Re: Kolmogorov complexity and logical languages

Source: http://sci.tech−archive.net/Archive/sci.lang/2009−05/msg00360.html

- *From*: linguist.in.hiding@xxxxxxxxx

- *Date*: Mon, 18 May 2009 11:10:04 −0700 (PDT)

I have looked into the Kolmogorov thingie a bit. So, from now on I'll say something about complexity in languages (I know, it is in the FAQ). If we don't take pidgins and language death into consideration, it is hard to say that one language is more complex than another. Some quotes follow.

First a search of sci.lang gives 27 (actually 26) occurences of 'Kolmogorov' (omits this thread, for now). Mostly it has nothing to do with linguistics.

1 from analys...@xxxxxxxxxxx

8 from Richard Herring:

−−

> Oliver Cromm wrote:
>
> > Quoth Jouni Filip Maho:
> >
> > > I just wanted to point out that "equally functional" isn't the same thing as "equally complex", even though that erroneous equation underlies many uses of the equally−complex assertion.
> >
> > That depends on your definition of complexity. The definition I am used to, maybe owing to my mathematical background, is something like "expressive power", and a value of complexity could be "context free".
> >
> > A lot of people here seem to speak about complicatedness (involvedness)

> instead of the more abstract notion of complexity.

> It's hard to find definitions of 'complexity' in linguistics, but normally it hasn't (or isn't supposed to have) any direct connection to expressiveness.

> It has more to do with the formal set−up/organisation of the "rule system". The more rules you need to describe the grammar/phonology/pragmatics of a language, and the more exceptions you need to establish to those rules, the more complex the language is. Something like that.

> I'm sure, though, that if you ask 100 linguists you may get 100 slightly different answers, but they would all (or nearly all) be saying something about the system of "rules".

That sounds like Kolmogorov complexity. Something like the size of the shortest description which encapsulates all salient features of the system.

Deciding what is a salient feature is left as an exercise for the reader...

−−

2 from Hans Aberg

1 from izzy (not relevant)

2 from Marc Adler (not relevant)

2 from Jouni Filip Maho (not relevant)

2 from LEE Sau Dan

4 from Yusuf B Gursey (not relevant)

4 from H.M. Hubey (what needs to be said, you'll see)

I went a bit deeper, and got the following quotations:

> From http://groups.google.fi/group/sci.lang/msg/3581ff4d5f4c3d8e

−−

> Laws of probability theory take precedence over simple
> heuristic rules of 19th century historical linguists.

And evidence takes precedence over probability theory. But even without the evidence it should be obvious that the cap problem is a poor model of the linguistic reality.

Brian M. Scott

−−

Hubey said in http://groups.google.fi/group/sci.lang/msg/3a3b69ac016327ea

−−

There are different methods of solving equations of
this type, such as mean square methods and Fokker−Planck−Kolmogorov methods. There is no reason why these methods cannot be used in linguistics.

−−

We then have Herman Rubin in the same thread
http://groups.google.fi/group/sci.lang/msg/3bacbb5e3bf8591d

−−

Statements such as language is a stochastic process are useless. All this
means is that there is some joint probability distrubution over all utterences, taking into account alse the times of those utterances. It
would provide no information whatever to say that all observations by any
observer form a stochastic process.

It is only when restrictions are put on the process that there is any content to the formulation. To say that the process is a process of independent random variables, or a Markov process of a given order, provides a restriction which presumably can be tested.

However, any assumptions about the process should be made on linguistic
grounds, not mathematical convenience.
−−
Herman Rubin, Dept. of Statistics, Purdue Univ., West Lafayette

IN47907−1399
Phone: (317)494−6054
hru...@xxxxxxxxxxxxxxxxxx (Internet, bitnet)
{purdue,pur−ee}!pop.stat!hrubin(UUCP)

−−

Paul L. Allen years later says http://groups.google.fi/group/sci.lang/msg/c77ccdcd9dfe5e95

−−

> One way to deal with this is via the Fokker−Planck−Kolmogorov methods.
> ONe obtains a partial differential equation for the probability density
> of the process. IT is of the diffusive type so one can easily imagine
> this process taking place in a high dimensional space, resembling a kind
> of fluid flow.

You can imagine all you like. But to get anyone here to believe you,
you'll first have to show that your model is both credible and matches
known data. It looks very much to me like you've invented a model and
now you're busy trying to force the data to fit. In fact, you appear
to be an archetypal net kook.

−−

To have Hubey ranting (about Fokker−Plank−Kolmogorov equations) got my
alarm bell ringing. The thing is that the mathemathical models
proposed are hopelessly inadequate for real life. So, as the
frustrated Mikael Thompson stated clearly in
http://groups.google.fi/group/sci.lang/msg/e3f40283e0885573

−−

I must say, you are one of the most arrogant, ignorant posters on this
group. Now then, before your Highness gets all riled up at yet
another
person so very much more ignorant than he is, I'll list my
qualifications: I studied physics at Princeton for two years; before
then I was a research assistant (non−linear dynamics) in the physics
department at TCU (yes, during high school); I learned calculus my
freshman year in high school and taught myself ordinary differential
equations the summer of my sophomore year in high school.
Unfortunately, or to my way of thinking, fortunately, I ran out of
money
and had to work for five years, during which time I discovered I love
linguistics and history even more than I do physics. I have a pretty
good grasp of mathematics then, and all I have to say on that score is
that your mathematical model is primitive in the extreme and doesn't
fit

the facts that have been presented to you, and all you do is sit there with the typical snobbery of a second−rate mathematician who at least can take satisfaction in thinking he knows more math than those dumb−ass humanities drudges. People have pointed out to you many times that you don't know very much at all about linguistics, and certainly not historical linguistics: You sneer at the comparative method as mere heuristics, you attack the idea of regular sound change, you don't know the first thing about the processes of language change (witness your question as to why a common word would have five syllables, your question why words don't all just erode away through sound change, and so on), etc.; and when you do ask those questions, your obvious desire to trip other people up with stupid judo arguments peeks through your equally obvious misunderstandings of the fragments of linguistics books that you've read. And then you have the gall to sit there insulting us because *your* model is irrelevant to the real world! I refuse to give obeissance to you, as you so obviously wish, as the Mathematician−Savior come to redeem the infidel linguists still dwelling in the darkness. If the linguists on this newsgroup were as woefully ignorant of mathematics as you are of linguistics, they'd be broke and on the streets.
So tell me, when you get in controversies in mathematics, do you abuse your opponents with streams of feculent discharge like this bundle of whines below, or do you reserve that for those people that mathematics types like yourself are acculturated to sneer at−−since, by the very fact of going into the humanities, they are obviously second−rate (or worse) mediocrities who just couldn't cut it? Your previously fastidiously−concealed disdain is blatantly obvious now.

Mikael Thompson
H. M. Hubey wrote:

> Ross Clark <d...@xxxxxxxxxxxxxxxxxxxxxx> writes:
>
> > No Mark, you don't get it. The task of generalizing this to make it
> > universally applicable is one _you_ have set yourself. The "results I
> > obtained" were done by exactly the same two−line exercise you went

Look, it is you who does not get it yet.

You remind me of the type of people who sit around all day long, cannot accomplish anything, are not good at anything, but then they are always around to criticize the world around them.

Just produce your solutions. I am getting bored.

I can already see that you have never solved a problem in your life; not a freshman physics problem, not a junior level engineering problem, not a computer programming problem, not a probability problem, and you cannot even recognize solutions when you see them.

do you have any idea of the math and simplification that goes into the solution of everyday problems from the car you drive every day to the computer you use, to the TV you watch and a myriad of other things?

I bet you have no clue. If the people who have produced all of these things waited for the perfect solution to hit them and sat on their butts all day criticizing things they cannot even comprehend, you'd be behind a plow pulled by a couple of oxen.

And so coarse too. The first thing that comes out is "Mark, you still don't understand."

What makes you think that there is anything in this world that someone of your incompetence can understand that I cannot? Do you think that I do not get students in my classes who know about as much math as you?

> Do you have any idea how many of them I have seen since
> 1983?
>
> What is your big problem? I know that there are many
> things I don't know. I know that there are many fields of
> math, QM, biochemistry etc that I do not know. But I do
> know what I know, and I do know what I do not know. That
> is the mark of an expert.
>
> That is why I can recognize people who do not know
> what they do not know, and do not have enough sense
> to stop.
>
> SHOW US YOUR SOLUTION. FESS UP. IT'S SIMPLE ENOUGH.
>
> YOU DO NOT LIKE MY SOLUTION. GIVE US YOURS. I AM
> HAPPILY AWAITING IT.

––

In one post Brian M. Scott sums up my view on the issue (although I am not "willing to accept the possibility that there is a meaningful sense in which one language is more complex than another" for many reasons) http://groups.google.fi/group/sci.lang/msg/8cb65e9244bbabba

––

On 3 May 2003 03:14:41 GMT, mr...@xxxxxxxxxxx (Mark J. Reed) wrote:

> In article <jbysxveflzcngvpbpna.heai6c0.pmin...@xxxxxxxxxxxxxxxxxx>,
> Wolf Kirchmeir <wwolf...@xxxxxxxxxxxxx> wrote:
>
> > The trouble with this definition is that it equates inflection with
> > complexity.
>
> Yes, it does, but I don't see that as trouble.

That appears to be because you simply don't recognize the complexity inherent in other parts of a language.

> > Why should the presence of inflections make a language more
> > complex than their absence?
>
> Well, by definition. :) Also, note that I'm not equating complexity
> with difficulty, which is far more subjective.

You haven't answered the question.

> > From what I remember of my middle school Latin,
> > inflections hugely simplified syntax −− I could stick an adjective in almost
> > anywhere − I didn't have to put them in front, as in English (and in a fixed
> > order, too!), or behind, as in French (except when they were put in front, in
> > which case they meant something else, and you'd better know which
> > adjectives
> > you could put in front and which ones you couldn't.) Not simple at all!
>
> I would not say that English syntax is complicated by the
> lack of inflections. Its word order is less flexible, but I
> would say that makes it simpler, not more complex.

The information conveyed by inflexions in, say, Latin or Old Norse still (by and large) has to be conveyed in English, for instance by word order and prepositional usage. The complexity is simply transferred from one part of the language to another.

[...]

> > The fact that English syntax is not as simple as it looks is proven by the
> > fact that no elementary (=school) text I've have _ever_ seen describes it as
> > it is actually done.
>
> Okay, so the texts are prescriptive rather than descriptive; that doesn't
> imply that either the "official" or the "actual" grammar is
> particularly complex.

It isn't a matter of prescription versus description; they don't describe what is actually done even in prescriptively correct usage.

> > But I didn't say English was simple, either,
> > just that it's simpler than some other languages, more complex than others,
> > and that it's not that hard to tell which is which.

I am willing to accept the possibility that there is a meaningful
sense in which one language is more complex than another, but it
will have to involve the entire language, not just the
morphology.

Brian

−−

Brian M. Scott continues http://groups.google.fi/group/sci.lang/msg/73f9ec3cf265638f

−−

On 3 May 2003 14:58:17 GMT, mr...@xxxxxxxxxxxx (Mark J. Reed)
wrote:

> WK = Wolf Kirchmeir <wwolf...@xxxxxxxxxxxxx>
> MJR = Mark J. Reed <mr...@xxxxxxxxxxxx> (Me)
> BMS = Brian M. Scott <b.sc...@xxxxxxxxxxx>
> WK> The trouble with this definition [more inflections = more complexity]
> WK> is that it equates inflection with complexity.
> MJR> Yes, it does, but I don't see that as trouble.
> BMS> That appears to be because you simply don't recognize the
> BMS> complexity inherent in other parts of a language.
> BMS> The information conveyed by inflexions in, say, Latin or Old
> BMS> Norse still (by and large) has to be conveyed in English, for
> BMS> instance by word order and prepositional usage. The complexity
> BMS> is simply transferred from one part of the language to another.
> Yes, but rearranging words does not alter their complexity.

So what? If you want to measure the complexity of a language,
you can't limit yourself to single words; you must at the very
least consider sentences. Besides, it's extremely difficult to
come up with a definition of 'word' that makes sense
cross−linguistically. Is English 'pull off' (as in 'Can he pull
it off?') one word or two? Does it really make sense to say that
German <Ringfinger> 'ring finger' is one word, while its almost
identical English translation is two? Is French <je le vois> one
word or three? What about polysynthetic languages?

> The individual words are still readily recognizable in their new
> location, without having to be decoded from whatever inflected form
> they happen to take. With prepositional phrases, each
> word has only one form and each preposition has one form.
> This is mathematically less complex than the situation with inflections.

That is not at all clear.

> BMS> I am willing to accept the possibility that there is a meaningful
> BMS> sense in which one language is more complex than another, but it
> BMS> will have to involve the entire language, not just the
> BMS> morphology.
> Fair enough. I don't think there's any argument that Latin is more
> complex morphologically than English, so let's look at some of the ways
> English might be considered more complex than Latin:
> Articles − English has them, Latin doesn't, and it's very difficult to
> explain when their use is called for.

Indeed; so difficult that no one's done it yet.

> Multiple forms of each tense − English makes heavy use of the progressive
> forms in lieu of the simple ones, and again it is difficult to explain
> exactly when. On the other hand, English doesn't usually bother with the
> indicative/subjunctive distinction.
> Orthography − English spelling and pronunciation are at first glance
> somewhat abitrary; even when you learn the rules, they are complex and
> full of exceptions. Syllable breaks are difficult to identify, and
> even when you can identify them, the emphasis is not easily predictable.

Complexity of writing system is independent of complexity of language; after all, most languages have never been written at all, and some have been written in multiple systems. Stress assignment, on the other hand, is part of the language proper, and its irregularity does indeed add complexity.

> What else?

English word order rules. Even ordering a string of adjectives correctly is non−trivial, however natural it may seem to one who's grown up with it. Possibly rules of usage peculiar to a few lexical items in a given category; I don't know how common this is in Latin. (E.g., the 'he is to blame' construction under discussion elsewhere and the unacceptability in most varieties of 'he might could do it'.) Proper use of the auxiliary 'do'.
Phrasal verbs that despite appearances are distinct lexical items (e.g., 'to run out' of something).

> Is there enough complexity here to balance out the
> extra complexity in Latin morphology?

Brian

−−

I don't know if I really need to add more after the next comment on
the same thread by Jukka K. Korpela http://groups.google.fi/group/sci.lang/msg/035275c36d4ebff1

−−

> There is no doubt that one language
> may have greater overall grammatical complexity and/or a
> communicative advantage in a certain sphere, over another. But
> this is a topic for a separate book.

Let's not hold our breath.

Surely there are _some_ ways in which the complexity of a language can
be defined in an objective and even measurable manner. For example,
the
number of cases is a measure, and so is the number of essentially
different meanings that word order can express, and the number of
different phonemes, etc. But it would be worse than futile to study
such issues if the real goal is to declare some languages as more
complex, more advanced, more communicative, etc., than others. Even if
we limit ourselves to mere complexity (and why would _that_ be
interesting, really?), an overall complexity would be just a weighed
sum of individual complexities − and the results would tell more about
the opinions of people who set the weights than anything else.

−−

With friends like Hubey defending the Kolmogorov thingie... Just to be
precise, the Kolmogorov thingie has been used in linguistics.

http://maxbane.com/wp−content/uploads/2007/09/wccfl−slides.pdf

"Quantifying and Measuring Morphological Complexity"

I just had the feeling that the person that wrote this is a Hubey
reincarnate...

I quote:

−−

General method: Count the occurrences of a variety of
hand−picked, intuitively justified properties of the linguistic
system.

Phonological Complexity
Size of phoneme/syllable inventory.
Number of marked phonemes.
Number of rules/alternations.
Morphological Complexity
Number of possible inflection points in a typical sentence.
Number of inflectional categories, morpheme types.
AUTOTYP synthesis (Bickel & Nichols 2005).
Syntactic Complexity
Number of parameters deviating from default.

−−

"the results would tell more about the opinions of people who set the weights than anything else." Indeed.

So we have a complexity figure of 35.51 % for Latin, 19.51 % for Dutch, 16.88 % for English, 0.05 % for Vietnamese and so on. Now, this indeed describes even the intuitive sense of complex vrs simple morphology.

Sorry for not being impressed but how has the assigning of percent values with the use of the Kolmogorov thingie advanced our knowledge of linguistics... You could as well have measured them the old−fashioned way like counting the mean of affixes to a "word" in a text (of course there are difficulties in that too). I doubt the order of the languages would have changed considerably. And even if it did... So what? I fail to see that that method is intrinsically superior.

As a concluding remark. Let me just quote Kolmogorov himself, W. Andries van Helden Case and Gender. Concept Formation between Morphology and Syntax I, II, in

http://www.slavistiek.nl/ssgl/ssgl20−21.htm

−−

In 1957 the mathematician Kolmogorov confronted the participants of a seminar on mathematical linguistics with a few pilot questions, such as "what exactly do we mean when we say that two words are in the same case?" The rigorous answers which the Set−theoretical School worked out for Kolmogorov's questions turned out to have far−reaching implications for linguistic theory.

−−

Has that really had "far−reaching implications for linguistic theory"? I myself am skeptical. Besides, it is quite rich to assume, as implied here, that linguists had not thought about that particular question. No satisfactory answer exists to the question, anyway. To think that to have an answer (of course there are answers to the question, but

not quite rigorous for mathematics) to that question (in no time)
shows that Kolmogorov himself was quite naive about linguistics.
.