

Terms of Endearment

Juan Uriagereka, reply by Evelina Leivada

In response to "[Misused Terms in Linguistics](#)" (Vol. 5, No. 2).

To the editors:

I would like to start by thanking Evelina Leivada for an essay that is useful in both scope and tone. Civil discourse over matters that matter—and terms certainly matter—is something we can all benefit from. Apologies on behalf of the field, also, about the amazing anecdote that closes the piece. If linguistics is not an interdisciplinary science, no science is.

A pedantic nitpick: what Leivada calls Universal Grammar (UG) is really a *theory* of the mental organ it is supposed to describe. Noam Chomsky coined the term I-language in 1986, long before anyone, so far as I know, started to put “i” in front of just about anything. I am not sure whether that term is any less confusing. I have no troubles with UG, so long as we understand that there is a biological entity of some sort, and then a theory about that entity. This theory has, of course, changed over the years. Indeed, the theory Leivada refers to as involving different kinds of factors, including genetic endowment, history, and economy principles—usually called the minimalist program—has only been around for the last quarter century.

Since I find myself in the same minority as Leivada, I will make the rest of my comments from that admittedly narrow perspective, triggered in large part by the need to safely place linguistics among the disciplines with which it interacts, as well as the permanent questions the field has faced about learnability and evolvability.

Many of Leivada’s questions and clarifications stem from our theories of variation, which I entirely agree need attention. In fact, the situation may be even worse than she alluded to, whether we are speaking of parameters, features, or, for that matter, rules, as linguists did in the not-too-distant past. The numerical questions raised during the 1970s continue to be as pressing today as they were then. That is, so long as we assume that children acquiring a first language cannot rely on instructions, or what is commonly called negative data.

The problem emerges from the functions related to lan-

guage variation. Call the number of objects of variation, whether parameters, features, rules, etc., X . Take n to be the number of variations that number X allows, which is assumed to be at least 2, often 3 for privative features, or occasionally a scale for degrees of variation. That will grow as n^X . It matters much if said objects of variation are optional or obligatory, which yields 2^X possibilities, and whether the elements that add up to X are ordered among themselves, which yields $X!$ permutations. At that point, the explosion is served. This is why we know that either the entire approach is wrong or X has to be very small.

It is always possible to be wrong, although the issue then is to find an alternative theory that still does not rely on negative data or some other miracle. As for limiting X , one can stomp one’s feet about this, in various ways. My parameters are of course better than yours, just as my dialect sounds cooler than yours. The issue is how to make progress in a manner that is nonparochial and testable by current techniques.

Back in the 1980s, some of us already thought features could be interesting, so long as they avoid Norbert Hornstein’s pitfall, which Leivada mentions. Incidentally, Hornstein contributed to this pitfall, in my view, as did several others among my best friends, by taking θ -roles as features. Since that is just a dispute among siblings, I won’t spill any blood here. But a serious issue remains: features should be what features are when observed across the world’s languages. Period.

We probably will not debate whether “tense” or “person” or “number” and so on are features that simply show up across dialects, a question that drove philosophers in the Vienna Circle crazy. These are the imperfections of language. Linguists like me salivate over them. Not because we are library rats, checking dusty grammars for this kind of thing, but because it is the equivalent of astronomers finding a weird object out there in space. We presume there is some I-language, which UG is supposed to theorize about, and when we find its offspring, we get teary eyed and open the champagne. Our theories are such stuff as those imperfections are made on.

Incidentally, this approach has created a number of headaches for those of us, happy savages, who come from

the worlds of exotic languages. The deal was normally this: if your feature from a godforsaken dialect proved some pet theory of the moment—the *moment* being Chomsky’s Fall Class at MIT, which some of us religiously attended whenever we possibly could—you were the toast of the town for the next fifteen minutes. Or the next fifteen years, if you milked it properly. Alas, should the feature make the opposite case, it depended on how well you behaved. You could bring it up in the hope that someone in an ivory tower would find a reasonable way to incorporate it, sooner or later, at which point you might get some credit *if* the feature came from your own native language. But if you decided to make an amendment to the theory yourself on the basis of your feature... may all your papers be in proper order.

I kid because I care, and because frankly things have improved, in my view, largely due to the herculean effort of Ken Hale. He crisscrossed the world over in search of said features, in the process making friends with the locals whose languages he acquired. Note, I haven’t said “learned”; he somehow *acquired* them! You simply cannot describe as *appropriating* anything done by a friendly man who comes to live in your hut and fully acquire the local language, and at the same time introduces the principles of linguistics. I can name several groups of linguists, some very close to my heart, who came of age through Ken’s vision.

Anyway: *that’s* a feature. Nothing more, nothing less. If you postulate an abstract feature to make your trains run, cool: you do what you have to do. But be ready to work with others who may help your theory, to find the damn thing in some form or another in the world’s languages. If it ain’t there, it ain’t a feature. Not yet anyway. Call that, if you will, Uriagereka’s razor: *Do not claim a feature you need, unless you find it.*

In thinking about the matter for half a minute and without trying to be exhaustive, I came up with the following inventory of features that can be found in language after language.

Features generally presupposed in generative systems: perspective, wh-/focus, tense, negation/emphasis, mood, aspect, voice, person, number, gender, case, definiteness.

I’m sure we can all think of more, but that is plenty to make my point, simply by making X equal, say, 12, for the 12 features just listed.

I understand that classical learnability, based on something like the subset principle—that learners guess the smallest possible formal language compatible with the evidence received—is irrelevant in I-language terms, as it is unclear what the term smallest possible means when not dealing with set-theoretic objects, as in an E-language. Still, there has to be a time *t*, positive of course, that it takes a learner to figure out a parametric option, whether a rule, a parameter, a feature, or whatever varies. Say you

are at an intersection looking at Google Maps and hesitate whether to follow its directions or take that other left the nice neighbor offered as an alternative. How large is *t*? Since we are dealing with a cognitive process, possibly small. Slower than a mere reflex, but perhaps faster than an immune response.¹

If we are going to claim physiological clout for the discipline, we may as well take it moderately seriously. Say you determine, for example, one minute for *t* on average, just to make the calculations easy. Of course, you must also assume that children pass the time doing things other than acquiring language; they sleep, eat, poop, and spend an enormous amount of time messing around, attempting to steal their parents’ gadgets, etc. I have no idea what would be a realistic average time that a normal developing child devotes to language acquisition. Again, just for simplicity, let’s assume, unrealistically, that a child devotes eight hours a day to language acquisition, or 480 minutes. In a good day, the child can make 480 linguistic decisions.

Returning to the above list of features generally presupposed in generative systems, those features, if binary, would take $2^{12} = 4,096$ minutes to set, under the assumptions just run. Not bad, right? That very efficient child, working eight hours a day on the task, would be done in less than ten days. But is the case feature really binary? How about person? Perhaps it is a bit more than that. Just for clarity, $2^{10} \times 3^2 = 9,216$, even though it is not clear that case features, at least in principle, are merely ternary; you can be inherent (of several types), structural (of usually four types: nominaccusative, ergabsolutive, dative, and genitive, two of which divide further), without going into lexical cases. Similar issues can be raised about person values. These often vary across the world’s languages in terms of including or excluding the addressee, and in Thai—just to drive the point home—whether you are addressing His Royal Majesty. Depending on how you count, that’s really anywhere between four and ten values that UG provides! And, again for clarity: $2^{10} \times 4^2 = 16,384$, $2^{10} \times 10^2 = 102,400$. Well, 213 (eight-hour) days is probably still not too bad—after all, kids also work weekends.

Except that we haven’t even started. Remember the optionality of features? Jacqueline Lecarme was one of those out in the bush who publicized the remarkable fact that Somali codes tense *within noun phrases*, contrary to what our most revered forefather Marcus Terentius Varro preached. In his *De Lingua Latina*, Varro first taught us how to divide speech into four parts, one in which the words have cases, a second in which they have indications of time, a third in which they have neither, and a fourth in which they have both.² Varro was wrong: UG allows languages to indicate time even if they *also* indicate case, and Somali is there to prove the point. Ah, but that means features may or may not be in categories that a learner is trying to acquire. This puts our ongoing calculation at

something more like $102,400 \times 2^{12} = 419,430,400$ minutes. 2,394 years to learn a language is a lot of time, even for children working very hard at it. If it took a second, instead of a minute, to set those options, it would all amount to forty years. Didn't it take Someone thirty to come of age?

I am afraid we may not be done yet: the features have yet to be ordered. This is being pursued seriously by some theorists, so as to determine how agreement relations work. If the ordering follows from something else, such as syntactic configuration, which is itself dependent on more elementary matters, perhaps there is nothing to worry about. But if it is truly the case that features can *also* be linearly ordered—in addition to possibly having multiple values, *and* allowing for obligatory (tense in V) versus optional (tense in N) manifestations—then the crazy number arrived at in the previous paragraph must be multiplied by the factorial function that arises from permuting the features. And, boy, does that baby grow. If only one or two features need to be ordered, it is not a big deal. If it is half of the 12 features listed above, it is already 720 possibilities. With a dozen features, the number of combinations is 479,001,600. That result alone is a nightmare; even if assigning one second to time *t* for evaluating each of these many options, the task would take over 45 years. When one combines that with the considerations above, we lack a theory.

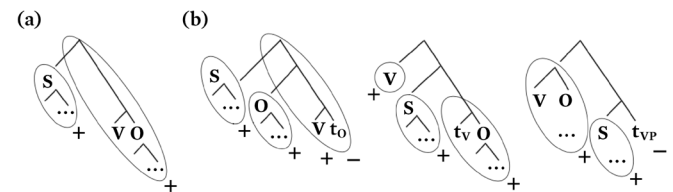
None of these issues are new. A textbook summary can be found in the second chapter of *Syntactic Structures Revisited* by Howard Lasnik et al., published in 2000.³ These were the simple mathematical considerations that helped us move from rules to principles in the 1970s. To be honest, I am not sure how to address this crisis. One can reduce the time *t* it would take to evaluate these options to the order of nanoseconds, whatever that means in neurophysiological and cognitive terms, or else we need to restrict our theories. This was known half a century ago, and it is no less true today.

If I were to make a positive suggestion to sharpen Uriagereka's razor, it would be with the following statement: you need parameters that correspond to actual development.

This touches on Leivada's comments about development, which are entirely apropos regarding metaphors, although they also can be qualified via what we know about development. For instance, if a neural language network is dictating matters, as Angela Friederici has claimed,⁴ that ought to determine a definite phase one starting in utero. Adam Wilkins, Richard Wrangham, and W. Tecumseh Fitch, as well as Cedric Boeckx et al., have provocatively suggested that neural crest development could be playing a role much earlier, related to what may have been aspects of domestication, in the technical sense, in language evolution.⁵ This is all highly speculative, but, from the point of view of the working linguist, highly plausible too, in possibly reducing the class of variations from the get-go.

Suppose that some version of Richard Kayne's linear correspondence axiom, or just about any other variant that maps hierarchical structures to linear orders in phonology, can be assumed.⁶ Key to these general ideas is that they massively change phrasal topology. Kayne's version,⁷ whose ultimate veracity is entirely immaterial to the argument I am trying to make, would yield something like the following.

Figure 1.



a) Linear correspondence axiom default, b) marked possibilities arising via movement.

The point is this: Kayne, or anyone else attempting what he tried, is plausibly predicting rather different (broad) prosodic envelopes for the objects in Figure 1, each object corresponding to one of the ovals. These objects are meant to separate default clausal orderings in a language such as, say, Chinese, as represented in Figure 1a, from those in, for instance, Japanese, Irish, or Malagasy, in that order for the examples in Figure 1b. I personally find it moving that Kathleen Wermke et al. have demonstrated babies cry slightly differently in the context of different languages, presumably as a consequence of setting these very early, entirely core, parameters.⁸ I like to think of these as dark parameters.

Suppose the picture—dare I say theory?—of parameters is something along the lines of the model presented in Table 1, which I obscurely proposed back in 2007:⁹

Table 1.

...Time...	a. Development	b. Acquisition	c. Learning
Neurobiological	Cranio-facial dev.?	Neural language network	Prefrontal cortex
Psychological	Sleep?	Sub-case parameters	Idioms, fads
Sociological	Mother	Speaker populations	Cliques

A model of language development, acquisition, and learning.

Again, it does not matter whether this model is precisely correct. The only point it makes, in relation to the numerical explosions above, is that it reduces parametric possibilities to

1. those in genuine development, as in dark parameters,
2. those in genuine acquisition via standard positive data, as in sub-case parameters, and
3. those that can be tweaked up until adolescence, as the pre-frontal cortex continues to mature and group identities are formed, as in microvariation.

In that order. Note that order here would not be strictly *acquired*, as it is dictated by development as one transitions from being a baby, to a child, to a teenager.

Homework is needed, though. We still want to see what sorts of features make it where and why. How, for instance, case features may have a bearing on the orderings that stem from Kayne's rationalization, in which case (pun intended), those would be effectively dark-featured. As opposed, perhaps, to adjustments on default values that continue pretty much throughout life. I, for one, have had as much trouble adjusting to using default gender values more appropriate to today's sensitivities as I did to using Zoom the last few weeks. I think of that as bona fide learning of the sort driven by purely statistical considerations, while earlier decisions in the model in Table 1 would be more akin to growing.

Last but not least, the theory presumed in Table 1 also makes predictions about acceptability, another important point Leivada discusses. A violation of grammar corresponding to superficial matters, such as those in column (c), need not be even remotely in the same league as violations corresponding to the dark parameters of column (a). In physiology in general, it seems kind of goofy to speak in terms of absolute grammaticality, as if a fever of over 100°F is ungrammatical but one under is only dispreferred. It all is what it is, and degrees of acceptability hopefully correspond to the stabs we are taking at modeling it all, whether as in Table 1 or any other approach one may reasonably attempt. I have tried to convince my experimentalist friends, so far without success, that this is a good idea to test.

I don't mean to plug my own work or that of my associates. But these are the perspectives that I have learned from the scholars cited in this commentary, plus a few others, especially Željko Bošković, Stephen Crain, Bill Idsardi, Tony Kroch, David Lightfoot, Massimo Piattelli-Palmarini, Eduardo Raposo, Ian Roberts, Doug Saddy, and William Snyder. To me, it all suggests that we are not as far from one another as it may at first seem, even if it is not always easy to agree on the terms we use, whether to endear ourselves with one another's proposals or to challenge them—all of which is, of course, useful.

Juan Uriagereka

Evelina Leivada replies:

I want to start by thanking Juan Uriagereka for the very thoughtful and interesting letter that usefully expands on

all the important points of my essay. This will be a very short reply, not because I do not have more things to say on the topic of misused, ambiguous, and polysemous terms in linguistics,¹⁰ but because I want to limit myself to the contents of his reply, and, unsurprisingly, it seems that I agree with everything he wrote. I want to highlight three points of agreement that I find crucial.

First, it is indeed true that the situation may be even worse than what I described, especially if we bring under scrutiny the primitives of inventories that feed one another: parameters that are localized to functional heads and heads that multiply to accommodate new features, according to the one feature, one head approach. I should perhaps explain that I had good reasons to tone down this bit of the discussion. The first reason has to do with personal preferences over tone and discourse. The second reason has to do with the reaction I received when presenting this work in conferences. One of the comments I got was that a junior linguist is in no position to tell other linguists how to do linguistics. Although I hope that my essay has made clear that this is far from my intention, I find the logic of this argument to be part of the problem. Uriagereka is right; the situation *is* worse, because certain terms of endearment are defended to the degree that attempts to track progress or inconsistencies in their use over the years is criticized by some as unnecessary nitpicking. I find the logic of this argument interesting too, especially when it comes from theoretical linguists. Following the exact same reasoning, who was Noam Chomsky in 1959 when he published his seminal review of B. F. Skinner's *Verbal Behavior*, if not a recently graduated young academic who reviewed the work of a prominent and much senior figure in the field? Scientific claims do not gain veracity based on how long they, or their proponents, have been around.

The second point on which I agree with Uriagereka is the claim that the problem arises mainly from the study of language variation. Enter "understudied varieties" here and amazing things happen. Uriagereka's description of the feature-discovery process is exquisite and it summarizes the situation much better than I did. Uriagereka's Razor, a term that I hope will find its way to linguistic textbooks—*Do not claim a feature you need, unless you find it*—is spot-on too. If I could add something to this formulation, it would read: *Do not claim a feature you need, unless you find it. And do not claim you found it unless its existence has been independently and repeatedly verified*—meaning the feature is seen by someone other than you, your three students, and your two lab associates. If that is not the case, it is fine, and nothing prevents you from still working with it. Call it a working notation based on your reading of the data, call it a hypothesis to be explored, perhaps describe its anticipated cognitive function—why does the system you describe need it?—but don't just say that this is a new feature A that describes a structure B, nefariously building on the unmentioned and unproven assumption

that this feature is innate and you have just discovered a previously unknown primitive of human cognition or biology.

The last point of agreement I want to highlight has to do with the notion of endearment. Linguistics is a small field, or at least it seems so if one looks at the number of grants that are given to linguists from big funding schemes in Europe.¹¹ Do we want it to remain small? Having a strong preference for terms that have been around for decades is perfectly understandable. However, the need to resolve issues that pertain to terminological coherence in order to boost the influence that linguistics exerts on neighboring fields should be stronger than one's affection for certain terms, precisely because the field is small and "we are not as far from one another as it may at first seem," as Uriagereka correctly observed.

Juan Uriagereka is a Professor in the Department of Linguistics at the University of Maryland.

Evelina Leivada is a psycholinguist at the Universitat Rovira i Virgili in Tarragona, Spain.



1. A tradition dating back to the work of B. Elan Dresher and Jonathan Kaye in the 1990s argues for a cue-based approach to these issues, whereby the learner is supposed to tackle not the primary linguistic data, but some, more sophisticated, higher order data (the cue), to settle on given options to the exclusion of others. B. Elan Dresher, and Jonathan Kaye, "[A Computational Learning Model for Metrical Phonology](#)," *Cognition* 34 (1990): 137–95, doi:10.1016/0010-0277(90)90042-I. In such a model, the time for each decision is not multiplied, but merely summed, to the others. The jury is still out on how such an approach would work in the sorts of syntactic instances we care about in syntax, particularly for unbiased learners whose data is of the sort human children arguably encounter. See, for example, Lisa Pearl, Timothy Ho, and Zephyr Detrano, "[An Argument From Acquisition: Comparing English Metrical Stress Representations by How Learnable They Are From Child-Directed Speech](#)," *Language Acquisition: A Journal of Developmental Linguistics* 24, no. 4 (2017): 307–42, doi:10.1080/10489223.2016.1194422 for discussion.

2. Jacqueline Lecarme, "Tense in the Nominal System: the Somali DP," in *Studies in Afroasiatic Grammar*, ed. Lecarme, Jean Lowenstamm, and Ur Shlonsky (The Hague: Holland Academic Graphic, 1996).
3. Howard Lasnik, Marcela Depiante, and Arthur Stepanov, *Syntactic Structures Revisited* (Cambridge: MIT Press, 2000).
4. Angela Friederici, *Language in our Brain* (Cambridge: MIT Press, 2017).
5. Adam Wilkins, Richard Wrangham, and W. Tecumseh Fitch, "[The 'Domestication Syndrome' in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics](#)," *Genetics* 197, no. 3 (2014), doi:10.1534/genetics.114.165423; Cedric Boeckx et al. "[Dosage Analysis of the 7q11.23 Williams Region Identifies BAZ1B as a Major Human Gene Patterning the Modern Human Face and Underlying Self-Domestication](#)," *Science Advances* 5, no. 12 (2019), doi:10.1126/sciadv.aaw7908.
6. Richard Kayne, *The Antisymmetry of Syntax* (Cambridge: MIT Press, 1994).
7. The ultimate veracity of Kayne's version is entirely immaterial to the argument I am trying to make.
8. Kathleen Wermke et al., "[Fundamental Frequency Variation in Crying of Mandarin and German Neonates](#)," *Voice* 31, no. 2 (2016), doi:10.1016/j.jvoice.2016.06.009.
9. Juan Uriagereka, "Clarifying the Notion 'Parameter,'" *Biolinguistics* 1 (2007): 99–113.
10. In fact, as I also say in my response to Anna Maria Di Sciullo's letter, many more terms can be critically discussed for content and use.
11. From the 6,907 projects funded by the European Research Council in 2017, 2,445 belonged to the panel Life Sciences, 3,113 to the panel Physical Sciences and Engineering, and 1,349 (a mere 19.5%) to the panel Social Sciences and Humanities, of which linguistics is a small part. Eva Kondorosi, "[The ERC: Funding Investigator Driven Frontier Research](#)," presentation at the European Research Council's Scientific Council Working Group on Widening European Participation, Tbilisi, Georgia, April 20–21, 2017.