1. Introduction. I fully agree with Davis, Gillon, and Matthewson’s target article (DG&M; 2014) on some of their key points: (i) that sophisticated first-hand work on small languages should be a priority for contemporary linguistics, (ii) that hypothesis-driven elicitation is a very important technique of descriptive (= language-specific) linguistics, (iii) that many regularities of language are difficult to discover on the basis of small or moderate-sized corpora, and (iv) that comparative linguistics is not linked to a particular (meta)theoretical approach, and that generative linguists have made very important contributions to this field.

So where is the controversy? I think that the main problem is that DG&M have not framed the divisions in our field (as highlighted in Levinson & Evans 2010) in the right way. It is not hard to see that linguists who work on linguistic diversity tend to fall into two very rough sociological groups: those who are more likely to attend conferences like the Association for Linguistic Typology and publish in journals like Linguistic Discovery, and those who are more likely to attend the West Coast Conference on Formal Linguistics (WCCFL) and publish in Natural Language and Linguistic Theory (NLLT). But what kind of intellectual difference, if any, corresponds to this grouping?

My feeling is that the intellectual difference between the two sociological groups is not well understood in our field, and that many linguists who tend to hang out in one of the groups more than in the other are perhaps not committed to a particular intellectual orientation. This may be particularly so with linguists who work in depth on a few small, little-studied languages. In previous work (Haspelmath 2014), I have argued that the most interesting intellectual difference is between linguists who are committed to an aprioristic approach (working with crosslinguistic categories that are given in advance by the restrictive framework) and linguists who have no such commitment and are open to discovering completely new categories, and who are also open to diverse ways of explaining the generalizations they find. It seems to me that to a large extent, DG&M’s work (and the other work they report) falls into this latter category. However, they adopt one key idea from the aprioristic approach: that descriptive hypothesis testing is the same as comparative hypothesis testing. I discuss this assumption in §2 and argue that descriptive hypothesis testing is very different from comparative hypothesis testing.

That DG&M are critical of large-scale comparison as epitomized by the World atlas of language structures (WALS; Haspelmath et al. 2005) is not surprising, as we have received a fair amount of criticism for supposed inaccuracies from descriptive linguists. I address their criticism of WALS in §3. DG&M reject the idea that a special set of comparative concepts must be used for crosslinguistic comparison (see their n. 5, p. e185), but I would argue that their own work also appeals to such special comparative con-

---

1 DG&M use the term ‘minority language’, but I prefer small language (or little-studied language, where this means that not more than a dozen linguists have done serious work on them), because many ‘minority languages’ are quite large, and in some countries like India or Nigeria it is not clear that there are any ‘majority languages’.

Printed with the permission of Martin Haspelmath. © 2014.
cepts (see §4). I end by making a few more comments on the relation between socio-
logical and intellectual divisions within the field of diversity linguistics (§5).

2. Descriptive vs. comparative hypothesis testing.

2.1. Descriptive hypothesis testing. DG&M rightly observe that a thorough de-
scription (or analysis) of a language system requires experimental as well as observa-
tional evidence: to determine the precise boundaries of a generalization, negative
judgments are very helpful and sometimes indispensable. Hypotheses about generaliza-
tions must be checked and replaced by new hypotheses, these must be checked again,
and so on.

But it is not necessary to base the hypotheses on what is known from other languages,
as DG&M seem to imply. In each of their case studies, they start with a hypothesis that
derives from another language, generally English. In case studies 1 and 2, they start with
the idea that their Pacific Northwest languages are like or unlike all other languages, and
in case studies 3–5, they start with the idea that their languages are like or unlike English.
Of course, this is the simplest method, but it is also perfectly possible to examine a range
of data (from a small corpus) and to form an initial hypothesis that is not based in any way
on other languages. Linguists have again and again discovered completely new phe-
nomena, for which no positive or negative model existed, either in crosslinguistically
general phenomena or in phenomena well known from a prestige language: the inclu-
sive/exclusive distinction (unknown in European languages and first discovered around
1550 in Quechua), grammatical evidentials, verbal focus inflection, ejective consonants,
and so on. Describing these categories required the formulation and testing of hypothe-
ses, but no reference to other languages or to some general theory or substantive frame-
work. Thus, linguistic analysis can be truly framework-free (Haspelmath 2010b).

Surprisingly, DG&M do not seem to distinguish between descriptive and compara-
tive hypothesis testing: most of the time in their discussion of methodology they talk
about the analysis or description of particular languages, but in §2.2 they say that ‘some
hypothesis must be adopted at or near the beginning of any crosslinguistic inves-
tigation’ (p. e184). So they seem to see fieldwork as part of a crosslinguistic investiga-

tion, that is, as a comparative rather than a descriptive enterprise.2 This is in line with the
widespread view in generative grammar that analyzing a particular language and com-
paring diverse languages is really the same enterprise: in both cases, the goal is to char-
acterize the linguistic knowledge of a speaker and how it developed from the initial
state of acquisition, which is universal grammar (our a priori knowledge of grammar).
In this respect, DG&M talk like typical aprioristic Chomskyan.

However, while nobody would deny that knowledge of languages from around the
world is very useful for fieldworkers,3 it would be a serious omission if hypothesis test-
ing in fieldwork were limited to phenomena that have some crosslinguistic motivation.
Languages have many phenomena that cannot be easily compared with similar phe-
nomena in other languages, and these need to be described just as thoroughly as the
more readily comparable phenomena. Think of German discourse particles like denn,
ja, or doch: describing these in semantic-pragmatic terms has proved a major challenge,
despite a lot of hypothesis-testing work by descriptive linguists, and crosslinguistic

2 Note that I regard both descriptive (= language-specific) and comparative work as theoretical (if it is car-
rried out without concern for practical applications), so I reject the dichotomy between ‘descriptive’ and ‘the-
oretical’ that is sometimes made (cf. Haspelmath 2010b:§2.3).

3 Shopen’s (2007) set of three volumes on comparative syntax was specifically written to assist field-
workers.
comparisons has not helped. Languages like Latin, French, or English do not have much that is analogous to these particles, and Gitksan and Nsyílxcen may not be of any help either. Thus, linguists who conflate descriptive hypothesis testing with comparative hypothesis testing run the risk of ignoring phenomena that have no counterparts in other languages that we know about.

2.2. Comparative hypothesis testing. Testing a comparative or crosslinguistic hypothesis is an entirely different matter from testing a language-specific hypothesis. While the latter involves a generalization within a language system that the speakers must have internalized in some form, crosslinguistic hypothesis testing involves generalizations over sets of human languages, and these generalizations have no direct link to the linguistic knowledge of individual speakers. Since the majority of generalizations with global scope that linguists find interesting seem to have exceptions (cf. Dryer 1997), testing their validity often involves large samples of languages and difficult issues of representativeness and statistical significance (e.g. Dryer 1989, 1992, Bickel 2013). This presupposes that we have data from a wide variety of languages, and so far the most practical method for arriving at such data has been the grammar-mining method practiced by most WALS authors: scanning existing comprehensive grammatical descriptions or grammar sketches for the relevant information.

DG&M seem to present their fieldwork-based method as an alternative to the grammar-mining method, but it is only in one highly specific situation that in-depth fieldwork on one or a few languages can give insight into larger crosslinguistic issues: when the research question is whether a proposed generalization is an absolute universal. Thus, in a context where the ban on c-commanded pronoun antecedents (DG&M’s 9, p. e191, an aspect of ‘principle C’) is widely taken as an absolute universal, it is of course very interesting that there seem to be a few languages that do not conform to this generalization. But this is certainly not a common situation in fieldwork, because most phenomena are not exceptions, and there are very few absolute universals to begin with. (It may be the most prestigious situation, however, which is most likely to give the fieldworker a lot of attention.)

Thus for most larger generalizations—generalizations with global scope or generalizations about interesting geographical distributions—there is still no good alternative to the Greenbergian grammar-mining approach. In practice, almost all of the hypothesis testing done by fieldworkers will be descriptive hypothesis testing.

2.3. Comparison need not be hypothesis-driven. Language description must aim to be exhaustive, and hence must consist entirely of universal statements (‘All verbs of the second inflection class … ’; ‘all postverbal noun phrases … ’). Thus, describers must set up hypotheses of universal scope and test these by systematic corpus observation and experiments (i.e. elicitation). For this reason, description must adopt a hypothesis-testing approach, and in principle requires negative evidence.

Crosslinguistic comparison is different, because it does not aim for exhaustiveness: we cannot compare all languages, and we are not just interested in universal statements that make predictions about all languages. Often we are interested in areal generalizations (e.g. ‘Lexical tone is far more widespread in Africa than South America’), and existential statements are interesting as well (e.g. case assignment can be an unbounded rule in some languages; Vainikka & Brattico 2014). Thus, comparison is not necessarily hypothesis-driven. We can simply note a distinction between two languages A and B, and ask which other languages are like A and which languages are like B.4

4 One could, of course, reformulate every statement as a test of some hypothesis. Thus, the statement that Amharic is an SOV language like Turkish can be reformulated as the falsification of the hypothesis that...
3. Large-scale comparison. DG&M are of course right that ‘[t]ypology can only be as good as its primary sources’ (p. e216), but on what basis do they assert that ‘[d]escriptive grammars are (naturally) not typically based on targeted fieldwork’ (p. e216)? It is true that a few authors have claimed that their grammars are largely corpus-based (e.g. Heath 1984), but the usual approach in grammar writing has always been to combine elicitation with corpus-based work. This is what all the handbooks say (e.g. Ameka et al. 2006, Chelliah & de Reuse 2011), so why do DG&M imply that the sources of large-scale comparative works like the WALS chapters do not rely on hypothesis testing?

It may be that the kind of hypothesis testing that is routinely done when linguists work out inflectional paradigms and gender categories is seen as somewhat trivial by DG&M and not considered worthy of being called ‘hypothesis testing’. But it is not at all different in nature from what they describe. Corpora are rarely sufficient to give us evidence of all of the cells in inflectional paradigms and all of the inflection classes, and textual evidence for gender categories is often insufficient, so linguists need to resort to elicitation. The situation is not different when it comes to describing copula constructions, relative clauses, or adnominal possession. Most grammars contain a wealth of information about these topics, and this information can be used by grammar-mining approaches to crosslinguistic comparison.

Of course, monograph grammars cannot be sufficiently detailed to answer all of the questions a comparative linguist might have, and the distinction between cumulative and proportional readings of quantifiers has not yet made it into the canon of topics that a grammar (or perhaps rather dictionary, in the case of quantifiers) is supposed to cover. But this canon is getting broader and broader, and grammars have been getting better and better. So the grammar-mining approach will continue to hold a lot of promise, especially since it is very easy and cheap to do (compared to the alternative of asking a range of linguists who are experts on diverse languages to study a particular phenomenon from the same perspective)\(^5\).

DG&M also criticize WALS for ‘overlooking diversity’ in the case of the modality chapter (van der Auwerda & Ammann 2005). They are, of course, right that the three-way distinction in the WALS chapter on epistemic and situational modality ‘misses significant differences among languages’ (p. e214), and their additional three subtypes in 82 (p. e215) could have been distinguished. But it is important to recognize that this is true of all WALS chapters, and indeed of all language comparisons, so it is not a ‘problem’. One can never include in one’s comparison all of the distinctions that someone else might have found relevant.\(^6\) DG&M include necessity and possibility in their (very interesting) comparison, but as the label ‘(weak) necessity’ in their Table 3 indicates (p. e213), there are further distinctions of modal force, and likewise of modality type. Of course one can split further and further, but one also needs to lump if one wants to see any patterns at all. And it is simply not true that by adopting a formal-semantics approach, one does not ‘overlook’ any distinctions. There are many dimensions of semantic variation that formal semanticists have not (yet) taken any interest in (and as long as

---

Amharic is an SVO language like English. DG&M’s procedure in 76a,b (p. e212) is very much like this. The content of the notion of hypothesis testing becomes rather empty in this way.

\(^5\) Specialist-based typology also exists (e.g. Hartmann et al. 2013, Michaelis et al. 2013), but it requires a far greater organizational effort.

\(^6\) Thus, it is not ‘a serious problem’ that there is no category for languages whose determiners encode distinctions other than (in)definiteness in Dryer’s WALS chapters on articles (DG&M’s n. 34, p. e205). If enough information were available, nothing would prevent us from adding this information in a next step.
formal semanticists do not claim that their approach is all there is to say about semantics, this is not a problem.

As for the alleged inaccuracies of WALS, a rough estimate is that at most a quarter of all cases where critics have claimed an inaccuracy are due to real errors. In the majority of cases, reporters of errors simply did not understand the formulation of the comparative concepts well. This seems to be due to the fact that many linguists seem to take the short labels for the WALS values as definitions, but in order to understand the claims made by the WALS classifications, one has to read the chapter texts very carefully. There are many different ways in which languages can be classified into types by comparative concepts, and there is often no obvious choice of parameters.

It is understandable that descriptive linguists like DG&M are sometimes frustrated by perceived shortcomings of large-scale comparative work, also because the latter tends to get a lot of attention, and there is a danger that general linguists might not sufficiently appreciate the hard work of descriptivists on the ground. But comparative linguistics will continue to be done, and the descriptivists will want the comparativists to make use of their work. So as DG&M say, the only solution is to produce better and better descriptive work, and to base our comparisons on the best work that is available. It would make little sense to praise large-scale comparative work as a better way of discovering linguistic diversity. It is good at giving us a bird’s eye view of some aspects of diversity, but it is complementary to the approach that DG&M advocate, and not in competition with it.

4. A PRIORISTIC CATEGORIES AND THE NEED FOR SPECIAL COMPARATIVE CONCEPTS.
DG&M stress that their approach ‘does not force an unfamiliar language into a eurocentric mold’ (p. e218), and this is indeed mostly the case for their semantic case studies 3–5. But for their syntactic case studies 1 (‘principle C’) and 2 (noun-verb distinction), many would disagree. The ‘independence condition’ in 9 formulates the restriction on antecedent-pronoun relationships in terms of c-command, but in many languages, there is no independent evidence for VP, so it cannot be claimed without circularity that the subject c-commands the object (but since the violations that DG&M discuss concern relationships between main and subordinate clauses, not between subject and object, it would be possible to reformulate the independence condition in less eurocentric terms).

The most striking case of a eurocentric analysis is the discussion of the noun-verb distinction in case study 2. DG&M’s argumentation here is very similar to Chung’s (2012) claims for Chamorro. Their examples 15–22 (p. e196) clearly show that St’át’imcets is interestingly different from English in that it exhibits predicate-argument flexibility. This is what earlier linguists had in mind when they said that St’át’imcets has a single ‘verb-noun’ class, rather than a ‘noun’ and a ‘verb’ class. This is an interesting typological difference, which does not disappear just because of the realization that there are specific contexts in which two subclasses of the verb-noun class need to be distinguished. It is well known that grammatical classes typically have subclasses (and sub-subclasses), and there is usually quite a bit of cross-classification. Hence there is no difference of substance between saying that St’át’imcets has a single verb-noun class with subclasses and a noun class that is distinct from a verb class. It is only if one assumes aprioristic cate-

---

7 Public comments on WALS online can be found at http://blog.wals.info/ under ‘Recent comments’.

8 Note that when Levinson and Evans (2010:2747) say that ‘the explosion of new research based on new questions has been spectacular’ (cited in n. 46, p. e216), they are referring to quantitative crosslinguistic (and cross-population) studies such as Dediu & Ladd 2007, which were not possible before large-scale databases like WALS were available. This should not be controversial, even though one may of course remain skeptical about the validity of the results.
gories that the difference matters (the arguments in Haspelmath 2012b against Chung’s claims thus apply here in exactly the same way; see also Haspelmath 2012a).

Instead of assuming preestablished crosslinguistic categories, rigorous comparative linguistics makes use of a special set of COMPARATIVE CONCEPTS, that is, concepts that one would not need for analyzing the patterns of a language, but that are indispensable for comparing languages (Haspelmath 2010a). And indeed, DG&M make use of such concepts, implicitly or explicitly.

Implicitly, they make use of the comparative concepts ‘thing/person’ and ‘action/process’ (as well as ‘property’), because otherwise they would have no basis for saying that the class of words that includes ‘bird’ and ‘chief’ is the noun (sub)class, and that the class of words that includes ‘arrive’ and ‘see’ is the verb (sub)class. Explicitly, they use the comparative concept of ‘determiners’, which they define as ‘elements that introduce argument noun phrases and cannot occur on their own (in contrast to demonstratives)’ (p. e200). While aprioristic categories need not be defined but can be diagnosed by their symptoms (Zwicky 1985, Haspelmath 2015), comparative concepts need to be defined, so DG&M’s definition of ‘determiner’ shows that they are not taking it as a category of universal grammar.9

Other important comparative concepts that DG&M use are ‘pronoun/antecedent’, ‘predicate/argument’, ‘definite’, and ‘quantifier’. While one could imagine restrictive frameworks in which these concepts have crucial roles as a priori categories, in practice they are comparative concepts that every linguist understands but that are not expected to correspond closely to language-specific categories (for this reason, DG&M did not have to specify the framework that they assume; comparative concepts typically work well across formal frameworks).

Thus, even though DG&M officially subscribe to the mainstream generative view that there is a single set of universally available categories that all languages make use of (cf. their rejection of comparative concepts in n. 5), in practice their successful work does make use of such comparative concepts.

5. TYPES OF DIVERSITY LINGUISTS. DG&M say that they ‘draw equal inspiration from Boas and Chomsky’ (p. e217), and this can indeed be seen very clearly in their work. The idea that there is an a priori set of categories comes from Chomsky (Boas would have rejected it), while the idea that studying small languages is valuable because it gives us access to previously unsuspected diversity comes from Boas (it is not clear whether Chomsky would care). When Levinson and Evans (2010:2734) note the lack of interest in diversity among some Chomskyan linguists, they clearly do not have in mind fieldworkers pursuing DG&M’s program. So why do DG&M interpret Levinson and Evans’s program as an attack on their approach? This may have to do with Levinson and Evans’s distinction between ‘C-linguists’ (abstract minimalists) and ‘D-linguists’ (most others), where ‘C’ appears to include everyone working in the Chomskyan tradition, that is, all linguists who attend WCCFL and submit their work to NLLT. But since the 1960s, many linguists who do not necessarily share Chomsky’s philosophical concerns with the nature of human cognition have been working in this tradition, and this group has been grow-

9 Note that in English, demonstratives are normally categorized as Determiners, because they occur in the same slot as articles, regardless of whether they can occur on their own. (In fact, this was the original reason for creating the term determiner, distinct from article.) Thus, not all Determiners in English would count as determiners in DG&M’s sense—that is, this is a clear case of the disparity between descriptive categories and comparative concepts that was highlighted in Haspelmath 2010a.
ing and growing. So it is natural that it now also includes many Boasians, such as Davis, Gillon, and Matthewson.

DG&M themselves say that most generative fieldworkers do not work within the minimalist program in the narrow sense, and this presumably includes them as well. So in what sense are they Chomskyans? In their n. 1 (p. e182), they mention three ingredients that they regard as crucial: (i) an approach to fieldwork ‘in which predictions are generated by an explicit formal model and then tested’, (ii) ‘a commitment to the use of grammatical intuitions’, and (iii) a view of the grammar ‘as a set of representations in the minds of individual speakers’. But we saw that (i) and (ii) are not at all specific to linguists working in the Chomskyan tradition, and (iii) seems to be fairly irrelevant to most fieldworkers’ concerns (it also conflicts with the roots of formal semantics, which lie in explicitly noncognitive philosophical considerations by authors like Gottlob Frege and Alfred Tarski).

My feeling is that DG&M’s work is very non-Chomskyan in that it does not seem to have any concern for universals of human cognition. Quite unlike other generative critics of Evans and Levinson (2009) who have defended the program of a restrictive theory of universal grammar (e.g. Abels & Neeleman 2010), DG&M repeatedly emphasize their interest in discovering diversity, and they highlight several kinds of counterevidence to universal claims made by Chomskyans. They make no attempts to salvage whatever might be left of the presumed universals, but seem content with recording the differences between the languages. So their work is not only very Boasian in spirit, but also comes across as similar in spirit to Evans & Levinson 2009, except for case study 2 (noun-verb distinction), which is more on the universalist side (though their conclusions are very similar to the conclusions of Evans & Osada 2005).

But whether one is primarily intrigued by universals or most revels in diversity is a matter of inclinations, not an interesting intellectual difference. It seems to me that the most important intellectual distinction is between categorial universalists and nonapriorists, or between an approach where description and comparison is conflated and an approach where they are regarded as separate enterprises. DG&M do not focus on this distinction at all, simply stating that they adopt the standard Chomskyan aprioristic view. But as we saw in §4, in practice they work with comparative concepts as well, and much of their work could be slightly reformulated in such a way that it would fit into a conference like that of the Association for Linguistic Typology or a journal like Studies in Language. So while I think that they have not framed the divisions in the field in the right way in their target article, their work is mostly quite compatible with the work of nongenerative descriptive and comparative linguists. I look forward to more of this type of work from them and their students.

REFERENCES


Davis, Henry; Carrie Gillon; and Lisa Matthewson. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. Language 90.4.e180–e226.


Hartmann, Iren; Martin Haspelmath; and Bradley Taylor (eds.) 2013. Valency Patterns Leipzig: A database of valency information in 35 languages around the world. Leipzig: MPI for Evolutionary Anthropology. Online: http://valpal.info/.


Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6
D-04105 Leipzig, Germany
[haspelmath@eva.mpg.de]