

Applying Logic to the Study of Human Language Syntax

Geoffrey K. Pullum

Brown University and the University of Edinburgh

August 2012

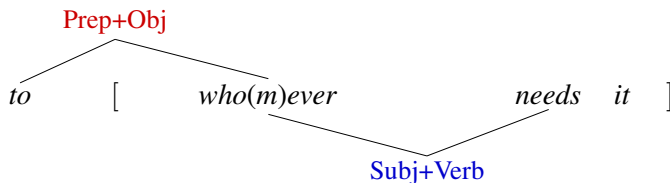
East Asian School in Logic, Language and Computation

Lecture 5: Implications of Model-Theoretic Description

?*I shall give it to whoever needs it.*

?*I shall give it to whomever needs it.*

- A pronoun form must be accusative if it is the head of an NP that is the object of a preposition
- A pronoun form must be nominative if it is the head of an NP that is the subject of a finite verb.



Constraint A: NP determiners are genitive.

Constraint B: Coordinate daughter NPs take the case assigned to the whole NP-coordination.

Constraint C: Genitive case on multi-word full NPs is marked by a suffixed 's.

Constraint D: The suffixed 's on multi-word NPs is encliticized to the last word.

Constraint E: The 's suffix is never encliticized to a personal pronoun that has its own irregular genitive form.

These are not mutually contradictory; but they entail **no coordinate determiner NPs in which the last coordinate is a pronoun with an irregular genitive form.**

Quandaries

- ??He and I's book is coming out soon.
- ??He and me's book is coming out soon.
- ??He and my book is coming out soon.
- ??He and my's book is coming out soon.
- ??Him and I's book is coming out soon.
- ??Him and me's book is coming out soon.
- ??Him and my book is coming out soon.
- ??Him and my's book is coming out soon.
- ??His and I book is coming out soon.
- ??His and I's book is coming out soon.
- ??His and me book is coming out soon.
- ??His and me's book is coming out soon.
- ??His and my book is coming out soon.
- ??His and my's book is coming out soon.

**I think these are George and yours.*

**I think these are George and you's.*

- **Maximality:** Non-heads are maximal.

In FO: $\forall x[\neg(\text{Head}(x)) \rightarrow \text{Maximal}(x)]$

- **Maximality:** Non-heads are maximal.

In FO: $\forall x[\neg(\text{Head}(x)) \rightarrow \text{Maximal}(x)]$

- **Optionality:** Non-heads are optional.

In FO: ? ? ? ? ?

Interstructural constraints

In all cases I know of, interstructural constraints are just a mistake.

- Optionality is universally ignored by X-bar theorists anyway
- There are no valid avoidance-of-ambiguity constraints (Hankamer, responded to by Langendoen)
- There are no exceptionless blocking constraints
- The “economy” constraints of minimalism are misconceived (see Johnson & Lappin 1999)...
... and so on

MTS makes it particularly explicit that there cannot be interstructural constraints, and that seems like a good result.

Strange recent remarks by linguists (1)

“Infinity is one of the most fundamental properties of human languages, maybe the most fundamental one. People debate what the true universals of language are, but indisputably, infinity is central.”

(Howard Lasnik, 2000)

Strange recent remarks by linguists (3)

“This property of discrete infinity characterizes EVERY human language; none consists of a finite set of sentences. The unchanged central goal of linguistic theory over the last fifty years has been and remains to give a precise, formal characterization of this property and then to explain how humans develop (or grow) and use discretely infinite linguistic systems.”

(Sam Epstein and Norbert Hornstein, 2005)

Strange recent remarks by linguists (3)

“[M]any have argued that the property of recursive infinity is perhaps *the* defining feature of our gift for language.”

(Charles Yang, 2006)

The supposed inductive argument for the claim that English has infinitely many grammatical expressions:

- *very nice* is grammatical
- adding one *very* makes *very very nice*, which is grammatical
- adding another *very* makes *very very very nice*, which is grammatical.
- ... (and so on) ...
- So by induction, for every natural number n , adding one extra *very* to *veryⁿ nice* makes an expression *veryⁿ⁺¹ nice* which is also grammatical.

Infinite

But “for every natural number n ” gives the game away: the question has been begged.

The decision that induction on the natural numbers can be used in this domain has ALREADY PRESUPPOSED that the domain is infinite.

On domains where we know the infinitude conclusion cannot be correct, we simply reject the appropriateness of the reasoning.

For example . . .

A stupid argument in human biology:

- 1 year is a biologically possible age for humans.
- Adding one year of life to a human of age 1 gives an age of 2, which is also biologically possible.
- Adding one further year of life gives an age of 3, which is also biologically possible.
- ... (and so on) ...
- So by induction, for every natural number n , adding one extra year of life to a human of age n gives an age of $n + 1$, which is also biologically possible.

(Conclusion false because of the Hayflick limit.)

A stupid argument in evolutionary biology:

- This organism is of the species *Canis lupus familiaris*.
- Its female ancestor one generation back was a female organism also of the species *Canis lupus familiaris*.
- Its female ancestor one generation before that was a female organism also of the species *Canis lupus familiaris*.
- ... (and so on) ...
- So by induction, for every natural number n , at n generations back its ancestor $n + 1$ generations back was a female organism also of the species *Canis lupus familiaris*.

(Conclusion false because dogs were only domesticated from the gray wolf about 15,000 years ago.)

We have to ask how we know that the argument used for the claim that English has infinitely many sentences is a sensible one, not one of the many stupid ones.

We need grounds for claiming (a):

- (a) extension in sentence length and complexity goes on forever without altering grammaticality

rather than claiming (b):

- (b) extension in sentence length tapers off gradually and ceases to preserve grammaticality after some (rather vaguely defined) point is reached.

We don't have any such non-question-begging grounds.

Even in pure mathematics we know of cases where a long succession of cases where some claim is true can be followed by infinitely many more where it is false.

Take the prime-counting function $\pi(x)$ and the logarithmic integral function $\text{li}(x)$.

It has been shown computationally by Kotnik (2008) that there are no values of x below 10^{14} for which $\pi(x) > \text{li}(x)$.

Yet Stanley Skewes proved long ago that eventually there are values of x where $\pi(x) > \text{li}(x)$ (in fact there are infinitely many crossing points).

So where are the calculations by linguists on the matter of maximum expression complexity? There aren't any.

Or rather, when evidence is gathered or calculations are done, linguists tend to ignore both.

- Fred Karlsson searched carefully for sentences with significant depths of initial or center-embedding, and found hardly anything.

But linguists continue to believe what they believed before: that initial embedding and center-embedding to any degree are grammatical and the set of sentences exhibiting them is **infinite**.

- András Kornai did some statistical analysis on the frequencies of attested words and showed that the data clearly have the profile you would expect from an **infinite** population of words.

But linguists continue to believe what they believed before: that the set of words is **finite**.

What consequences flow from the supposed infinite number of sentences in human languages?

None.

- Nothing follows about use of the language
- No theoretical claims build interestingly upon it
- No evidence directly confirms it.
- No evidence refutes it, or ever could.

Only one suggestion has much plausibility.

A generative grammar for a large finite set of expressions is very tedious to construct. Walter Savitch has shown that infinitely many finite stringsets have infinite extensions with exponentially shorter grammars.

Recursive rule application is the obvious solution to many descriptive problems. And where there is non-trivial recursive rule application, a generative grammar will generate infinitely many strings (the cases where this does not happen can be regarded as somewhat pathological).

If we assume linguists have mistaken the effects of their descriptive technology for a property of their subject matter, we have an explanation for their otherwise strange infatuation with infinitude.

If they are not simply being misled by generative grammars, we need to ask why linguists cling to the belief that human languages have infinitely many expressions when

- (i) it may well be false of some languages (e.g., Pirahã), and
- (ii) it is empirically unsupported and unsupportable even for English, and
- (iii) if true it would make no difference.

They may feel infinitude is closely tied to the **creativity** of language use: People make up, utter, and understand sentences that have never been encountered before.

But connecting creativity to infinity is a mistake. Think of (i) chess, (ii) bridge, or (iii) composing sonnets or *haiku*.

The connection to an implication of model-theoretic syntax is very straightforward.

How many graphs are there that satisfy the transitivity condition $(\forall x, y, z)[E(x, y) \wedge E(y, z) \rightarrow E(x, z)]$?

As many as you want to say there are. Given a finite class of finite candidate models (say, the set of graphs representing sets of human beings who know each other), it is some finite number.

Given the class of all finite graphs as candidates, is countably infinite (though vanishingly small asymptotically as a proportion: as larger and larger randomly constructed graphs are considered, the probability of a graph satisfying transitivity falls away to become zero in the limit).

Just so with linguistic expressions. If English has just finitely many expressions and they are of finite size, then only finitely many structures will satisfy the grammar.

If there is no limitation to a finite number, then perhaps infinitely many satisfy the grammar.

The rules of the grammar, the syntactic constraints, will be the same in either case.

We do not need to stipulate an answer: *we can describe syntactic structure in a way that does not entail any commitment regarding how many expressions exhibit that structure.*

Conclusions

- The past 50 years of syntactic theory have been almost entirely preoccupied with GES

Conclusions

- The past 50 years of syntactic theory have been almost entirely preoccupied with GES
- The procedural metaphors that GES naturally engenders and encourages are rife among syntacticians

Conclusions

- The past 50 years of syntactic theory have been almost entirely preoccupied with GES
- The procedural metaphors that GES naturally engenders and encourages are rife among syntacticians
- The mathematical apparatus is available for full and thoroughgoing development of a model-theoretic alternative

Conclusions

- The past 50 years of syntactic theory have been almost entirely preoccupied with GES
- The procedural metaphors that GES naturally engenders and encourages are rife among syntacticians
- The mathematical apparatus is available for full and thoroughgoing development of a model-theoretic alternative
- At least half a dozen phenomena and several conceptual clarifications suggest that an MTS mode of formalization might be better for theorizing about human languages (as opposed to artificial ones, for which GES is excellent).

THANK YOU!

It has been a privilege to be here
in China to give these lectures.

Geoffrey K. Pullum • **gpullum@ling.ed.ac.uk** or **pullum@gmail.com**
<http://ling.ed.ac.uk/~gpullum>

Department of Cognitive, Linguistic,
and Psychological Sciences
Brown University
190 Thayer Street
Providence, RI 02912

School of Philosophy, Psychology,
and Language Sciences
University of Edinburgh
3 Charles Street
Edinburgh EH8 9AD, UK