cannot produce an infinitely long sentence. Let us place an arbitrary limit on sentence length of a million words. To create such an arbitrary limit means that our grammar has a rule which says any sentence that is more than a million words long is ungrammatical. Alas, this will not resolve our problem. Suppose that we created a one million and one word long sentence that, but for the sentence length rule, complied with all other grammatical rules. Is it good or bad? Depending on one's attachment to sentence length rules, one might say bad while another might say good. Who is correct? There are salient arguments either way. The problem cannot be resolved. If we say that the sentence length rules should be modified to say that a sentence is good if it complies with all grammar rules even though it is longer than a million words, we create a meaningless rule because, by definition in a formal language, a

sentence (string) is bad if it violates any rule. If the sentence length rule is to be strictly enforced, we have created a system of grammar which discards otherwise good sentences because of some arbitrary size limit. In short, we have created a system of grammar which does not meet its own objective: to create a system which differentiates good sentences from bad and accepts all and only good sentences.

Ultimately, the IS claim is a claim about time. To say that language can create an infinitely long sentence is to say that language is not constrained by time. This may or may not be true for a theoretical entity like language but it is certainly not true for people. People are constrained very strictly by time. If all we could say is this, it is obvious that any linguistic theory that relies on the existence of an IS generator (to date, all grammatical theories) does not and cannot accurately model the way that people communicate with other people.

4. A FINAL PROBLEM. There is a dispute about whether language can produce an infinite number of sentences. This dispute cannot be resolved because the IS claim is not a claim about the real world. Many, if not most, linguists recognize this. However, the IS claim is the basis for the claim that language has grammar. Language needs grammar to determine whether the products of the IS generator are acceptable or not. There is no reason to believe that when people communicate with other people, an IS generator is involved. If, as HSL researchers believe, none is involved, what then is the basis for the belief that when people communicate with other people, grammar is somehow involved?

5. CONCLUSION. There is no evidence that human beings have, as part of their physical makeup, an IS generator. Being finite, we cannot generate infinitely long sentences. The IS claim arises from the domain confusion inherent in traditional linguistics (confusing people communicating with a hypothetical construct called language). Traditional linguistic theories ground the need for a grammar on the IS claim. Grammar is a system of rules needed to distinguish good from bad sentences (actually strings). If the IS claim actually modeled some aspect of the real world, people could not communicate with other people in a timely fashion, because the IS generator would waste too much time generating nonsense. Occasionally it would get stuck and generate an infinitely long sentence or, at least, one that would take longer to generate than the remaining life span of the unfortunate human involved. Grammar cannot work on an infinitely long sentence. Given the lack of support for the real world existence of IS generators and the questionable theoretical basis for the claim, the claim that humans need grammar or have a grammar encoded in our genes must be viewed with great skepticism. If the claim that we humans need or have grammar must be viewed with skepticism, all other claims which have a grammatical component must also be viewed with skepticism. Traditional linguistics has a significant problem which cannot be ignored any longer.

REFERENCES

CHAITIN, GREGORY. 1998. *The limits of mathematics.* Singapore: Springer Verlag.

——. 2001. *Exploring randomness.* London: Springer Verlag.

HAUGEN, EINAR, ed. 1974. *Language as a human problem.* New York: Norton.

LYONS, JOHN. 1981/1992. *Language and linguistics, an introduction.* Cambridge: Cambridge University Press.

MOULTON, WILLIAM. 1974. *The nature of language.* In Haugen 1974, 3-21.

PAGELS, HEINZ. 1989. *The dreams of reason – The computer and the rise of the sciences of complexity.* New York: Bantam.

SYPNIEWSKI, BERNARD. 2007. *Domain confusion.* LACUS Forum XXXIII.399-404.

YNGVE, VICTOR. 1996. *From grammar to science.* Amsterdam: John Benjamins.