

How to tear down the walls that separate linguists: Continuing the quest for clarity about general linguistics

MARTIN HASPELMATH

Max Planck Institute for Evolutionary Anthropology

This paper is a response to eight commentaries (for *Theoretical Linguistics*) on my target paper “General linguistics must be based on universals (or nonconventional aspects of language)” by David Adger, Balthasar Bickel, Roberta D’Alessandro, Diana Forker, José-Luis Mendivil-Giró, Susanne Fuchs & Ludger Paschen, Adam Tallman, and Dietmar Zaefferer.

The primary motivation for my target paper (“General linguistics must be based on universals”) was a perceived need for greater clarity in general linguistics. Many linguists seem to share the feeling that we often talk past each other because we understand terms like “theory”, “framework”, “explanation”, “analysis” and “description” in seemingly different ways. But even if different linguists (necessarily) prefer different methods, we need not disagree about these basic terms and concepts, and I think that some “walls” that separate linguists from different communities could be “torn down” if we became more aware of what unites all theoretical linguists: that we want to understand particular languages, and that we need them in order to understand Human Language in general.

I am grateful to the commentators for their interesting contributions, and here I continue the discussion by responding to some of their remarks. To recapitulate, I made three key points in my target article:

(A) language-particular description (p-linguistics) is no less theoretical than general linguistics (g-linguistics)

(B) general linguistic claims must be generally testable because they make universal predictions

(C) testing of these predictions can be based on independently defined universal yardsticks (e.g. Croft 2003), or on a natural-kinds programme (Baker 2001)

I also noted that the original natural-kinds programme is not being widely pursued anymore by generative linguists, and that this is a problem for the generality of the claims associated with generative analyses of particular languages.

The commentators challenge a number of these points and conceptual distinctions and also make some comments on nonconventional aspects of language. However, none of them presents a comprehensive alternative vision for the field, so I would like to continue to pursue the agenda of helping to unify the field a bit more.

On g-linguistics, p-linguistics, and theories

Bickel recognizes that the distinction between general linguistics and particular linguistics is almost trivially important, and it is easy to agree with him that while there is a clear conceptual distinction, “the two enterprises are heavily intertwined”: We cannot compare languages if we haven’t previously described individual languages, and our comparative knowledge can give us ideas about what to look for in particular languages (comparison can “inspire us”, as I put it in Haspelmath 2020). Bickel then cites the

examples (i) of labiodental consonants emerging in agricultural societies and (ii) of ambient humidity favouring tone, and he raises the objection that these are general insights that are not based on universals. However, he himself offers the correct solution to the apparent problem: “One can ... recast these findings in the form of conditional universals (e.g., “for each language, if its speakers have mostly overbite and overjet tooth configuration, it is more likely to develop and maintain labiodental sounds”).” This connection, as well as the connection between humidity and tone (and other explanations of non-universal linguistic traits by non-universal non-linguistic traits), has the same kind of “universal rule” character as the Greenbergian implicational universals (cf. Brown (2004: 48-49) on the analogous distinction between universals and “universal rules” in anthropology).

Even though my target article takes pains to emphasize the distinction between p-theories and g-theories, D’Alessandro, Mendivil-Giró and Adger largely ignore it. **D’Alessandro** says that “a theory makes predictions, it doesn’t limit itself to descriptions”, but she does not address my point that language-particular descriptions also make predictions, namely about speaker/signer behaviour. So when she (imprecisely) says “a theory”, she means “a general theory”. This point is not unimportant for political reasons, because describers of particular languages (especially small and endangered languages) should not be made to feel that their work does not deserve the prestigious attribute “theoretical”. However, it is also conceptually necessary, because there is no doubt that a grammatical description of a p-language is a theory of this language (which explains the behaviour of its speakers/signers). So the use of “theoretical” instead of “general” is not accurate and can create confusion.

Mendivil-Giró (§2) emphasizes the generativist goal of constructing “theoretical models of these knowledge systems” (i.e. mental p-grammars), something that I have not questioned, but he does not say why these theoretical models should consist of the same building blocks in different languages (e.g. on what basis concepts such as “clitic pronoun” or “noun compound” should be used in different languages with the same sense). He merely says that when such building blocks are not very likely to have been learned from the environment, “it is possible to end up postulating that they are innate” (§5), but he ignores the usual practice of generative grammar, where categories that are readily learnable from the environment are routinely taken to be available to any language (e.g. the distinction between dependent and oblique case, as in Smith et al. 2019). If such categories were not innate, it would not be meaningful to claim that category X that was found in language L_A also exists in language L_B . But linguists make such claims all the time, presupposing a rich set of innate building blocks.

Adger basically equates “theory” and “theoretical linguistics” with “general-theoretical linguistics”, and he reserves “analysis” for language-particular studies. But this is not in line with Chomsky (2017: 2) (“the theory of each individual language is called its Generative Grammar”), and it contradicts a very similar statement in his own earlier book (Adger 2003: 11). Adger wants to distinguish between theoretical and general linguistics: “What makes theoretical linguistics theoretical is that it is concerned with building theories. This is quite distinct from General Linguistics.” But I did not understand what the basis for this is, and what general linguistics is in Adger’s view. Any kind of theoretical endeavor is concerned with building theories, and general linguistics builds general theories. So I still think that my conceptual distinctions (theoretical vs. applied; general vs. particular) are the only sound ones. **Tallman** agrees: If descriptive linguistics is not considered atheoretical by definition (i.e. if theory is not equated with general linguistics), then it is “obviously theoretical, because descriptive claims are testable” (§2), by the usual methods of documentary and descriptive linguistics.

On explanation (and erecting walls)

Zaefferer says that my claim that a descriptive grammar explains the behaviour of speakers/signers is “gravely mistaken”, because he thinks that “in order to really explain these regularities ... one has to go back to their causes” (§3.2). However, one can say that the social conventions are the causes of the speakers’ behaviour. For example, the last word of the preceding sentence (*behaviour*) ends in *-our* (not *-or*) because I follow the British spelling conventions. These conventions are thus the cause of my behaviour, and they can be said to (really) explain its regularity. (This example is from the domain of spelling, but written language is not crucially different from spoken language in its conventionality. Spelling rules are often taught explicitly, but this is not always the case, and it also applies to some widely observed grammatical rules.)

Admittedly, terminological usage varies, but Zaefferer does not seem to like my terminology, and he says that “[Haspelmath’s] concept of explanation is rather different from mine”. But in fact, there are no “conceptual” differences – he just uses the term *explanation* in a more restricted way. This becomes clear when he says:

“I can understand it only as buying into the annoying generative custom of calling syntactic analyses explanations.”

By talking about an “annoying custom” of a community of scholars, Zaefferer erects a wall between him and them, and it is these kinds of walls that I am hoping we can eventually tear down. Indeed, I am here adopting a widespread usage in the field, according to which a description/analysis of a grammatical regularity is a kind of explanation. This implies that descriptive (or p-) linguistics is explanatory and thus theoretical, and again Zaefferer is not happy with this terminology: He prefers to say that only g-linguistics is “purely theoretical”, though he grants that p-linguistics “can be theory-laden”. However, this distinction remain obscure, and it does not accord with the usual practice in large parts of the discipline (especially generative linguistics).

Describing particular languages

Adger says that a theoretical framework (“a theory”) is not only used for analyses in order to test the theory (i.e. to test the proposal that the framework is innate), but that some linguists might want to “use the theory as a fundamentally descriptive tool”. However, he does not say what advantages that might have (as I noted in the target article (§5.2), framework-bound analyses are hard to understand, and they are often preceded by generally comprehensible descriptions anyway). Tellingly, Adger talks about “the theory [the linguist] learned in graduate school”, because this is what happens in practice: Instead of introducing students to general notions, findings and problems of syntax, university courses in syntax typically “teach a theory” together with basic phenomena, concepts, and formalisms (e.g. Adger 2003; Koenenman & Zeijlstra 2015; Börjars et al. 2019). The discipline is not yet very familiar with the alternative idea that general concepts of syntax might be teachable separately from a particular theory (= theoretical framework) of syntax, even though nobody believes that one of the available theories might be the correct one. Thus, while Adger conceptually distinguishes between (i) contributing to a (natural-kinds) theory and (ii) using a conventional framework as a descriptive tool (because of its “wide acceptance”), many linguists are apparently unaware of this crucial distinction. And this leads to the odd situation that many valuable descriptive

contributions are constructed around concepts and ideas that will be without much value in a decade's time, when different ideas will be on the general-theoretical agenda.¹

D'Alessandro highlights a more practical question, namely whether one should approach a new language as if nothing were known:

“Imagine you have to describe an undocumented language... Do you start looking for pronouns or not?... Should one start from scratch for every new language description[?] This [would] mean that all our knowledge is going to be completely useless when describing a new language, because we are not allowed to postulate the existence of a category just because it exists in another language.”

The answer is that one should be inspired by the existing descriptions (including the known cross-linguistic generalizations), but the justification of an analysis must come from language-particular considerations (see also Haspelmath 2020 on the uniqueness of languages, and the value of comparison for description).

Bickel's general perspective is completely different from the generativists, but he also downplays the difference between the conceptual tools for general linguistics and for p-descriptions, saying (in a section heading) that “comparison is description”. But he himself recognizes that comparative research normally focuses “only on one property” (identifiable in all languages), while language-particular analysis often uses categories that are identified by “bundles of properties” that do not necessarily recur across languages (e.g. the English Passive construction, which is defined as including the English Past participle and the English verb *be*). So comparison is not the same as description. But Bickel is of course right that the more properties we take into account, the richer our comparisons become. For instance, if we investigate more than “dominant word order” in our word order typologies, we will know more about word order in general. However, regardless of how much we know about general word order phenomena and principles, language-particular descriptions (such as the peculiar verb-second order in German) will never fall out from this knowledge, and must be stated separately. Likewise, even if we know more and more cross-linguistic details about “patient arguments” and their kin, the idiosyncrasies of the Latin Accusative case will not go away and will call for insightful p-analysis (including uses of the Accusative that have nothing to do with the patient role). Moreover, even if we have more and more fine-grained metalanguages, we will still need coarse-grained comparative concepts, such as “ergative alignment” (e.g. Sauppe et al. 2021). This is as elsewhere in cultural sciences, e.g. in comparative musicology, where very fine-grained comparative concepts (such as the hundreds of distinctions made by the Hornbostel-Sachs classification of musical instruments)² do not make the larger categories superfluous (e.g. “membranophones”, “chordophones”, etc.) P-linguistics and g-linguistics must go hand in hand, and metalanguages must be allowed to evolve and be sensitive to varying goals, but the metalanguages of the two kinds of activities cannot be the same. Etic concepts are different from emic categories, because etic concepts must be measurable or determinable in an independent way, whereas emic categories can only be described as parts of larger culture-specific systems.

¹ I mean contributions such as Rudnev's (2017) paper on Avar reflexive pronouns, Baron's (2019) paper on order in the nominal phrase in Nafara, and Campos Castro's (2020) paper on incorporation in Tenetehára.

² <https://en.wikipedia.org/wiki/Hornbostel%E2%80%93Sachs>

Etic comparative concepts and emic descriptive categories

Zaefferer spends several pages (§5.3) criticizing my earlier distinction between descriptive categories of p-linguistics (emic categories) and comparative (etic) concepts of g-linguistics (Haspelmath 2010; 2018), fearing that the distinction “hampers progress” in typology. In the end, he even asks me (tongue-in-cheek) to “tear down this wall”, and since I do not like walls between people and communities, I decided to take this exhortation seriously, and I incorporated the idea into the title of this response paper.

The lesson that I think I have learned is that walls between people sometimes arise by misunderstandings, and by distinguishing clearly between p-linguistics (with its descriptive categories) and g-linguistics (with its etic or natural-kind categories), we can understand the different practices in comparative grammar as potentially complementary: Some works (e.g. the contributions to the *World atlas of language structures*) rely on etic concepts for comparison, while others (works in the Bakerian tradition) attempt to compare and analyze languages with the same building blocks (hypothesized to be innate, as natural kinds). Zaefferer, by contrast, works with a distinction between a “view of language and languages based on innate factors” and “usage-based approaches that see languages as dynamic systems” (§2). But this is a wrong dichotomy, because everyone agrees that there are some innate factors and that language systems are dynamic in some sense. The dichotomy describes different communities (separated by social walls), but not clearly different methods or “views”. By recognizing that the Greenbergian and Bakerian methodologies are different but possibly complementary, we can tear down the ideological wall between people, because both methodologies could be right at the same time (for different phenomena).

Zaefferer thinks that the practice of current typology is “induced by the noble motive of preventing comparison-based bias in language description” (§7), and he says that “procrustophobia is a bad counselor” (§5.2). But the reality on the ground is something that Zaefferer may not even be aware of: that many younger linguists seem to think that reference-grammar-style descriptions are somehow incomplete and must be complemented with framework-based “analyses” that employ highly technical machinery (cf. note 1). They think that such additional technical treatments are required to make their work “theoretical”, even though it is typically unclear whether any additional predictions follow from them (as **Tallman** notes in his §2). I agree with Zaefferer that the 16th-century Latin-based grammars of German did not do “irreparable damage” (§5.2), but this is a very low bar indeed. We want to do better in the 21st century.

Testing general claims

A general test must involve a representative sample of the world’s languages, so it is not true that “in order to establish whether something is universal it must be checked in all languages”, as **D’Alessandro** thinks. Clearly, studying all existing languages is not possible in practice, and studying all possible languages is not even possible in principle. What we need is predictions that can be tested on any language, and actual tests of these predictions. The predictions must specify “what cannot happen” (as D’Alessandro says), i.e. which logically possible languages are expected not to be found in the world.³

³ Oddly, D’Alessandro’s example for a generalization or prediction that her earlier work has made is “that all Italo-Romance varieties have something in common, namely some extra features”. But this is not a general prediction, as it is a statement about one group of historically related languages (so it is a p-linguistic claim). Presumably, these similarities have a historical explanation, which need not make reference to aspects of human linguisticity.

Adger and Tallman address this central issue of testing general claims with data from a variety of languages, though only Tallman uses the word “testing”. **Adger** describes his collaborative work on Kiowa and says that its value consists in helping to evaluate (= test) competing theories, and I agree: To the extent that the preferred analysis of Kiowa requires a particular set of building blocks (and not others), it supports a theory that proposes these building blocks as innate natural kinds. But this is so only on the 20th century view of generative grammar (where a rich set of building blocks is thought to be innate), and it presupposes that all other languages can be analyzed in a way that supports these analyses. So on this view, the analysis of Kiowa is only one small piece in a larger Bakerian “universals research” programme, and this is fully consistent with what I say. When I said “must be based on universals”, I did not mean that the universals must be established *before* further work can be carried out, but that language-particular work must be taken as contributing to testing universal claims.⁴ So when I said in the paper’s abstract that “one must study universals”, this was an admittedly inadequate formulation (as **Zaefferer** points out in §3.1)

So there is no disagreement here with Adger, and I am in full agreement with **Tallman** that comparative linguists should worry not only about falsifiability in principle, but also about falsifiability in practice (§3). This is often difficult in Bakerian comparative linguistics, but it is not impossible, as Tallman illustrates.

A widespread idea about testing cross-linguistic claims is that one should start out with the hypothesis that some phenomenon is universal and then look for possible disconfirming evidence (e.g. Davis et al. 2014). **D’Alessandro** puts it as follows:

“Isn’t it easier to start from postulating that Italian has no number and no wh-movement based on Chinese, and then look for evidence to the contrary, ... to conclude that there is wh-movement in Italian?”

What she means is that it is probably easier to describe a language by hypothesizing some state of affairs than by starting from scratch for each language, and of course we should not literally start from scratch, but we should be inspired by what we know from other languages (Haspelmath 2020). However, in practice, generative linguists very often find things in languages from around the world that are also present in English (e.g. they find VPs, finiteness, case licensing, movement, even though these phenomena are not apparent in these languages; see Haspelmath 2012). The reverse is much less common, and this should make us suspicious.

The problem is that if we start with the hypothesis that language B has the same category C that we identified in language A, we need to be clear about our criteria (we need to engage in “severe testing”, as **Tallman** puts it). Since the days of structuralism, we have known that different languages have different structures, which means that we cannot easily apply the same criteria in all languages. Scientists know that they should avoid cherry-picking of data, and the corresponding problem in linguistics is what I now call “**criteria selection bias**” (in 2018, I called it “diagnostic fishing”). This problem is much less well known: There is an open-ended range of phenomena that one might take into account in formulating a generalization, and we are often biased toward particular conclusions (for example, in my first p-syntax paper (Haspelmath 1991), I was biased toward finding an ergative construction in Lezgian, and to be sure, I found evidence for it). So there is strictly speaking no alternative to “going back to Aristotle” (who described Greek without being biased by earlier descriptions of other languages, because he was

⁴ Maybe the formulation was not clear enough, so in (B) in the second paragraph above I say: “General claims must be generally testable”, i.e. in principle by studying any language. As Zaefferer puts it (and I agree): “General linguistics must attempt to find and establish empirical universals.”

unaware of the findings of Babylonian linguists). Again, this does not mean that we should not be inspired by what other languages do, but we cannot simply assume that different languages have the same categories when they are manifested differently.

In-depth analyses

D'Alessandro gives the example of subject clitics and claims that if we observed them merely as a “superficial category”, without “looking into their structure and the features they encode”, and “if we did not proceed by reductionism”, the predictions of generative grammar would not be possible. This has indeed been the general view in generative linguistics for a long time: Abstract analyses are not only necessary because we want to capture all the relevant p-generalizations, but also because such analyses make generalizations possible that would otherwise be overlooked. (Essentially this view was articulated by Coopmans (1983) in his highly critical review of Comrie (1981).) Again, this does not contradict what I say, and it follows from the natural-kinds programme (which I have also called the Mendeleyevian Vision): Just as chemists broke up diverse stuffs into their constituent elements and found them recurring in other compounds, linguists have been looking for abstract analyses using recurring elements that are hypothesized to be innate. This was an important part of 20th century linguistics, and there is no conceptual problem with it. The only question is which innate building blocks have been found so far (not many, as far as I am aware, although there have been many unsuccessful attempts).

Has the 20th century natural-kinds programme not been abandoned?

Adger and Mendivil-Giró question whether there has been a “21st century shift” (as I said in §5.1 of the target article), from an earlier phase of generative grammar in which a rich set of substantive universals was assumed, to a new phase, where there are no substantive universals and where the innate component is minimized. They agree that generative grammar is “naturalistic inquiry”, and that the same theoretical entity found in one language can be found in another language, which is my criterion for a natural kind.⁵ But it is clear that the idea of a rich innate toolkit, familiar from quotations like the following, is no longer widely shared.

“That there must be a rich system of a priori properties – of essential linguistic universals – is fairly obvious... It is useful to divide linguistic universals roughly into two categories. There are, first of all, certain “formal universals” that determine the structure of grammars... In addition, there are “substantive universals” that define the sets of elements that may figure in particular grammars.”
(Chomsky & Halle 1968: 4)

Mendivil-Giró agrees that this conception is “incompatible with the minimalist approach”, but he does not say why the “rich grammar blueprint” was so clearly present in Chomsky’s earlier writings. All he says about innateness is that “it would be very surprising if [the biological capacity for language] did not restrict the design space

⁵ A minor point is whether an “operation” such as Merge can be a “kind” (Adger finds this odd), and whether cognitive entities can be natural kinds (Mendivil-Giró observes that they are not made of cells or molecules). I do not see this as an issue, as psychologists have widely discussed whether emotions (anger, surprise, disgust, etc.) are natural kinds or not (Barrett 2006). A natural kind is an entity of naturalistic inquiry that is taken to exist in advance and independently of inquiry. (Strangely, Adger cites Spike (2020), who rejects the very possibility of natural kinds in linguistics, thus taking a clearly anti-generative stance, contrasting with my conciliatory conceptualization of the research methodologies.)

available to children”, without further argument.⁶ He also says that “Chomsky cannot give up assumptions that he has never defended”, but I do not see why. Chomsky may not have “defended” the assumption of “substantive universals that define the sets of elements that may figure in particular grammars”, but he and his followers surely made this assumption for several decades.

Adger’s discussion of rich innate elements is more interesting, because he argues that one can “build richer sets of categories out of poorer ones”, and that we can have fairly minimal sets of units and operations but still maintain the explanatory scope of the theory. The discussion is very abstract, but he gives the nice example of the “rich set of nine building blocks” to express natural numbers: the numerals *one, two, three, four, five, six, seven, eight, and nine*, which can be contrasted with the binary system: 0, 1, 10, 11, 100, 101, and so on. There are only two units in the binary system plus one principle of positional interpretation, so the system is not as rich, but similarly expressive. However, here one might note that it is not sufficiently restrictive, because unlike numerals, the binary system necessarily includes zero (not a natural number, at least not by all definitions). Moreover, in many languages, conventional numerals do not go higher than a thousand or so, but the set of numbers in the binary system is infinite. But in general, there is no question that the goal of reducing a richer set of innate elements to a smaller one is a worthy one.

However, if the general approach no longer posits principles and parameters of the the 20th century, it is difficult to see how it can be as restrictive as that approach claimed to be (e.g. Roberts 1996). For example, classical X-bar theory excludes specifiers and complements that are not maximal phrases, and I would ask: How can an approach that abandons X-bar theory derive the same restrictions? In Haspelmath (2014), I contrasted the Bakerian restrictivist approach with my non-aprioristic approach (based on comparative concepts), but the Bakerian approach is not minimalist. Adger’s perspective says nothing about explaining limits of cross-linguistic variation.

However, in practice, most working grammarians do assume a fairly rich set of generally available (and hence presumably innate) elements. For example, **D’Alessandro** says that “since [wh-movement] is possible, it can be found in languages other than English because we are all human, and we all make use of the same restricted set of grammatical tools”, and this idea of “a restricted set of grammatical tools” stands for the 20th century view. So the tension between the widespread 20th century practice and the 21st century ideology remains unresolved, as far as I can see.

The secondary role of natural-kind explanations

D’Alessandro makes the following interesting comment on the relation between domain-specific natural-kind explanations and more general explanations:

“A formal theory is built through the convergence of a number of proven hypotheses... We reduce [several] phenomena to a more general phenomenon..., until we hit the general “law”... The general law could be due to cognitive requirements, it could have functional explanations. Only if we do not find any such functional explanation or general cognitive explanation can we attribute this “law” to UG.”

⁶ I would not find this surprising, because one can easily imagine that the design space is primarily (or perhaps even entirely) restricted by the communicative function of language. There are obvious communicative limits on language structures, and it is the biological limits that are less obvious (though of course still very plausible).

This is fully in line with what I have called “cost scale of explanatory constraints” (Haspelmath 2019a: §7): Natural-kind explanations are inherently less likely (and thus more costly) than explanations appealing to more general factors, so the latter should be preferred if possible. Natural-kind explanations are secondary in that they come into play when other explanations are not available. This is widely reflected in the practice of non-generative linguists, but for some reason, many generative linguists continue to assume that natural-kind explanations are primary in that they are relevant even when alternative, less costly explanations are available (see, e.g., Bány & Kalin 2020 on differential object marking).

Tacit and explicit knowledge of social rule systems

Forker rightly points out that there are different kinds of social rule systems, and that the control we have over our linguistic rules is somewhat limited. Traffic rules and rules of games such as soccer are taught explicitly and can be changed by authoritative bodies, but languages are different. Authorities can change technical terminology (and spelling rules), but everyday vocabulary and grammatical rules are implicit knowledge (or “tacit knowledge”). However, there are aspects of social behaviour which differ across cultures and which are learned implicitly and are unconscious, e.g. differences in preferred interpersonal distance (Sorokowska et al. 2017). And within linguistic conventions, there are degrees of accessibility to consciousness (e.g. words are more conscious than grammatical markers, concrete markers are more conscious than abstract markers or marking by apophony). Thus, while it is perhaps true that linguistic conventions are particularly remote from consciousness, they are not the only type of socially learned unconscious rule system. (Forker also asks whether languages are abstract objects or not, but I have nothing to contribute to this discussion here.)

Nonconventional and conventional aspects of language

Fuchs & Paschen provide a fascinating discussion of a range of phenomena where the line between conventional and nonconventional aspects of language may not be clear. At first blush, it might seem that conversational turn-taking is universal among humans and not subject to variation, but cross-linguistic studies on turn-taking such as Stivers et al. (2009) have shown that here, too, there are cross-cultural differences (as well as some universal trends). And while I know of no relevant research, it would not be too surprising if slips of the tongue were found to be more frequent in some speech communities than in others, e.g. where speech tempo is higher, cf. Fenk-Oczlon & Fenk 2010). But what would follow from such demonstrations that linguistic behaviour is more culturally variable than one might think?

Clearly, such discoveries would strengthen my central point that we cannot simply study a particular language and draw immediate conclusions about Human Language (what I called the general linguistics paradox). When there is cross-linguistic variability, our general claims must be tested in a general way, i.e. in a wide range of languages and in a representative way. Thus, Fuchs & Paschen’s commentary nicely complements my paper and underlines its main thrust.⁷ There is no need for conventional and nonconventional aspects of language to be a strict dichotomy, because the point is that when linguistic behaviour is (partly) conventional, it cannot serve immediately as data for general theorizing (but must be filtered through a broadly comparative approach). And

⁷ However, I would not say that “child language dictionaries” are indicative of the conventionality of children’s innovative language use. The works cited by Fuchs & Paschen seem to be individual data collections, not recordings of conventions (like ordinary dictionaries).

to the extent that even our brains (or genes) are shaped by our diverse cultural behaviours, it becomes still more important to compare human populations systematically in order to determine what is truly general. Psychologists used to think that their experimental tasks are culture-independent, but since Henrich et al. (2010), they have understood that there may be a problem in psychology as well (compare also cross-cultural research such as Sorokowska et al. on interpersonal distance mentioned above). Behaviour may be culturally variable in a systematic way even when it is not (clearly) conventional, and all such variability challenges simple inferences from one language or culture to the human population in general. We must first look for universals of behaviour.

Social grammars and mental grammars

By “language-particular theoretical linguistics”, I mean both the description of mental grammars (internal knowledge systems) and the description of sets of social conventions (“social grammars”). This is in line with what linguists have done over many decades, though since the 1960s, mental grammars have been emphasized more. **D’Alessandro** imagines a scenario in which “a phenomenon is produced by one speaker of one language (say, Spanish)”, and of course, this is worthy of attention. Every grammatical system is a possible system and must be considered, regardless of how many speakers or signers there are. Worldwide comparative linguistics does not take the number of speakers into account, and of course many languages nowadays have very few speakers left.

Mendívil-Giró emphasizes the difference between mental grammars and social grammars and objects to my proposals on the grounds that generative grammar is not interested in social grammars, and should not be evaluated in these terms. But while this is true in principle, it is not true in practice. Like all other linguists, generative linguists talk about socially defined entities like “Spanish” (e.g. Mendívil-Giró 2021), and “Spanish” is not only used as an abbreviation for the name of the I-language of an idealized speaker-hearer. When confronted with the criticism that the acceptability judgements in syntax are often unreliable, generative linguists typically cite experimental work such as Sprouse & Almeida (2012) which shows that overwhelmingly, the judgements found in the literature stand up to further scrutiny. Now crucially, this further scrutiny involves multiple speakers (or signers) of the same language, i.e. *the same social grammar*. If the purpose were merely to describe a single speaker’s internalized knowledge system, then other speakers (or quantitative evidence) would be irrelevant.

In practice, generative linguists work with the same kinds of data as every other linguist (namely with human behavioural data), and the resulting descriptions have the same status: They describe social regularities or norms, which must of course be internalized in order to be followed (there is of course an internal grammar in each speaker).⁸

Adger asks why he and I end up with rather different conceptions of linguistics despite many shared assumptions, and he speculates that like the philosopher W.V.O. Quine, I “take a grammar to be ultimately a theory of behaviour”, whereas for Chomsky, “a grammar is a theory of an object of the natural world (a cognitive state)”. But this misses a crucial intermediate entity that linguists rarely talk about: systems of social conventions. Language internalization (i.e. acquisition) by a child presupposes not only a vague “input”, but an elaborate system of social conventions. A theoretical social grammar is

⁸ Generative linguists often act as if language users’ acceptability judgements somehow provided direct evidence for their internal knowledge systems, and they use the term “introspection”, as if the users could somehow look inside their heads. In fact, what the users make judgements about is the social (normative) acceptability of an experimental sentence (for more discussion, see this blogpost: <https://dlc.hypotheses.org/2433>).

thus a theory of the system of social linguistic conventions of a speech community, a theory of a “social state”, which corresponds to the “cognitive states” of the individual speakers.⁹ Chomsky subsumes Quine’s approach under “behaviourism”, which may (or may not) be right, and I do not know what Quine meant when he talked about “dispositions to verbal behaviour”, but I see no reason to give primacy to cognitive states over social states. Moreover, since each speaker community is part of the population of humans, which is itself part of the natural world, it is not clear to me why systems of social conventions would somehow not be “objects of the natural world”.

In any event, I do not see how the difference between social grammars and mental grammars might contribute to solving the general linguistics paradox. If an I-language is a cognitive state of an individual, then a theory of that cognitive state does not automatically say anything about Human Language (or human linguisticity). Without broad and systematic study of the variation across the cognitive states of very diverse individuals, we cannot hope to understand its limits (unless we appeal to stimulus poverty, as I noted in §4.1 of my target article).¹⁰

Science as a never-ending conversation

Adger notes that in the kind of analysis-based theory-testing that he advocates (and exemplified with his own analysis of Kiowa complex verbs), “many things are in play at once, and juggling complex data, analyses, and theories is a delicate and difficult business”. Another prominent generative linguist noted with respect to the same approach in phonology that “reconciling typologies with descriptive frameworks and the analyses dictated by them can involve a labyrinth of choices” (Kiparsky 2018: 54). As **Tallman** notes, this often involves “fishing” for diagnostics, and “there are too many degrees of freedom in the interpretation of the baseline theoretical categories for us to easily test the hypotheses in practice” (§4).

If we were sure that there must be innate building blocks, then this arduous path would of course still be worthwhile, because it would be the only possibility. But the “independent-yardstick” approach (what I called “non-aprioristic” in Haspelmath 2014, and “measurement uniformity” in Haspelmath 2019b: §4) is another option that one can adopt in comparative grammar research and that avoids the kind of “juggling” and “labyrinth” that generative grammar implies.

Adger is of course right that “science is a never-ending conversation”, but we do sometimes see intermediate successes, which can be recognized as such when they lead to practical applications, or to further insights that then lead to applications. In linguistics, there is not much of this (though the introduction of the IPA in the 1890s has probably led to pedagogical progress), so it is hard to use “success” as a criterion. According to Adger, “the criteria for success of a programme are whether it opens up new empirical phenomena for study and provides insight into the object of study”, but this is a very low bar. “Insight” is mostly subjective, and “opening up new phenomena” can happen in a wide variety of ways. Thus, general and theoretical linguists will probably continue to disagree about the most promising methodologies, but perhaps they will at least

⁹ **Mendivil-Giró** says that “languages are complex cognitive objects in which biology and culture are mixed”, but one could equally say that languages are *complex social objects* in which cognition and culture are mixed.

¹⁰ **Zaefferer** also mentions idiolects (§5.1), which are perhaps similar to I-languages. But I find his assertion puzzling that “inferences from observable manifestations of a native speaker’s idiolect to a particular language are far more problematic than the step from a particular language to language per se”. Idiolects of individual speakers always largely conform to a social norm, whereas the forms and rules of p-languages are not in an obvious relationship to language universals.

eventually be able to agree on the meanings of basic terms such as “theory”, “explanation”, “grammar”, “analysis”, “description”, “general linguistics”, and “capacity for language” (*linguisticity*). If we continue to use these labels in drastically different ways, there is little hope that the walls that separate the different communities will come down.

References

- Adger, David. 2003. *Core syntax: A minimalist approach*. Oxford: Oxford University Press.
- Baker, Mark C. 2001. *The atoms of language*. New York: Basic Books.
- Bárány, András & Laura Kalin. 2020. Introduction. In Bárány, András & Kalin, Laura (eds.), *Case, agreement, and their interactions: New perspectives on differential argument marking*, 1–26. Berlin: De Gruyter Mouton. (<https://doi.org/10.1515/9783110666137-003>)
- Baron, Bertille. 2019. DP-internal structure and agreement in Nafara. In Lotven, Samson & Bongiovanni, Silvina & Weirich, Phillip & Botne, Robert & Obeng, Samuel Gyasi (eds.), *African linguistics across the disciplines: Selected papers from the 48th Annual Conference on African Linguistics* (Contemporary African Linguistics). Berlin: Language Science Press. (<https://langsci-press.org/catalog/book/226>)
- Barrett, Lisa Feldman. 2006. Are emotions natural kinds? *Perspectives on Psychological Science* 1(1). 28–58. (doi:10.1111/j.1745-6916.2006.00003.x)
- Börjars, Kersti & Rachel Nordlinger & Louisa Sadler. 2019. *Lexical-Functional Grammar: An introduction*. Cambridge: Cambridge University Press.
- Brown, Donald E. 2004. Human universals, human nature & human culture. *Daedalus*. MIT Press 133(4). 47–54.
- Campos Castro, Ricardo. 2020. Incorporação nominal e aspecto lexical em Tenetehára (Tupí-Guaraní). *LIAMES: Línguas Indígenas Americanas* 20. e020012–e020012. (doi:10.20396/liames.v20i0.8659153)
- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Chomsky, Noam. 2017. The Galilean challenge: Architecture and evolution of language. *Journal of Physics: Conference Series* 880. 012015. (doi:10.1088/1742-6596/880/1/012015)
- Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.
- Coopmans, Peter. 1983. Review of Comrie, Bernard (1981) “Language universals and linguistic typology.” *Journal of Linguistics* 19(2). 455–473.
- Croft, William. 2003. *Typology and universals*. 2nd edition. Cambridge: Cambridge University Press.
- Davis, Henry & Carrie Gillon & Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90(4). e180–e226.
- Fenk-Oczlon, Gertraud & August Fenk. 2010. Measuring basic tempo across languages and some implications for speech rhythm. *INTERSPEECH-2010*, 1537–1540. (doi:10.1353/lan.2014.0076)
- Haspelmath, Martin. 1991. On the question of deep ergativity: The evidence from Lezgian. 44/45(1–2). 5–27. (doi:10.5281/zenodo.225289)
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. (doi:10.1353/lan.2010.0021)

- Haspelmath, Martin. 2012. Escaping ethnocentrism in the study of word-class universals. *Theoretical Linguistics* 38(1–2). (doi:10.1515/tl-2012-0004)
- Haspelmath, Martin. 2014. Comparative syntax. In Carnie, Andrew & Sato, Yosuke & Siddiqi, Dan (eds.), *The Routledge handbook of syntax*, 490–508. London: Routledge. (<https://zenodo.org/record/344909>)
- Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different. In Van Olmen, Daniël & Mortelmans, Tanja & Brisard, Frank (eds.), *Aspects of linguistic variation: Studies in honor of Johan van der Auwera*, 83–113. Berlin: De Gruyter Mouton. (<https://zenodo.org/record/3519206>)
- Haspelmath, Martin. 2019a. Can cross-linguistic regularities be explained by constraints on change? In Schmidtke-Bode, Karsten & Levshina, Natalia & Michaelis, Susanne Maria & Seržant, Ilja A. (eds.), *Competing explanations in linguistic typology*, 1–23. Berlin: Language Science Press. (<http://langsci-press.org/catalog/book/220>)
- Haspelmath, Martin. 2019b. Ergativity and depth of analysis. *Rhema* 2019(4). 108–130. (doi:10.31862/2500-2953-2019-4-108-130)
- Haspelmath, Martin. 2020. The structural uniqueness of languages and the value of comparison for description. *Asian Languages and Linguistics* 1(2). 346–366.
- Henrich, Joseph & Steven J. Heine & Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2–3). 61–83. (doi:10.1017/S0140525X0999152X)
- Kiparsky, Paul. 2018. Formal and empirical issues in phonological typology. *Phonological typology*, vol. 23. Berlin: De Gruyter Mouton. (doi:10.1515/9783110451931-003)
- Koenen, Olaf & Hedde Zeijlstra. 2017. *Syntax*. Cambridge: Cambridge University Press.
- Mendivil-Giró, José-Luis. 2021. Inflection, derivation and compounding: Issues of delimitation. In Fábregas, Antonio & Cuervo, María Cristina & Pujol, Isabel & Acedo-Matellán, Víctor & Armstrong, Grant (eds.), *The Routledge handbook of Spanish morphology*. Abingdon: Routledge, to appear.
- Roberts, Ian. 1996. *Comparative syntax*. London: Edward Arnold.
- Rudnev, Pavel. 2017. Minimal pronouns, logophoricity and long-distance reflexivisation in Avar. *Studia Linguistica* 71(1–2). 154–177. (doi:10.1111/stul.12066)
- Sauppe, Sebastian et al. 2021. Neural signatures of syntactic variation in speech planning. *PLOS Biology*. Public Library of Science 19(1). e3001038. (doi:10.1371/journal.pbio.3001038)
- Smith, Peter W. et al. 2019. Case and number suppletion in pronouns. *Natural Language & Linguistic Theory* 37(3). 1029–1101. (doi:10.1007/s11049-018-9425-0)
- Sorokowska, Agnieszka et al. 2017. Preferred interpersonal distances: A global comparison. *Journal of Cross-Cultural Psychology* 48(4). 577–592. (doi:10.1177/0022022117698039)
- Spike, Matthew. 2020. Fifty shades of grue: Indeterminate categories and induction in and out of the language sciences. *Linguistic Typology*. De Gruyter Mouton 24(3). 465–488. (doi:10.1515/lingty-2020-2061)
- Sprouse, Jon & Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger’s Core Syntax1. *Journal of Linguistics*. Cambridge University Press 48(3). 609–652. (doi:10.1017/S0022226712000011)
- Stivers, Tanya et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106(26). 10587–10592. (doi:10.1073/pnas.0903616106)