Is Language Infinite?

This question has been discussed in several places, with both mathematical and linguistic aspects to the debate.

Classic textbook discussions are similar to what is found in your textbook.

Standard (Generative Grammar) Assumptions: (about any natural language)

1. The length of any sentence is finite. (whether letters, phonemes, morphemes, or words.)
2. There is no longest sentence. (because of recursion)

From these two assumptions it follows (as in our textbook) that the cardinality of the set of sentences in any natural language is aleph-null: the set of sentences is denumerably infinite.

- Note: Sentence length, on the assumptions above, is unbounded, but not infinite.

Debates:

1. Is language finite? Some claim that the cardinality of the set of sentences in any natural language is ***finite***. See, for instance, the posting in Linguist List 2.551 by Jacques Guy, which touched off an extended discussion/debate in September 1991.

Go to http://www.linguistlist.org/issues/2/2-551.html

To see the full debate, run a Search within Linguist List on "Is Language Finite". That will get you all of the issues which contain contributions to the debate except for the first two, which are the link above plus: http://www.linguistlist.org/issues/2/2-554.html .

I found another site, http://www.usingenglish.com , where there was a more recent debate on the same topic, and where there was even a poll done on the question! And in that poll, there were 17 votes: 2 Yes, 15 No. There's no way to know how many opposed the "standard assumptions" and how many questioned the math. Comments on both sites suggest that there are some of each.

2. Is language non-denumerably infinite? This claim was made by Paul Postal and D. Terence Langendoen in their book, Langendoen, D. T. and P. M. Postal (1986). The Vastness of Natural Languages. Oxford, Basil Blackwell.

They challenged the first *assumption*, and suggested that an optimal grammar makes no constraint on the length of sentences, not even a requirement that each sentence be finite. And they assumed that conjunction can conjoin any number of clauses, finite or infinite.

(Something of that sort is required to generate non-denumerably infinitely many sentences.)

The Langendoen and Postal position is also debated online, although few appear to be up to the mathematics involved, and in fact among the debaters one frequently finds confusion between the claim that the length of sentences in unbounded (but finite) and the claim that there can be sentences that are infinitely long. The first is the standard assumption, the second is Langendoen and Postal's. For some discussion of the Langendoen and Postal claim see the posting by Terry Langendoen in http://www.linguistlist.org/issues/2/2-602.html , the posting by Alexis Manaster-Ramer in http://www.linguistlist.org/issues/2/2-619.html, and the posting by Avery Andrews in http://www.linguistlist.org/issues/2/2-634.html .

I will append a sample of excerpts from the discussions for us to start our own in-class debate. It's interesting and useful to see what the issues are, and it's really important to see how the argument breaks down when people don't have clear ideas about the mathematics of infinite sets.

From Linguist List 2.551

Message 3: Compositionality: the smallness of natural language

Date: Mon, 23 Sep 91 12:12:05 EST
From: Jacques Guy <j.guy@trl.oz.au>
Subject: Compositionality: the smallness of natural language

```
Alexis Manaster Ramer <USERGDD8@WAYNEMTS.BITNET> wrote about
compositionality: "...whereas the compositional rules generate an
infinite set of expressions."
No. The set of possible expressions, utterances, what-have-you can be
infinite
      (1) if and only if the number of language elements (phonemes, or
      morphemes or whatever, depending at what level you look at it) is
      infinite,
          OR
      (2) if and only if there exist utterances of infinite length.

Both conditions are contrary to fact. This point, however trivial, must
be worth making, seeing that Langendoen and Postal argued in a book
titled "The Vastness of Natural Language" that the cardinality of the set
of utterances was not only infinite, but greater than aleph-null, aleph-
one, aleph-two, in fact, if I remember correctly, greater than any
conceivable transfinite number. It is, in fact, not only infinitely
smaller than aleph-null, but very very much smaller than one googolplex
(but it is perhaps greater than one googol).
```

Question for you -- Is he right about the mathematics? (a) Would having an infinite set of words in the language guarantee the existence of an infinite set of sentences? (b) If there are finitely many basic elements (e.g., words), then do there have to be sentences of infinite length in order for there to be infinitely many sentences?

See all the replies to him in issue 2.554!

Tom Lai, and Tom Lai corrected, in issue 2.582:

Message 3: Language is infinite

Date: Sat, 28 Sep 91 00:14 +8
From: Tom Lai <ALTOMLAI@CPHKVX.BITNET>
Subject: Language is infinite

```
Language is infinite. This is a scientific truth. Jacques Guy is using
the word _infinite_ in a sense that no scientist (at least mathematicians
and computer scientists) should (excuse me for using this word, but I
can't say _will_) use this word in. Jacques Guy's reckoning of the length
of possible utterances and his talking program show exactly that the no
finite maximum exists for the possible length of an utterance. In
scientific jargon, the length of the longest possible (theoretically
speaking) utterance is then at least _countably_ infinite. In fact, this
should be _uncountably_ infinite, but I am not going into details as it
is infinite anyway.
Tom Lai.
```

Message 5: Is language countable?

Date: Sun, 29 Sep 91 00:24 +8
From: Tom Lai <ALTOMLAI@CPHKVX.BITNET>
Subject: Is language countable?

```
The set of all possible utterances in a language is infinite. In earlier
postings I said the cardinality of this set is uncountable. It seems that
I was wrong. The cardinality of the set of all possible utterances in a
language is countable. This should follow from the observation that for
any positive integer n, the number of different utterances of length n is
finite. This is no big deal. But I would appreciate confirmation by
people out there who are familiar with these things.
Tom Lai.
```

Alexis Manaster-Ramer on the question of whether this debate is an empirical one, in Linguist List 2.595.

Message 1: 2.582 Is Linguist Infinite?

Date: Sun, 29 Sep 91 11:54:14 EDT
From: <Alexis_Manaster_Ramer@MTS.CC.WAYNE.EDU>
Subject: 2.582 Is Linguist Infinite?

```
Michael Kac writes:
        Anyone interested in this question might find it worthwhile to
look at  the paper by Rounds et al. in *Mathematics of Language*, ed. A.
```

Manaster- Ramer (Benjamins, 1987), particularly the last para. of sec. 4 (p. 354).

As one of the authors of the article, I followed the instructions and read the last para. of sec. 4. It turns out that we pointed out that it is indeed reasonable to take languages as finite. However, this only makes sense if we treat them not as fixed finite sets but as families of finite sets with no fixed bound on size. As far as I know, Yuri Gurevich at U. of Michigan and some other people have developed a formal model of this sort to serve as a model of the behavior of computers. But I should caution that not much rides on the distinction between infinite sets and such families of finite sets, although there are some arguments for the latter view as more realistic (both for computers and for human beings).

I would also add that I am surprised by the vehemence with which several correspondence assert as fact or as scientific truth something that I would regard as a matter for indirect argument at best and perhaps only of mathematical convenience. The same applies, of course, the Langendoen/Postal work: I still cannot believe that Terry and Paul could seriously claim that the issue of whether NL sentences are merely of unbounded length or of infinite length (and whether the collection of English sentences was countable or not) a factual question. But by the same token I cannot understand the self-righteous dismissal of their proposals by those who somehow possess the certitude (that I so notably lack) that NL sentences are only of finite length.

The same applies to the question of whether there are any truly analog phenomena in nature. I do not feel at all certain that nature is ultimately digital as some seem to.

Finally, in response to Tom Lai's query, the following is mathematically correct (this we can be certain of): If we assume that there is no upper bound on the length of English sentences, then there is an infinite (but only countable) set of these, since they can then be placed in one-to-one correspondence with the natural numbers. If we assume that English sentences can be of actually infinite length, then (depending on what else we assume), they may form either a countable or an uncountable set (it is crucial that we assume something more, because you can have a finite (even a singleton) set of infinite-length sentences).

Q and A about Langendoen and Postal.

Q from Macrakis in 2.595:

Message 4: 2.582 Is Linguist Infinite?

Date: Mon, 30 Sep 91 14:04:18 EDT
From: <macrakis@osf.org>
Subject: 2.582 Is Linguist Infinite?

Kac says:

> Langendoen and Postal's *The Vastness of Natural Languages* ... argue that NL's are NONDENUMERABLY infinite -- indeed, that they are maximally so (that is, that the number of sentences in a NL is greater than any cardinal number).

Assuming finite-length sentences and finite numbers of combining elements (whether phonemes or something else), set theory won't get you anywhere beyond countable infinity. How do they get anything

```
bigger?
By the way, although mathematics talks <<about>> vast cardinals, only
a countable number of mathematical objects can ever be individually
named.  Is this the paradox they play on, perhaps?
        -s
*************************
```

Answer from Michael Kac in 2. 602:

Message 1: Re: 2.595 Is Language Infinite?

Date: Mon, 30 Sep 91 17:36:29 -0500
From: "Michael Kac" <kac@cs.umn.edu>
Subject: Re: 2.595 Is Language Infinite?

```
Answer to Linguist subscriber Macrakis on how Langendoen and Postal get
uncountably infinite languages: the answer is that they argue, contrary
to the standard position, that there are sentences of infinite length.
(See Chapter 3 of their book.)
        Alexis Manaster-Ramer makes a couple of comments in his most
recent posting that I want to strongly second. The first is that the
question of finiteness/nonfiniteness of languages is not a factual one;
the second is his expression of discomfort with the cavalier way in which
the Langendoen-Postal thesis is rejected by most linguists. I'm reluctant
to take up a lot more space with my endless 2 cents worth but encourage
Linguist subscribers who are interested in the issue to correspond with
me personally. My address is kac@cs.umnedu.
Michael Kac
```

Langendoen replies to Manaster-Ramer in 2.602:

Message 3: is language infinite?

Date: Mon, 30 Sep 91 17:02:17 MST
From: Terry Langendoen <LANGENDT@ARIZVM1.CCIT.ARIZONA.EDU>
Subject: is language infinite?

```
In a recent LINGUIST posting, Alexis Manaster-Ramer writes:
         The same applies, of course, the Langendoen/Postal work: I still
      cannot believe that Terry and Paul could seriously claim that the
      issue of whether NL sentences are merely of unbounded length or of
      infinite   length (and whether the collection of English sentences
      was countable or not) a factual question.  But by the same token I
      cannot understand the self-righteous dismissal of their proposals
      by those who somehow possess the certitude (that I so notably lack)
      that NL sentences are only of finite length.
I cannot speak for Paul, but I agree with Alexis that the question of the
size of natural languages is a theoretical one.  As a matter of fact, we
can all agree (even Jacques Guy), that NLs are at least of some finite
size.  How much larger they are depends on whether one considers that
grammars of natural languages contain what Paul and I called "size laws",
of which the following are possibilities. (For simplicity, I assume that
these all deal with the notion of "length".)
```

1. There is some fixed, finite n, such that all sentences of all natural languages are no longer than n. Something like this assumption was made by Peter Reich in a paper in LANGUAGE in 1969.
2. All sentences of all natural languages are of finite length, but there is no n, as in (1). This is the standard Chomskyan assumption.
3. There are no size laws. That is there are no priniciples of grammar that either explicitly or implicitly limit the length of sentences in natural languages. This is the assumption that Paul and I defend, basically on simplicity grounds.

Paul and I show that given (3), and given a principle of grammar concerning coordinate compounding, there are transfinitely many sentences in all natural languages for which that principle holds.

The principle itself is quite simple and we think uncontroversial. It says in effect if expressions of type X (such as the type of declarative sentences) are compoundable by coordination, then for any two or more expressions of that type, there is at least one coordinate compound in the language, also of type X, made up of those expressions. To see how this works in a simple case, let us limit our (uncompounded) expressions of type X to the countably infinite set B = <<(I know that)**n Babar is happy: n >= 0>>. (Note: I use << and >> as set delimiters, to avoid transmission difficulties with the curly brace symbols.) From the principle, there is an expression of type X for every member of the power set of B, excluding the empty set. Call this set B'. For example, B' contains <<Babar is happy, I know that I know that Babar is happy>>, and corresponding to this member is the expression, also of the appropriate type, Babar is happy and I know that I know that Babar is happy. (In fact, there are other expressions as well, but we ignore these.) By standard assumptions of set theory, B' has greater cardinality than B. Paul's and my result then follows, unless one imposes a size law such as (2), which effectively (and in our view arbitrarily) limits the application of the coordinate compounding principle to sets of expressions of finite cardinality (that is, one replaces the "two or more" in my formulation above by the range 1 < x < infinity, where x ranges over the number of expressions that are coordinately compounded). Paul and I argue that the best theories of natural languages are those which do not countenance size laws, which we claim have two properties:
1. they are not constructive (e.g., generative);
2. they support a realist (i.e., Platonist) view of natural languages.
The first property certainly must hold, but given the spirit of much contemporary theorizing is less controversial than it was when we published our book in 1984, if it was even controversial then, given the rise of "principles and parameters" based theories of natural languages during the 1980s. The second property, it now strikes me, does not hold. Paul and I spilled a lot of ink trying to clarify Chomsky's "conceptualist" view of natural languages, and having clarified it, to show that it is incorrect. But it now strikes me as quite easy to work out a coherent conceptualist theory of natural languages that does not incorporate a size law. In fact, any principles and parameters theory that does not do so, but that does include what strikes me as the patently correct principle of coordinate compounding, will yield the result that natural languages (albeit, E-languages, in Chomsky's terminology) are of transfinite size.

Avery Andrews on why it could matter to linguists, and which questions are important and why, in 2.634 and 2.663.

2.634, Message 1: finiteless of language

Date: Tue, 8 Oct 1991 13:05:05 GMT
From: AVERY D ANDREWS <ADA612@csc1.anu.edu.au>
Subject: finiteless of language

It seems to me that both Jacques Guy and Postal/Langendoen are missing something important in their positions on NL cardinality. What Guy is missing is that the idea that sentence length has no finite bound is an interesting idealization of the behavior of an actual device, namely, us. If one increases the quality of the life-support systems and the strength of motivation, the maximum attainable sentence length will increase, with no fixed limit. Such limitations as there are on sentence length are not inherent in the structure of the language faculty, but derive from other features of human existence. On the other hand, what Postal/Langendoen miss is that a sentence of infinite length is not an idealization of our behavioral capacities: no amount of life-support & proferred rewards are going to get an infinite-length sentence out of anyone's mouth. The mathematics of infinite languages might be amusing for its own sake, but it has no bearing on the goals of generative grammar.

Avery Andrews (ada612@csc.anu.edu.au)

2.663, Message 2: Is language finite?

Date: Tue, 15 Oct 1991 10:35:10 GMT
From: AVERY D ANDREWS <ADA612@csc1.anu.edu.au>
Subject: Is language finite?

I don't really like disagreeing with Alexis Manaster-Ramer, but I do. The important issue, I think is not cardinality of NLs per se, but what the aim of generative grammar is supposed to be. I would go with Chomsky in saying the aim of figuring out how an actual device works and is organized. Given this, I find finite but not finitely bounded sentence length to be the `obviously' appropriate idealization, since the limits on actual sentence length are (a) not well defined (b) clearly not due to the structure of the gadgetry responsible for the facts of grammar. I see (a) and (b) as empirical issues, though the step from them to what the appropriate idealization is is not an empirical issue (and probably not a very important one either).

Infinite sentence lengths are on the other hand not attainable by any kind of gadgetry whatsoever, and therefore, I would claim, are irrelevant to investigations into the function and structure of mechanisms.

The infinite size of languages was perhaps a more important issue thirty years ago than it was today, since nowadays it is widely accepted that grammars have to capture generalizations, while in those days the idea that a grammar might be a huge, stupid and unorganized list of sentence patterns was more of a serious contender.

Avery Andrews (ada612@csc.anu.edu.au)

Excerpts from long reply by Manaster-Ramer to Avery (and others) in 2.678.

Message 1: Infinite Languages

Date: Tue, 15 Oct 91 11:49:30 EDT
From: <Alexis_Manaster_Ramer@mts.cc.wayne.edu>
Subject: Infinite Languages

LINGUIST has once again made it possible for a significant issue to be dissected in a way which does not seem possible in any other forum I can think of! Avery Andrews and I seem to disagree on the status of the claim that NLs are sets of sentences without an upper bound on their lengths (but always of finite length). Yet, as he points out in his latest, the real issue he is concerned with is not cardinality but rather the goals of linguistic theory as originally defined by Chomsky. I, on the other hand, was concerned specifically with the question of cardinality (and more generally with the issue of how appropriate it is to treat mathematical idealizations as literal claims about the real world).

Having said this, I would like to try and convince Avery and everyone else that we must make some distinctions that are not usually made in this area:

(1) To say that the limits on actual sentence length are undefined is not the same thing as saying that there aren't any. Let me use a simple analogy to illustrate this point. The set of citizens of the United States is finite and reasonably well-defined. The set of black citizens of the U.S. is not well-defined at all, yet it is clearly a subset of the former, and hence also finite.

(2) The fact that sentences as they become longer and more complex also become less acceptable can be captured in a number of ways, one of which is to divide up your theory into two components, one called competence, the other performance, and let the second only worry about this fact. However, this is NOT the only reasonable alternative. Another is to assume that the device (or gadgetry, as Avery calls it) that we have is one that does things in real time, and that competence is an idealization of it. Thus, the theory of competence is not about a specific mental organ separate from performance. It is about the one mental organ that there is (but it is idealized in certain important ways, which is perfectly reasonable, by the way. I have no quibble with idealizations.)