What to be realist about in linguistic science

Geoffrey K. Pullum

School of Philosophy, Psychology and Language Sciences
University of Edinburgh

This paper was presented at a workshop entitled 'Foundation of Linguistics: Languages as Abstract Objects' at the Technische Universität Carolo-Wilhelmina in Braunschweig, Germany, June 25–28.

> At the end of his viva, Hilary Putnam asked him, "And tell us, Mr. Boolos, what does the analytical hierarchy have to do with the real world?" Without hesitating Boolos replied, "It's part of it."
>
> [From the *Wikipedia* article on the MIT logician George Boolos]

Prefatory remarks on realism

I need to begin by clarifying what is meant (or at least, what I and others have meant) by 'realism'. That such clarification is necessary is indicated by the fact that the subtitle of this workshop ('Languages as Abstract Objects') clearly pays homage to Katz (1981) (*Language and Other Abstract Objects*). What the late Jerrold J. Katz meant by the word 'realism' does not comport with the usage of most philosophers.

Katz shows little interest in defining his terms. He barely even waves a hand in that direction. The third page of *Language and Other Abstract Objects* book announces that he has had an epiphany: there is "an alternative to both the American structuralists' and the Chomskian conception of what grammars are theories of, namely, the Platonic realist view that grammars are theories of abstract objects (sentences)" (Katz 1981, 3). Neither 'Platonic' nor 'realist' are analysed, explicated, or even glossed at this point. 'Platonic realism' is thereafter abbreviated to 'Platonism' and simply assumed to be understood.

Katz's index gives page 15 as the first citation for the terms 'nominalism' and 'conceptualism', but his index is mistaken. From page 10 onward both these terms for traditional alternatives to Platonism (and contradictions of it) are used without definition. We get just two parenthesized clarificatory phrases: Katz makes reference to "the highly articulated formulations of linguistic nominalism (in the structuralist tradition) or linguistic conceptualism (in the tradition of generative grammar)" (Katz 1981, 10), and that is that. We do find a definition, however, in his later work *Realistic Rationalism* (1998: 2):

> What makes someone a realist is his or her acceptance of abstract objects; what makes someone a realist of a particular kind is his or her acceptance of abstract objects of that kind.

But this is not at all what is standard in philosophy. Realism is standardly taken to be **relational**: you can be (or not be) realist with respect to the objects in some specific domain (say, trees or quarks or integers) but perhaps not some others (superstrings; moral properties; ghosts).

Moreover, realism about a class of objects is standardly taken to be a two-part claim: that (a) they actually exist, and (b) they exist independently of us. Thus realism about the objects referred to in mathematics holds that (a) the referents of mathematical terms actually exist (rather than being fictions), and that (b) their existence is independent of our thoughts, perceptions, and use of language.

What I think is going on is that Katz is presupposing an acquaintance with the problem of universals in late scholastic medieval philosophy, and takes the three standard solutions to be the only ones there are or could can be. For the late scholastics the worry was about whether properties exist, and if so, what exactly they are. Those who said yes, properties exist, have been referred to as Platonists, and those who said no, there are only particular things, have been called nominalists (confusingly, since there are other older definitions of that term). But of course nominalists would typically be realists about tables, chairs, physically detectable particles, and physical inscriptions of formulae, so 'realist' should not be taken as an antonym of 'nominalist'.

Abstract objects

Abstract objects are crucial to the whole of Katz's philosophy, so it is noteworthy that he doesn't define that notion either. One might think that he imagines Plato had done the work two thousand years ago, but in fact the abstract/concrete distinction is recent in philosophy: only in the 20th century did it begin to receive any attention. The notion of abstract objects emerged out of Frege's 'third realm' of objects that are neither mental nor sensorily perceivable.

Often abstract objects are defined as those that do not occupy any definite region of space or period of time. Alternatively, they may be defined as the causally inefficacious objects. Perhaps the two definitions are compatible and could be conjoined.

In the latter connection, George Boolos makes an interesting remark on the ways in which abstract objects might be relevant to us despite their causal inertness. It is worth quoting in extenso:

> We do not engage in physical interactions with them, in which energy is transmitted, or whatever. But we twentieth-century city dwellers deal with abstract objects *all the time*. We note with horror our *bank balances*. We listen to *radio programs*: *All Things Considered* is an abstract object. We read or write *reviews of books* and are depressed by *newspaper articles*. Some of us write *pieces of software*. Some of us compose *poems* or *palindromes*. We correct *mistakes*. And we draw *triangles* in the sand or on the board. Moreover, bank balances, reviews, palindromes, and triangles are "given" to us "in experience," whatever it may mean to say that. . . . [No] sense of "sensible" or "experience" has been shown to exist under which it is not correct to say that we can have sensible experience of such objects, such things as the zither melody in *Tales from the Vienna Woods*, the front page of the sports section of the morning's *Globe*, a broad grin, or a proof in set theory. . .
> (Boolos 1998: 128–129; emphases in original)

Being realistic

Some remarks of Chomsky's (characteristically) render things even more confused. Consider this passage:

> The shift of perspective from the technical concept E-language to the technical concept I-language taken as the object of inquiry is therefore a shift toward realism in two respects: toward the study of a real object rather than an artificial construct, and toward the study of what we really mean by 'a language' or 'knowledge of language' in informal usage.

The first point contrasts "real object" with "artificial construct", where 'real' seems to mean either 'genuine' or 'physical', which has nothing to do with the usual philosophical sense of realism (an artificial leg is certainly an artificial construct, but is just as real as an regular biological limb). The second speaks of "what we really mean by" certain everyday informal locutions, which of course could lead us away from realism if the correct analysis of our informal usage construed the locutions 'a language' and 'knowledge of language' in antirealist terms.

Everyday uses of the adjective 'realistic' nearly always relate to savvy common-sense appraisal of prices or situations, e.g., assessing the cost of some course of action. That too is just a red herring, which makes the title of Katz's *Realistic Rationalism* amazingly unhelpful. The book should apparently have been called 'Platonist Rationalism'.

Mind-independence

Perhaps the most serious problem with Katz's exposition is his neglect of the issue of mind-independence. Katz often writes as if objects like numbers, propositions, and natural language sentences would exist even if we bipedal primates had never evolved, and nothing about our thoughts, opinions, definitions, or ways of speaking had any bearing or influence on their existence; but he says very little in direct terms about this claim, at least with respect to linguistic concepts. As a result there is a highly significant gap in his picture of the logical possibilities: we get absolutely no discussion of the possibility of objects that are abstract but not mind-independent.

For the Platonist realist, nine abstract objects are referenced by the beautiful equation $e^{i\pi} + 1 = 0$, known as Euler's identity: the additive identity element 0 and the multiplicative identity element 1 (both natural numbers), the real numbers e and π (both transcendental numbers), the binary equality relation '=' (a subset of $\mathbb{R} \times \mathbb{R}$), the addition operation '+', the multiplication operation symbolized by adjacency, and the exponentiation operation symbolized by superscripting the exponent (all ternary functions, i.e. subsets of $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$), and even the imaginary unit element i satisfying $i^2 = -1$. Moreover, $e^{i\pi} + 1$ is claimed to have borne the equality relation to 0 through all eternity: Euler's identity would have been true all along even if Euler had never been born, or if primates had been destroyed by an asteroid strike on the Rift Valley a million years ago and the human species had never evolved.

Maintaining this kind of thoroughgoing realism with respect to the expressions of natural languages would involve claiming that those expressions actually exist in a way that is independent of the existence or mental activities of human beings. Katz's view says that all of the natural languages that linguists study, from Aari to Zuni, would exist just the same in a universe where life had never evolved. So would infinitely many other languages over infinitely many other vocabularies. That is indeed what Katz maintained. But it is surely reasonable to doubt it.

Too many languages

One problem that as far as I can see Katz never addressed is that there are too many languages under this view, infinitely many of them being utterly unlike human languages, but all of them apparently being fit subjects for study by linguists. Katz doesn't just imply this, he insists upon it.

Sets of dialects of a natural language are just classes of sets of sentences, and "There is an infinite range of such classes, including English, French, Sanskrit, Engrench (i.e., a class of sentences with English syntactic structure but an anglicized French vocabulary), and infinitely many other languages, living, dead, unborn, conceivable, and inconceivable" (Katz 1981: 9). The set of well-formed sentences of Standard English in American spelling as exemplified in Katz's own writings is not distinguished from any of these imaginary languages:

(1) a. $h(\text{English})$ where h is a homomorphism mapping all nouns, verbs, and adjectives to *blah*. (Sample text: *We blah these blah to be blah, that all blah are blah blah, that they are blah by their blah with blah blah blah, that among these are blah, blah, and the blah of blah*.)

b. The concatenation of Japanese and French (i.e., the set $\{wx | w$ is a sentence of Japanese and x is a sentence of French $\}$).

c. The set $\{ii, iii, iiiii, iiiiiii, iiiiiiiiii, \ldots\}$ containing all and only the prime-length strings of the letter i.

Linguistics is not distinguished from formal language theory or logic or theoretical computer science under Katz's view: its truths as well as its methods are entirely a priori. As he says: "Platonists classify linguistics with the mathematical sciences, rather than with the social, biological, or physical sciences, as also about a reality outside of us and the physical world"; thus "statements about the grammatical structure of sentences are no more *empirical* than statements about numbers" (Katz 1981: 23).

As a posteriori philosophy of science this will not do: linguists simply do not work in a way that can possibly be regarded as rendering this description true. Asserting things about the grammatical structure of sentences in arbitrary invented languages does not count as work in linguistics. Nor does the development of true and provable theorems concerning such languages. It is certainly true that in any language where every clause contains a verb, and all verbs are octotransitive (i.e. they take eight obligatory object NPs), and subject NPs are obligatory in all clause types, and NPs minimally consist of a noun followed by a determiner, every sentence has at least 11 words; but this would not be recognized by any linguist as a syntactic discovery falling within linguistics.

Effability

Katz suggests only one universal property to offer to pick out the natural languages from all other sets of strings: his Principle of Effability.

(2) Each proposition (thought) is expressible by some sentence in every natural language.

But this condition is nowhere near enough to delimit the natural languages. In fact it seems neither necessary nor sufficient. It is not sufficient to define a language as a natural language because no linguist would allow (1a) to count as a natural language, yet it can express with single sentences all the propositions that English can express.

And it is almost certainly not a necessary property of natural languages because there are indefinitely many sentences of English that cannot reasonably be regarded as translatable into Pirahã. Pirahã has no numeral words whatever. There is a word used for the description of very small quantities such as 1 or perhaps 2; there is a word used for meager quantities such as 2 or 3; and there is a word used to denote more substantial quantities like 'plenty'. But when assessing the

quantities of successively large sequences of objects speakers do not even agree with the judgments they make on the same sequences considered in decreasing order. I think there is literally zero possibility of translating sentences like these into Pirahã:

(3) a. *Federal authorities moved today to toughen oversight over banks offering subprime mortgages.*

b. *The square root of two is irrational.*

c. *If you bought cheaper coffee and invested $3.50 a day into the stock market starting at age 25 you could have $225,000 extra when you retire.*

Thus there are obviously non-natural languages that can express every proposition expressible in English, and there are propositions that cannot be expressed in some extant natural languages. I conclude that Katz has no way of delimiting the objects of study for the field of linguistics.

Platonism and its rivals

If Platonism about linguistic objects is the view that sentences (etc.) are abstract and mind-independent, then it has more than the two rivals (nominalism and conceptualism) that Katz discusses. At least the following possible positions concerning English sentences conflict with Platonist realism:

(4) a. Katz's 'Nominalism': English sentences exist mind-independently but are concrete rather than abstract.

b. Katz's 'Conceptualism': English sentences exist as mental representations: they are both concrete and mind-dependent.

c. Pullum & Scholz's 'Constructivism': English sentences exist as abstract yet mind-dependent objects.

d. Fictionalism: English sentences do not exist at all, and statements entailing their existence are strictly false (though no doubt useful for purposes of organizing discourse about linguistic behaviour).

It should not be thought that any of these are self-evidently absurd, or already refuted by (e.g.) polemics in the 1960s against American structuralism or the arguments of Katz and Postal (1991) against Chomskian linguistics.

Katz's Nominalism

Nominalism in the sense of admitting absolutely nothing but physical objects to one's ontology does seem to be very hard to work out in detail if it is to provide room for classical mathematics to be done, as famously shown in the struggles of Goodman and Quine (1947) and the more recent labours of Hartry Field. But it is well motivated by worries of an epistemological nature: for objects that are not spatiotemporally located and that have no causal relations to us or our world, it is completely obscure how we could come to know about them at all. Nominalism posits only objects that we could in principle come to know about.

Katz really has nothing to offer on the epistemological point other than the idea of a special 'intuition' that can link us to an abstract object in some mysterious way, permitting it to influence and guide our behaviour without being causally implicated. And while it may seem strange to

think of statement inscriptions like '$\forall x[x + x = 2x]$' as not making reference to the integers or the addition operation or the number 2, and of being useful to us without even being true, opinions differ on that, and have differed for centuries. Nominalism cautiously restricts itself to stuff that uncontroversially exists and could in principle have causal roles. That may be inadvisable (and will of course seem so to the Platonist metaphysician), but the motivation for it is not risible. And as Katz notes, Chomsky once agreed, for his early work contains contains various explicitly nominalist remarks.

Mind-dependence

It is surely not irrational to suggest that if there were no thinking beings in the universe there would be no sentences of English or any language. It seems to me comparable to the notion that if there were no thinking beings in the universe there would be no money.

We should distinguish (as philosophical realists generally do) two kinds of mind-dependence: a trivial one that calls tables and chairs mind-dependent because someone with a mind had to invent them and manufacture them, and a deeper one that does not regard tables and chairs as mind-dependent because they continue to exist even when no one is thinking of them or even knows of their existence, and could in principle carry on existing after all life on the planet had been rendered extinct by a burst of gamma radiation from a nearby star. We are interested in the latter sense, of course.

We regard a diamond deep underground that no human has yet discovered as existing: no human being needs to know about it or think about it in order for it to exist. Nothing similar holds for a dream that no one has ever had, or for a poem that no one has yet written, or a sentence that no one has ever thought to frame or utter. I think it is reasonable to regard such things as dreams, poems, and sentences as abstract yet mind-dependent.

Pullum & Scholz (1997) suggested that 'constructivism' might do as a term (not necessarily ideal) for a view centred on this possibility. It involves a conjunction of claims that was missing entirely from Katz's worldview. George (1996) seems to attribute it to Chomsky, apparently presupposing that Katz and Postal are wrong in thinking that Chomsky posits only time-bound and spatially located finite mental objects. And George is not alone in assuming that abstract objects can be mind-dependent: Quine (1948) claims that "there are universals but they are mind-made." One can relate such views to the movement in mathematics known as **intuitionism**; and Chomsky can be found in at least one place (Huybregts and van Riemsjijk 1982, 13) suggesting an intuitionist basis for linguistics.

Fictionalism

Fictionalism, too, is a well-known stance that many have adopted with regard to domains such as mathematics and morals. The idea is that to say that there is a number greater than 7 is strictly false, because there actually are no numbers, but it is a special kind of falsehood that happens to improve our ability to do reasoning about quanitities. Thus Goodman & Quine (1947) say:

> [O]ur position is that the formulas of platonistic mathematics are, like the beads of an abacus, convenient computational aids which need involve no question of truth. What is meaningful and true in the case of platonistic mathematics as in the case of the abacus is not the apparatus itself, but only the description of it: the rules by which it is constructed and run. These rules we do understand, in the strict sense that we can express them in purely nominalistic language.

The idea seems to be that to assert the existence of a number greater than 7 is strictly to assert a falsehood, but it is nonetheless true that there exists a convention whereby we pretend there are numbers, and use the symbol '8' for the one that follows the one called '7', and follow the rule that treats '$2^3 = 8$' as true and '$2^3 = 7$' as false and so on.

Metaphors work in a similar way, and we are entirely familiar with them: "He was walking on air after Chomsky praised him" is strictly false (you can't walk on air), but it is useful in conjuring up a vivid image of someone so delighted and proud that they hold themselves tall and walk as if gliding, hardly touching the ground. Maps work in the same way too: it is certainly false that Glasgow is seven inches to the left of Edinburgh, but it helps enormously with our route planning if we operate under the rules of a practice where we pretend that it is for certain purposes. So let's not treat fictionalism as completely out of the game.

Pluralism

Pullum & Scholz (1997) aimed to open up some space for two further views: **heterogenism**, which says that more than one ontological variety of stuff is relevant when interpreting theories of linguistic structure, and **pluralism**, which says that separate and distinct homogeneous domains of discourse can participate in making the statements of a linguistic theory true.

Their thesis is that even if there is a platonic realm where sentences exist independently of time or space, that doesn't mean that linguistics has to talk about that domain and no other. Katz, Postal, Langendoen, and others are in danger of propounding a linguistics that, by the very definition of its foundations, has absolutely nothing to say about utterances that are actually attested in written or spoken form, or about the mental operations that only a language-using animal like us can perform.

What should we be realist about?

Katz's book title implies that **language** is an abstract object; but this seems to me a very strange and obscure claim, comparable to calling culture, or fire, or life, abstract objects. We do use what grammarians call abstract nouns to talk about the practice of structured propositional expression of thought, or ways of organizing human life, or the key property of belonging to the biological realm; but being named by a noun doesn't suffice to guarantee thinghood! It is astonishing that I should need to point this out, 80 years after Bloomfield and the Vienna Circle philosophers expressed the point so clearly:

> "Ordinary language ... uses the same part of speech, the substantive, for things ('apple') as well as for qualities ('hardness'), relations ('friendship'), and processes ('sleep'); therefore it misleads one into a thing-like conception of functional concepts." (Hahn et al. 1929)

Did Katz mean 'languages'?

Katz may have carelessly said 'language' when what he really meant was 'languages'. But I believe we should not be realist about languages because they are not the kind of thing for which scientifically reconstruction is appropriate. This needs some clarification.

What Katz was concerned to reject above all was the position advocated with increasing explicitness by Chomsky since about 1983. It is a strange position, latterly cloaked in an even stranger terminology. Chomsky holds that languages (as customarily understood) are not the object of inquiry for linguists. Instead, the things linguists study are what Alexander George (1989: 91) calls

physiogrammars, but which Chomsky has since 1986 chosen to refer to by the peculiarly perverse term 'I-language', where the 'I' is supposed to suggest all three of 'individual', 'internal', and 'intensional'.

A mentally represented generative grammar is (in George's terminology) a **psychogrammar**, and a physiogrammar is a physiologically instantiated inscription of a psychogrammar. (Since Chomsky appears to endorse the mind–brain identity thesis, these may be one and the same thing for him.) A psychogrammar is individual in that it inheres in a single person's brain states, not in a set of conventions maintained by a speech community; it is internal in that is located in the brain of a single speaker and nowhere else; and it is intensional in that it characterizes the sentence of a language in intension (Chomsky never really clarifies this point). His thoroughgoing materialist position on physiogrammars is one thing he has been admirably clear about. He claims they are spatiotemporally located and causally efficacious brain state components:

> [M]entally represented grammar and UG are real objects, part of the physical world, where we understand mental states and representations to be physically encoded in some manner. Statements about particular grammars or about UG are true or false statements about steady states attained or the initial state (assumed fixed for the species), each of which is a definite real-world object, situated in space-time and entering into causal relations.
> (Chomsky 1983: 156–157)

However, while he is fiercely realist about physiogrammars, he seems to be entirely antirealist with regard to linguistic expressions (sentences, phrases, etc.). For him they are evanescent or even epiphenomenal; they are sometimes referred to as 'representations' despite lacking anything to represent, and sometimes referred to as 'structural descriptions' despite lacking anything to structurally describe. Any appearance of existence they might have is due solely to the fact that I-languages 'generate' them. The sole function of an I-language is to generate a set of expressions which Chomsky calls an 'E-language':

> But sets are not in the mind/brain and grammars can be chosen freely so long as they enumerate the E-language, so the study of E-language, however construed, does not seem to bear on the truth about [what differentiates] speakers of English and Japanese: it is not even in principle, part of the natural sciences, and one might argue that it is a completely pointless pursuit, simply a matter of chasing after shadows.
> (Chomsky 1986: 34)

Part of the basis for Chomsky's insistence that E-languages are not a suitable object of study is his claim that it has never been clear where to draw the boundaries of any E-language, not even for Standard English:

> Consider what are sometimes called 'semi-grammatical sentences', such as the expression 'the child seems sleeping'. Is this expression in the language or outside it? Either answer is unacceptable. An English speaker interprets it instantaneously in a perfectly definite way, quite differently from the way this expression would be interpreted by a monolingual speaker of Japanese. Therefore, the expression cannot simply be excluded from the set 'E-English', though it is plainly not well-formed... But speakers

> of English and Japanese will also differ in how they interpret some sentence of Hindi. Therefore we conclude that all languages fall within English, a conclusion that makes no sense.
> (Chomsky 1987: 33)

This is an important observation, and I will return to it. But for now, notice that all of the concern is with identifying a particular **set** of expressions. But what of individual expressions? They go unconsidered and almost unmentioned through the debate between Chomsky on the one hand and Katz and Postal on the other about the metaphysics of linguistics.

What looks more plausible to me, as an a posteriori reconstruction of the practice of linguistic science, is a kind of diametrical opposite of Chomsky's view. What I think we should treat as the object of study for linguistics is **individual linguistic expressions and their structures**. We use linguistic expressions every day; they are in practice the very thing that linguists focus their attention on; and we can often be highly precise about their structural properties.

Grammars, by contrast, are surely best seen as theories, just as they traditionally were. They are descriptive theories about ways of structuring expressions. While we would want to be able to show that grammars are learnable under appropriate conditions, there is no reason to think that they will be paired one to one: a class of objects has many equivalent descriptions, and there are no signs whatever of speakers of a given language homing in on a **unique** grammar for it.

What are grammars like?

To the extent that Chomsky has ever made his conception of a grammar clear, he has taken grammars to be generative grammars. What are they?

—**Generative grammars are not functions**. The remark in Chomsky (1959: 137) about a generative grammar being a recursive function with the strings of the language as its range seems to be just a mistake: a generative grammar is not a function F that takes arguments and yields values. That is, although a generative grammar G may generate a set of strings that has *Everybody wins* as one of its members, there is no element b such that $G(b) =$ *Everybody wins*.

—**Generative grammars are not transducers**. The passage in Chomsky (1986: 26) about how a generative grammar assigns some kind of structure to every possible acoustic signal, is also a mistake. It simply isn't true that a generative grammar, whether of the sort Chomsky envisaged thirty years ago or of the minimalist sort he envisages now, takes acoustic inputs and transduces them to structural representations.

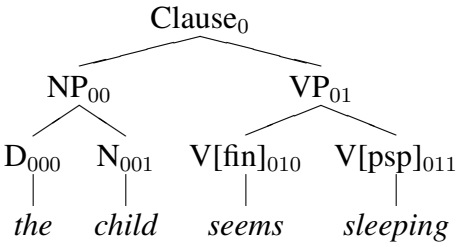
—**Generative grammars are random enumerators**. Generative grammars, as Chomsky has always outlined them in their differing versions, are nondeterministic random enumerators. A 'device' of this sort is said to 'generate' a natural language L iff (1) for every well-formed word sequence x in L there is some way of using the grammar to construct a derivation ending in x, and (2) for no ill-formed word sequence x in L is there any way of using the grammar to construct a derivation ending in x.

Model-theoretic syntax

The best way to render the notion of a grammar precise, I will argue, is basically the one introduced by Johnson and Postal (1980): model-theoretic syntax. Under the model-theoretic proposal a grammar is a theory (a finite and consistent set of statements closed under consequence) in a logical language interpreted on structures of the kind used for representing linguistic properties. The structures that satisfy the grammar are fully well-formed.

To the extent that a structure fails to satisfy the grammar, it is to that degree ill-formed (and of course many degrees of ill-formedness are possible). And here is the solution to Chomsky's puzzle about semi-grammatical expressions. Considered in the light of the constraints on structure that we would want to incorporate into a grammar of English, *The child seems sleeping* violates some constraint on choices for the complement of the verb *seem*.

(5)



With an ordinary NP as subject (not the dummy subject *it*), *seem* must have a complement, and that complement must be either

- a subjectless *to*-infinitival (*The child seems to be sleeping*) or
- a predicative adjective (*The child seems charming*) or
- an indefinite predicative NP (*The child seems a trouble-maker*) or
- an idiomatic predicative PP (*The child seems out of control*),

but not

- a finite content clause (**The child seems that we should leave*) or
- a bare infinitival (**The child seems be sleeping*) or
- a tensed VP (**The child seems sleeps*) or
- a subjectless past participial clause (**The child seems been sleeping*) or
- a subjectless gerund-participial clause (**The child seems sleeping*).

But the violation is at one node only (node 011 in (5), where the nodes are systematically numbered in Gorn tree-domain style), and only a single constraint is violated at that node. The expression is in violation of the grammar, but less so than these:

(6)

a. **Child the seems sleeping.*
(This also violates an ordering constraint: D precedes N.)

b. ***Child the sleeping seems.*
(This violates the same ordering constraint but also a second one, lexical head precedes complement.)

c. ****Of exterminates what the.*
(This, given any structure you like to posit, violates all sorts of constraints left, right, and centre.)

Lexical openness

The lexicon is a set of statements about specific terminal elements associated with meanings and syntactic, morphological, and phonological properties. An important consequence of MTS is that

syntactic regularities are freed from dependence on the lexicon. That is, a statement of syntactic regularities that are not lexically conditioned can be given in a way that is entirely separate from the accidents of which words belong to the language at any given time.

Consider the case of a language (like Pirahã) where there appear to be no complement clauses. An MTS account enables us to separate the issue of whether any sentences containing complement clauses exist from the question of whether there is a constraint excluding them. The syntactic constraints might allow for complement clauses exactly as in English, but if there happened to be no verbs currently in use that actually licence clausal complements (verbs like English *believe*, *conjecture*, *wonder*, *inquire*), sentences illustrating this syntactic possibility would be lacking. A verb with the relevant meaning might be introduced in a single act of lexical innovation, the result being that the language immediately had clausal complements, with no change whatever in its principles of syntactic organization. This strikes me as an interesting possibility, which GES seems to exclude.

But perhaps the most striking and straightforward observation about the independence of syntax from lexicon concerns the way we can understand expressions containing words we don't yet know. This point has been discussed by Postal (2004) in connection with the claim that the lexicon is open. Postal notes that the items that can appear in sentences may not even be denumerable: as Zellig Harris (1968: 11) pointed out, in utterances like '*He went* _____', the complement of *go* can be any of an in-principle nondenumerable infinitude of utterable sounds. Postal adds that they might even include gestures or grimaces. But the more important issue seems to me to bear on the point made earlier by Johnson and Postal (1980: 675–677) about how syntactic structure is to be described.

A number of different logicians and philosophers have remarked on our ability to grasp the structure of utterances in our native language that have lexical items unknown to us. For example:

(7) a. Lewis Carroll (*Alice in Wonderland*, 1865):
'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe . . .

b. Ogden and Richard (*The Meaning of Meaning*, 1923):
The gostak distims the doshes.

c. Carnap (*The Logical Syntax of Language*, 1937):
Pirots karulize elatically.

Such examples have often been quoted in the linguistics literature, but never to illustrate what seems to me to be the crucial point about grammatical description, which is that no GES grammar can provide any basis for an account of our ability to understand such sentences. They are not generated, because they are not even strings over the vocabulary of words used in the expressions that are generated (they are deliberately constructed to be). Yet we do understand them, not just as syntactically well-formed but as meaningful, with clear truth conditions.

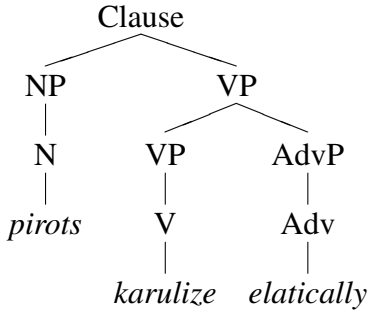
We learn from (7b), for example, that doshes can be distimmed, and that at least one gostak does this. How could we grasp such things if our knowledge of our language took the concrete form of a mentally inscribed GES grammar?

Under a GES description it is not even clear why utterances like those in (7) are recognized as being linguistic material rather than just noise. An expansion-oriented GES grammar will not

complete a derivation for any of them, and a composition-oriented grammar will not even get started — there can be no operation of combining *elatically* with *karulize* to produce a VP headed by the latter when neither is present in the lexicon so neither has any category.[1]

MTS offers at least a chance of explanation here. Consider a plausible structure for *Pirots karulize elatically*:

(8)



The point to notice is that there is nothing wrong with it according to any plausible set of syntactic constraints over the relevant set of categories. And that is true even for the terminal nodes. What constraint of English forbids *karulize* from being a verb? None — there are no constraints mentioning *karulize*. It is true that no dictionary includes *karulize*; but no dictionary comes with a firm guarantee that no other words exist.

We can account for such facts by giving a description in which syntactic constraints on NPs and VPs and adverbs are stated independently of anything about specific lexical items, and the conditions that define lexical items are stated as requirements place on particular phonological shapes. So there might be a constraint saying that *the* is to be used only as a definite determinative, or that *giraffe* is to be used only as a count noun species name distinct from *lion*, *zebra*, etc. Such a lexicon would simply not say anything at all about how *karulize* should be used. That would leave *karulize* free in principle to be a verb — though of course it is not linked to any meaning. What (7c) tells us is clear in a sense (that elatic karulization is one of the things that pirots do), but in another sense it tells us little, because a pirot could be anything, for all we know, and karulization could be any sort of process, and elaticity could be any kind of quality of karulization. We simply don't know what pirots are or what karulization or elaticity might be. But these are not questions about English!

Infinitude

Pullum and Scholz (2010) argue at length that the attitude generative linguists have taken toward infinitude as a property of languages has been completely cart-before-horse backwards. In works of various sorts (textbooks, popularizing books, encyclopedia articles, a letter to *Language*, a paper in *Science*) linguists have insisted that human languages are infinite in the sense of having infinitely many grammatical sentences, and have portrayed generative grammar as a theory capable of accounting for this revealed and remarkable fact.

Pullum & Scholz cite literature suggesting a complete lack of recursive hypotaxis and syndetic coordination in various languages: early Akkadian, Dyirbal, Hixkaryána, Pirahã, Proto-Uralic,

[1] Chomsky (1957: 104–105) discusses Carnap's example, and calls it a 'sentence', but only in the context of attacking the 'dubious' nature of appeals to structural meaning. He appears not to see that our ability to see anything at all in such a sentence goes entirely unexplained by a GES grammar.

Wargamay, and probably many others. But it is customary in current linguistics to simply ignore such testimony and stubbornly assert infinitude regardless, as a universal property of languages that calls for theoretical explanation.

The idea that languages could be antecedently discovered to be infinite rather than theoretically defined to be is of course untenable. No amount of evidence from any language could support or refute the claim of infinitude. What we know is that there seem to be no tight bounds seem to constrain us when framing complex thoughts in English, but that is not enough: No tight bounds seem to constrain us when composing haiku, but the set of possibilities is finite (the space of phonologically possible 17-mora Japanese haiku is roughly 10^{34}, a very small number compared to the number of atoms in the visible universe). The set of possible chess games, assuming the usual truncation rule to force a draw after a game descends into repeated sequences of moves, is also finite, though very hard to compute. Human languages have not been found empirically to be infinite. Rather, they are stipulated to contain a denumerable infinity of finite-length sentences by the facts of how generative grammars are defined. The finite-length limitation and the denumerability condition could be revised either upward or downward given appropriate theoretical alterations. What we really want, though, is a way of representing languages as having syntactic structure without committing ourselves to any motiveless upper or lower bound on the numerosity of sentences.

It is a signal virtue of model-theoretic syntax that it permits this. Stipulating the class of available models can be kept quite separate from the determination of the content of a grammar. A constraint saying that the verb heading the predicate VP of a tensed clause must bear the agreement form that is compatible with the head noun of the subject noun phrase of that clause will hold in any appropriate structure regardless of how big it is or how many others there might be. The content of a set of grammatical constraints does not have to say anything that either limits the size of possible structures or asserts that there is no such limit.

If the set of constraints does allow for nontrivial proper recursion in structures where every terminal node is a lexical item, then it may be possible for infinitely many distinct structures to satisfy the grammar; but that is not the same as saying that they have been shown to exist. If the language happened to be large but finite, the additional structures that satisfy the grammar would be merely structurally analogous to the ones belonging to the language and thus not specifically excluded, like the outcomes of games with the rules and pieces of chess but played on a board of more than 64 squares.

References

Boolos, G.: 1998, *Logic, Logic, and Logic*, Harvard University Press, Cambridge, MA.

Carnap, R.: 1937, *The Logical Syntax of Language*, Harcourt Brace, New York.

Carroll, L.: 1865, *Alice's Adventures in Wonderland*, Macmillan, London.

Chomsky, N.: 1957, *Syntactic Structures*, Mouton, The Hague.

Chomsky, N.: 1959, On certain formal properties of grammars, *Information and Control* **2**(2), 137–167. Reprinted in *Readings in Mathematical Psychology*, Volume II, ed. by R. Duncan Luce, Robert R. Bush, and Eugene Galanter, 125–155, New York: John Wiley & Sons, 1965 (citation to the original on p. 125 of this reprinting is incorrect).

Chomsky, N.: 1983, Some conceptual shifts in the study of language, *in* L. S. Cauman, I. Levi,

C. D. Parsons and R. Schwartz (eds), *How Many Questions? Essays in Honor of Sidney Morgenbesser*, Hackett, Indianapolis, IN, pp. 154–169.

Chomsky, N.: 1986, *Knowledge of Language: Its Origins, Nature, and Use*, Praeger, New York.

Chomsky, N.: 1987, *Language in a Psychological Setting*, number 22 in *Working Papers in Linguistics*, Sophia University, Tokyo.

George, A.: 1989, How not to become confused about linguistics, *in* A. George (ed.), *Reflections on Chomsky*, Basil Blackwell, Oxford, pp. 90–110.

George, A.: 1996, Katz astray, *Mind and Language* **11**, 295–305.

Goodman, N. and Quine, W. V. O.: 1947, Steps toward a constructive nominalism, *Journal of Symbolic Logic* **12**(4), 105–122.

Hahn, H., Neurath, O. and Carnap, R.: 1929, *Wissenschaftliche Weltauffassung – Der Wiener Kreis*, Wolf, Vienna. Quotation from the translation by Marie Neurath and Paul Foulkes published as 'The scientific conception of the world: the Vienna Circle' in Marie Neurath and Robert S. Cohen (eds.) 1973, *Empiricism and Sociology*, ch. 9, 299–318.

Harris, Z. S.: 1968, *Mathematical Structures of Language*, Interscience Tracts in Pure and Applied Mathematics, 21, Interscience Publishers, New York.

Huybregts, R. and van Riemsjijk, H.: 1982, *Noam Chomsky on the Generative Enterprise*, Foris, Dordrecht.

Johnson, D. E. and Postal, P. M.: 1980, *Arc Pair Grammar*, Princeton University Press, Princeton, NJ.

Katz, J. J.: 1981, *Language and Other Abstract Objects*, Rowman and Littlefield, Totowa, NJ.

Katz, J. J.: 1998, *Realistic Rationalism*, MIT Press, Cambridge, MA.

Katz, J. J. and Postal, P. M.: 1991, Realism vs. conceptualism in linguistics, *Linguistics and Philosophy* **14**, 515–554.

Ogden, C. K. and Richards, I. A.: 1923, *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*, Harcourt Brace, New York. Page reference is to the third edition, 1930.

Postal, P. M.: 2004, The openness of natural languages, *Skeptical Linguistic Essays*, Oxford University Press, New York, pp. 173–201.

Pullum, G. K. and Scholz, B. C.: 1997, Theoretical linguistics and the ontology of linguistic structure, *SKY 1997: Suomen Kielitieteelisen Yhdistyksen Vuosikirja 1997*, Linguistic Association of Finland, Turku, Finland, pp. 25–47.

Pullum, G. K. and Scholz, B. C.: 2010, Recursion and the infinitude claim, *in* H. van der Hulst (ed.), *Recursion in Human Language*, number 104 in *Studies in Generative Grammar*, Mouton de Gruyter, Berlin, pp. 113–138.

Quine, W. V. O.: 1948, On what there is, *Review of Metaphysics* **2**, 21–38.