Breadth versus depth: Theoretical reasons for system-independent comparison of languages

MARTIN HASPELMATH
Max Planck Institute for Evolutionary Anthropology

draft, June 2023

This paper examines the question of breadth vs. depth of description in theoretical and comparative grammar. I argue that there is no reason why comparison of languages with the aim of finding universal properties should have to be based on wide-ranging ("deep") distributional descriptions. Large-scale comparisons must first of all be based on uniformly applicable, system-independent concepts for comparison. By contrast, "deep" and wide-ranging language-particular descriptions are ideally based on language-particular categories (p-categories), in the spirit of "Boasian methodological relativism".

1. Introduction

This contribution argues that "analytical depth" at the level of particular languages is not necessary for successful general understanding of grammatical properties of Human Language. This goes directly against a widespread view, associated with the generative enterprise, according to which the correct analysis of particular patterns is a crucial prerequisite for finding genuine universals of grammar. If all language systems were made up of the same primitive building blocks, it would of course make good sense to use the same building blocks also for the comparison of languages and for the formulation of universal hypotheses. However, linguists have not been very successful in identifying universal building blocks of grammatical systems, so in practice, they cannot serve as concepts for comparison (yet). Instead, I will argue, comparative research that wants to achieve some amount of breadth must work with system-independent concepts (as it has done in the past, e.g. in well-known broadly comparative projects such as WALS and Grambank: Haspelmath et al. 2005; Skirgård et al. 2023).

I begin in §2 by describing the role of "depth" of description in generative grammar, which is usually presented as closely linked with "depth" of explanation. Questions about general grammar have led generativists to cross-linguistic considerations, but these have rarely resulted in broad worldwide studies. The traditional opposite of "depth" is "surface-orientation", and in §3 I discuss the idea that "surfacy" comparative study of languages does not yield good enough results. This has apparently prevented generativists from extending their research programme to a wide variety of languages worldwide.

But it seems that in practice, "in-depth" study of languages simply means wide-ranging distributional description (§4), which is valued in any kind of ambitious approach to grammatical description. What is peculiar to generative grammar is the idea that distributional description should be done in terms of a set of pre-established categories and features (in a "framework-bound" way, §5). But as I explain in §6, language-particular syntactic distributions must (or should) be described in terms of language-particular categories (p-categories), not in terms of the kinds of general concepts that we need for language comparison. Finally, I note that uniformity of grammatical categories and systems cannot be taken as the default (§7), at least as long as there are no constraints on the kinds of data that can be taken as evidence for analyses in terms of universal categories. I conclude that as long as we have not made good progress on identifying

innate building blocks of universal grammar, comparative grammar research is best served by system-independent comparative concepts, and there is no impediment to broad cross-linguistic comparison.

2. Depth of description and depth of explanation

Broad cross-linguistic comparison at a worldwide scale has been practiced at least since Schmidt (1926), who presented several world maps of grammatical features on the basis of many dozens of languages from around the world. This kind of research on "linguistic typology" and universals became well-known with Greenberg (1963), and the study of grammatical features in breadth has been an important part of general grammar ever since.

It is evident that for practical reasons, a single scholar (or a small team of scholars) cannot have the same depth of knowledge of dozens or hundreds of languages as they can have of one or two languages. Song (2018: 126) notes that there is a kind of trade-off:

> "typological databases may vary in terms of depth or breadth of coverage. Thus, some databases cover a large number of languages but the data that they provide may be restricted to selected structural properties, while others deal with a small number of languages, providing detailed data for each language."

This is a trivial practical observation, but in this paper I make a different point: that "depth" of description or "analysis" is not even desirable or possible for cross-linguistic comparison, because distributional analyses are based on categories within particular systems (§4), while comparison must be system-independent (§6).

Some readers may find this a startling claim, because it goes directly against what has been the mainstream view in generative grammar since the 1980s (when the Principles and Parameters programme became prominent, following Chomsky 1981). That grammatical descriptions should "achieve depth" has been emphasized repeatedly in the generative literature, e.g.

> "a great deal can be learned about UG [= universal grammar] from the study of a single language if such study achieves sufficient depth to put forth rules or principles that have explanatory force but are underdetermined by evidence available to the language learner." (Chomsky 1981: 6)

The notion of "depth" that appears here seems to have two distinct ingredients, which we may want to keep apart: On the one hand, it refers to larger generalizations that would not be found, for example, in a pedagogical grammar that has purely practical goals. A striking instance of this is Chomsky's (1981) application of his Binding Theory not only to reflexive pronouns like *herself*, but also to abstract elements such as "traces of A′ movement" (see Truswell 2014: 218). Thus, Chomsky claims that (1b) and (2b) are unacceptable for the same ("deep") reason: they both violate Principle A ("anaphors must be bound in their governing category"). (In (2), *t* is an unpronounced trace of movement, and the subscript links the fronted element to its trace.)

(1) a. *Susan thinks that Bill adores her.*
    b. **Susan thinks that Bill adores herself.*

(2) a. *Susan$_i$ seems t_i to adore Bill.*
    b. **Susan$_i$ seems that Bill adores t_i.*

This nicely exemplifies the kind of depth of description that is highly prized in the generative community. But Chomsky was (almost) only talking about English in his work on Binding Theory, and it quickly became clear that other languages demand a rather different approach to reflexive pronouns (see, e.g., Reinhart & Reuland 1993). Thus, he was still operating at the level of language-particular description (or "analysis"), and it turned out to be difficult to apply his description of English more generally.

But in addition to descriptive depth, Chomsky and other generativists have often emphasized depth of explanation. A quotation such as the following is quite typical:

> "A reasonable way to answer the question of domain specificity, given the current state of knowledge in cognitive science, is to develop theoretical approaches to linguistic phenomena which have as much empirical reach and explanatory depth as possible." (Adger & Svenonius 2015: §1)

As long as there is no direct claim that the highly general (or "deep") statements of a linguistic analysis can be linked to something outside a linguistic system (such as an innate, genetically specified blueprint for grammar), there is not really any difference between depth of description and depth of explanation: A shallow rule for a particular language (for example, as found in a pedagogical grammar) explains the behaviour of language users and their intuitions about aceptability in the same way as a deep rule does, if less elegantly.

However, since the early 1960s, Chomsky has also emphasized larger concerns: By "explanatory adequacy", he means that linguistic theories should ultimately be able to say which kinds of rules are possible at all in human languages. After 1981, this led Chomsky and other generativists increasingly in the direction of broadly comparative research on grammar, although stimulus poverty arguments have always been very important, too (see Pullum & Scholz 2002; Lasnik & Lidz 2017). The connection between the ultimate explanatory goal and broader, comparative work is summarized nicely as follows:

> "As soon as a grammatical property is ascribed to Universal Grammar on the basis of poverty of stimulus considerations, a hypothesis which can be legitimately formulated on the basis of the study of a single language, a comparative verification is immediately invited… So, in-depth research on individual languages immediately leads to comparative research, through the logical problem of language acquisition and the notion of Universal Grammar." (Belletti & Rizzi 2002: 9)

But generative grammarians never embraced the kind of breadth of worldwide comparison that we have seen in typological research of the Greenbergian tradition, where (to take some examples) Ultan (1978) studied adpossessive constructions in 74 languages, Stassen (1997) studied 410 languages, Dryer (2005) studied 1228 languages, and Grambank includes 2259 languages in its feature "Order of demonstrative and noun" (Skirgård et al. 2023).[1] Baker & McCloskey (2007: 286) compare generative ("formal") and Greenbergian typological research and note that "while crosslinguistic comparison is clearly on the rise in formal work, true typology is not."

But why is this so? We will see that "in-depth" analyses that are dependent on a framework (§5) are not well-suited for large-scale comparisons (§6). But first, we will look at the idea that "surfacy" descriptions are not adequate for the discovery of universal properties (§3).

[1] https://grambank.clld.org/parameters/GB025

3. The alleged problems of "surface orientation"

Generative linguists have often claimed that broadly cross-linguistic studies in the Greenbergian tradition are not very helpful (or even completely irrelevant) for the goal of finding the causes of the apparently deep unity of human languages. An idea in Chomsky's early work was that the deep structures of all human are the same, and this was still reflected in some of the discussion of the 1980s (e.g. Coopmans 1984; Hawkins 1985). Later, when the idea of universal deep structures was given up, generative linguists still emphasized the need to base one's comparative studies on "in-depth" analyses of the languages. The following paragraph from Rizzi (1989) deserves quotation:

> "[The new generative] comparative syntax cannot be done by comparing simple superficial properties of dozens of languages whose grammatical structure is poorly understood. Such a "surfacy" approach to comparison and typology is incompatible with the methodology and goals of the program of generative grammar. Simple superficial properties can be extremely deceptive: even an apparently straightforward task such as the determination of the fundamental word order between verb subject and object can be completely blurred by the existence of interfering factors (for instance the so called Verb Second Constraint in German, Dutch and many other languages). Such factors tend to be overlooked by "surface survey" type approaches, with disastrous consequences for the possible theoretical interpretation of the results. The broadening of the empirical basis of parametric theory (and, I think, of any theoretically sound approach to comparison and typology) can only proceed through the in-depth investigation of the individual grammatical systems to be compared. Only thoroughly analyzed properties of relatively well-understood grammatical systems can be the empirical material of a comparative approach seriously aiming at theoretical purposes." (Rizzi 1989: 74)

While "surface" used to refer to a specific level of representation in a formal grammar (surface structure, Chomsky 1965), it has now become a badge of comparative research that employs surveys of existing grammatical descriptions. A widespread hope is that while non-generative surveys such as WALS (the *World atlas of language structures*, Haspelmath et. al 2005; Dryer & Haspelmath 2013) result in confusing variety, true biocognitive universals will be found once one "thoroughly analyzes" the different languages. For example, Davis et al. (2014: e185) say that "linguistic diversity is often surface-visible, while uniformity is often more subtle, and only detected via hypothesis-driven discovery procedures". Another similar recent statement is Holmberg (2017: 359):

> "The grammatical properties that are investigated / compared [by Greenbergian typologists] are by necessity all easily observable 'surfacy' properties of the kind which are recorded even in sketchy descriptive grammars. One result of this is that the generalizations discovered have been probabilistic, riddled with exceptions, rather than absolute, because surfacy properties are subject to unpredictable variation to a greater extent than more abstract properties."

However, it seems that this is primarily based on a hope rather than on solid research results. Broad comparative research has discovered striking regularities in word order (e.g. the correlation between verb-initial order and prepositions; Greenberg 1963) and core argument marking (e.g. the strong tendency for ergative nominals to be overtly marked; Comrie 1978). On the other hand, subsequent generative research (taking into account a wider range of phenomena) has often found more diversity than was initially expected, e.g. among verb-initial languages (Carnie & Guilfoyle 2000) and among

languages with ergative case (Deal 2015). As a result, large-scale comparative work of the kind exemplified by *WALS* has been quite rare in the generative community.[2]

But the main point that I want to make in here (see §6 below) is that the system-independence of grammatical comparison (which may look like "surface orientation" to generative linguists) is not an unfortunate practical limitation, but a necessary feature of cross-linguistic research, for theoretical reasons.[3] By contrast, the claim that "in-depth" study of particular languages is a better basis for comparative studies has no theoretical basis, as far as I am aware.[4] Rizzi says that "simple superficial properties can be extremely deceptive" because even "the determination of the fundamental word order between verb subject and object can be completely blurred by the existence of interfering factors", but interfering factors are likely to occur in any type of research method. There is no a priori reason to expect greater uniformity if one compares "in-depth" descriptions.

Strikingly, nobody seems to have offered a theoretical characterization of the alleged difference between "detailed, in-depth" studies of grammatical phenomena and "shallow, surfacy" broadly comparative studies, even though this difference has been invoked repeatedly. The original notion of "surface structure" had become quite irrelevant by the 1980s, so it cannot serve as a foundation for the notion of "in-depth" study. Below we will see that the practical differences seem to be primarily due to the theoretically well-understood difference between general (comparative) grammar (g-linguistics) and language-particular analyses (p-linguistics) (see §6).

4. Depth as wide-ranging distributional description

The syntactic systems of human languages exhibit enormous complexity and can hardly be described exhaustively. Even very comprehensive grammars such as Abeillé & Godard (2021) (a 2600-page grammar of French) fall far short of covering all regularities. The complexity arises not so much from the number of basic syntactic constructions, but from the multiplicity of their conceivable interactions. For example, generative linguists have observed very subtle contrasts when question-word fronting in English interacts with a variety of contexts. Among many other factors, extraction from subordinate clauses depends on the meaning of the matrix verb (as seen in (3a-b)) and on the order of the subordinate clause with respect to the verb (as seen in (4a-b); see Hawkins 1999 for discussion).

(3) a. *How angry did Lee say that Kim was?*
b. *?*How angry did Lee whisper that Kim was?*

---

[2] Two notable exceptions are Cinque (1999) and Julien (2002).

[3] I should note here that some Greenbergain typologists have actually agreed with the generative view that language comparison should be based on language-particular analyses (this is very clear in Comrie 1984), and some have agreed that it is for practical reasons that the choice of comparative concepts cannot be based on analyses ("typologists find it useful to develop variables as close to observable data as possible and close to fieldwork. This is first of all a practical decision, because very abstractly defined variables are difficult to survey in sufficiently large samples"; Bickel 2007: 242). I disagree with these assessments by Comrie and Bickel.

[4] As an aside, I note that there does not seem to be any theoretical basis either for thinking that biocognitive constraints (of the "universal grammar" type) must lead to exceptionless universals. According to Roberts (2017: 8), "Greenbergian universals tend to be surface-oriented and tend to have counter-examples… Universals determined by UG have neither of these properties: clearly they must be exceptionless by definition", but I do not see why it should be clear that biocognitive constraints should be strict restrictions rather than violable constraints or weaker preferences (see Haspelmath 2019 for the terms *constraint, preference,* and *restriction*).

(4) a. *Who did it surprise Kim [that Lee disliked _]?*
b. **Who did [that Lee disliked _] surprise Kim?*

In the earlier grammatical literature, such contrasts were not often observed because the syntactic descriptions did not aim to investigate all possible interactions of syntactic constructions, including more far-fetched ones (such as *who* fronting and *that*-clause preposing in (4b)). For inflectional morphology, earlier linguists generally did aim to present complete paradigms (including far-fetched forms such as the second person plural of the past subjunctive), and it was an innovation associated with generative grammar that the paradigm-completing approach was extended to syntax (Haspelmath 2022).

So in practice, "in-depth study" generally means that a descriptive account of a pattern examines more interactions of the pattern with other patterns than some earlier work. "Depth" of description means wide-ranging description that takes a wide range of interactions into account, ideally arriving at larger generalizations. This sometimes yields surprising results, and linguists then typically try to find ways of making the results look less surprising (by suggesting that they might actually follow from some abstract concepts that were previously unknown or neglected, as in (1)-(2) above). But is this kind of research exclusive to the generative community?

As Croft (2001) noted, systematic distributional analysis in syntax was pioneered by American descriptivists (or "structuralists") since the 1930s, and in morphology, it was of course practiced widely in earlier times. Syntax was not as prestigious as phonology and morphology until the 1950s, but there is no reason to associate wide-ranging ("in-depth") study of grammatical patterns exclusively with the generative enterprise.

An important theoretical point made in descriptivist work of the earlier 20th century was that positing syntactic categories for a language must be justified by syntactic patterns of the language, and cannot be assumed on the basis of a presumed universal semantics or on the basis of other better-known languages. That different languages may have different categories had already been recognized in the 19th century, as illustrated by the following passage from Whitney (1875):

> "Every single language has thus its own peculiar framework of established distinctions, its shapes and forms of thought, into which, for the human being who learns that language as his 'mother-tongue', is cast the content and product of the mind, his store of impressions, however acquired, his experience and knowledge of the world. This is what is sometimes called the 'inner form' of language – the shape and cast of thought, as fitted to a certain body of expressions." (Whitney 1875: 21-22; cited from Joseph 2022: 24)

Since Boas (1911), many North Americanists have been aware that each language has its own morphosyntactic categories and therefore must be described "in its own terms" (i.e. using its own criteria for identification), and every fieldworker knows that this methodological point has not lost any of its relevance. Levinson & Evans (2010: 2734) call this principle "Boasian methodological relativism", and despite this grand name, the point is actually uncontroversial.[5] For example, an English Noun is not identified as "a word that denotes people, places and things" (as might be said in primary school), but as a word that can occur with a preceding Definite Article. Morphosyntactic classes are set

[5] Some authors have rightly emphasized the usefulness of cross-linguistic information for working out descriptions of particular languages (e.g. Himmelmann 2022), but strictly speaking, this is limited to inspiration from descriptions of other languages and to the choice of transparent terminology (see Haspelmath 2020: §3-4). In the end, the decisive criteria for setting up categories come from the language itself.

up to describe morphosyntactic distributions, and these are language-particular (i.e. they are p-categories).

Thus, rich in-depth distributional description is in no way peculiar to generative grammar, and in fact, one might say that it is less important in generative linguistics than framework-bound "analysis", as we will see in the next section.

5. Depth as framework-bound "analysis"?

Another way in which "in-depth study" can be understood in a generative context is in terms of the distinction between framework-independent description and framework-bound description (often called "analysis").

What is typical of generative papers on morphosyntax is that the phenomena are described twice: Oncee in a section that is often called "the data" (but that really contains a distributional description), and a second time in a section that is often called "the analysis". This latter section describes the same facts but by means of the abstract notions of a a general formal framework. For example, Maché (2022) first describes the compound patterns of Igbo (and other Benue-Kwa languages) in plain distributional terms, and then in §5 provides an "analysis" in terms of the HPSG framework. Another example is Ting (2023), who describes the N-bonding construction of Malagasy in §2-3 of her paper, before providing an "analysis" in terms of mainstream generative syntax in §4. The idea is that these re-descriptions add to our knowledge of general grammar by providing evidence in favour of, or against, some earlier proposals about universal grammar (see also Haspelmath 2021: §5.2 for discussion). There is little indication that they are gradually leading to a better and better understanding of the innate grammatical structures, but they are often still considered crucial ingredients of papers on language-particular phenomena.

Generative grammarians thus typically contrast "descriptive generalizations" with "analyses", where the latter are re-descriptions in terms of a hypothesized universal set of categories and features. In the next section, I will explain why I think that this approach has not had the success that was hoped for in the 1980s, but here I briefly note that the actual success of work by generative grammarians has been in terms of breadth of coverage rather than in terms of generality (or "depth") of theories (where few proposals have stood the test of time). This is admitted by Baker (2021):

> "the great contribution of Chomskyan linguistics has been to descriptive linguistics, as it has discovered a plethora of genuinely new facts about languages and language. I for one have been inspired by the kinds of discoveries that were being made in English and the Romance languages, and wanted to use a similar precision, sophistication, and attention to detail to make discoveries about Mohawk, Edo, Sakha, and Shipibo. (Baker 2021: 161)"

This new breadth of description was apparently triggered by the focus on completing the syntactic paradigms (as I noted in Haspelmath 2022), not really by Chomsky's specific interest in biocognitive constraints on language (or universal grammar). Thus, breadth may come about through a variety of different motivations that need not have much in common. In the next section, we will get to the main point of the paper, that comparisons must be system-independent.

6. Comparison must be system-independent

As I noted in the introduction, if all language systems were made up of the same primitive building blocks, it would make good sense to use the same building blocks also for the comparison of languages and for the formulation of universal hypotheses. However, linguists have not been very successful in identifying universal building blocks of grammatical systems, so in practice, they cannot serve as concepts for comparison (yet) (Haspelmath 2007).

As a result, comparison of languages requires a distinct set of comparative concepts, a kind of "auxiliary language", as noted by Levinson & Evans:

> "Language-specificity of categories raises problems especially for [cross-linguistic studies], as the typologists have become increasingly aware (Dryer 1997; Haspelmath 2007; 2010a). It doesn't follow that comparison is impossible, only that it has to be undertaken in an auxiliary language designed to generalize over language-specific categories." (Levinson & Evans 2010: 2738)

This is still underappreciated,[6] so let me briefly repeat the main point: Language-particular categories are identified by language-particular criteria, while comparative concepts are defined in the same way for all languages (Haspelmath 2018). This means that it is logically impossible to identify and define them in the same way. There are of course many similarities, and language-particular categories in different languages often have the same name, for good mnemonic reasons (Haspelmath 2020).

For example, English *the* is called Definite Article, and French *le/la/les* is also called Definite Article, and this has good practical reasons. They are similar semantically and syntactically (occurring prenominally, in complementary distribution with prenominal possessor forms), but they are not identified in the same way. The English Definite Article is simply defined as the segment sequence [ðə/ði]), and the French as the three related segment sequences *le, la, les.* When we look at the constructions in which they occur in more detail (or in greater "depth"), we see that they differ in a number of ways, as illustrated in (5a-c) from English and their French counterparts.

(5) English
   a. *Life is good.*
   b. *Bats are mammals.*
   c. ***The*** *sparrow is the most widespread bird.*

(6) French
   a. ***La*** *vie est belle.*
   b. ***Les*** *chauves-souris sont des mammifères.*
   c. ***Le*** *moineau est l'oiseau le plus répandu.*

French requires the Definite Article in all these generic contexts, while English uses the Definite Article only when the generic subject nominal is singular. Thus, what defines them is their shape, and not their function of encoding definiteness. They can be compared using the comparative concept "definite article" (e.g. Becker 2021), but they cannot be said to be the same category.

---

[6] For example, Bošković (2022: 3) says that "[Haspelmath] treats [comparative concepts] as artificial linguistic constructs and not real (in fact not part of the grammar of individual languages), but there is really no deep reason … why they should not be considered real." The text should make it clear that there is a "deep" reason for this.

Most language-particular syntactic classes are not defined in terms of their shape, but ultimately, they are all defined in language-particular terms. For example, German Demonstratives are defined as forms distinct from the Definite Article that occur prenominally but need not cooccur with a German Article or with a prenominal possessor form. This includes not only *dieser* 'this' (and *jener* 'that'), but also *folgender* 'the following', though not *obiger* 'the above', as seen in (7a-c).

(7) German

a. *dieses Problem*	'this problem'
b. *folgendes Problem*	'the following problem'
c. **obiges Problem*	'the above problem' (OK: *das obige Problem*)
d. **großes Problem*	'the big problem' (OK: *das große Problem*)

Cross-linguistically, demonstratives are defined by their deictic function (e.g. Diessel 1999),[7] not by specific syntactic properties such as the cooccurrence with a definite article.

Thus, it is by logical necessity that languages languages cannot be compared by language-particular concepts, and that the concepts for comparison must be independent of particular language systems. (Conversely, languages might be describable by comparative concepts to some extent, but such descriptions are almost always regarded as unsatsifactory by linguists.) It seems that the frequent misunderstandings arise from the failure to draw a consistent distinction between p-linguistics (the study of particular languages) and g-linguistics (the study of Human Language in general; Haspelmath 2021). As languages are similar in many ways, it is often insightful to carry over ideas from the analysis of one language to another language, but this does not mean that the descriptions can make use of the same categories.[8]

7. Uniformity cannot be the default

The "Boasian relativist" idea that each language system should be described with its own p-categories seems unintuitive to many linguists, as experience tells us that languages show substantial similarities, often in striking and surprising ways. Especially (but not only) generative linguists often find their expectation of categorial uniformity confirmed by additional properties of constructions that are not apparent at first glance. This leads to a danger of confirmation bias, and a temptation of cherry-picking the evidence that fits one's expectations. This problem is rarely addressed in the relevant literature.

As a concrete example, consider Faroese Dative Subjects (illustrated in (8)), which exhibit peculiar case-marking (Dative instead of Nominative), but behave like Nominative subjects in a variety of other ways, e.g. in that they can be the omitted argument in infinitival constructions, as seen in (8b).

[7] Actually, the German Definite Article can be used deictically when it is stressed, so there are two ways in which the German Demonstrative category does not coincide with the comparative concept "demonstrative". This is quite typical for the relationship between "category-like" comparative concepts and the language-particular categories that match them and are known by the same name.

[8] Croft (2009; 2010) draws attention to this problem and calls it "methodological opportunism". I agree that this is a problem, but I do not think that it results from sloppiness, as the word "opportunism" might suggest. Rather, it seems to arise from the fundamental (and largely unquestioned) assumption that all languages make use of the same categories.

(8) Faroese (Eythórsson & Barðdal 2005: 839)

a. *Mær vantar pengar.*
me.DAT lacks money.ACC
'I am short of money.' (Lit. 'To me it lacks money.')

b. *Það at Ø vanta pengar, er ikki gott.*
it to Ø.DAT lack money.ACC is not good
'Being short of money is not good.' (Lit. 'It lacking money is not good.')

Since Andrews (1976) and Zaenen et al. (1985), Icelandic and Faroese have typically been analyzed as having subjects with "quirky case" that do not show the typical argument marking properties of subjects, but are nevertheless regarded as subjects because of their other syntactic properties. The properties that Nominative Subjects and Dative Subjects share seem to be similar to properties of English Subjects, so from an English perspective, this conclusion is perhaps understandable. From a more general perspective, one could alternatively say that there is no subject in (10a) because there is no Nominative argument (cf. Haspelmath 2011, where I say that a subject in a general sense is the A-argument of a transitive clause or the S-argument of an intransitive clause, i.e. the Faroese Nominative argument).

So what justifies the claim that Faroese and Icelandic have subjects with dative case? Why is it possible to ignore the two most important features of subjects in older Indo-European languages, namely nominative case and control over verbal person-number marking? Chomsky (2001: 2) states the principle in (9).

(9) Uniformity Principle
In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances.

Given such a heuristic principle, it is easy to conclude that the Dative argument *mær* 'to me' in (8a) is a subject because this preserves uniformity between English and Faroese, and while it stands in the Dative case, this is not "compelling evidence" that it is not a subject. Its Dative marking is "easily detectable", and its more abstract subject status can be supported by further properties such as the existence of infinitival constructions like (8b).

However, adopting the Uniformity Principle is hardly compatible with a rigorous scientific approach. Whether the evidence to the contrary is "compelling" is a subjective matter, and science needs objective criteria. Whether some property of a construction is "easily detectable" is likewise not determinable in a straightforward way. In many or most situations, there are some similarities and some differences between constructions that one might want to compare across languages. Now if it is up to the individual researcher to decide which criteria are relevant and which criteria are left aside, the result is not objective and is likely to be coloured the analyst's expectations.[9] There is a serious danger of Eurocentric bias in syntactic analyses, because there is typically no compelling

[9] It is not uncommon for generative linguists to dispute the relevance of grammatical facts on the grounds that they are not "reliable" as tests (e.g. Woolford 2006: 117: "not all the tests that have been used in the literature are entirely reliable: some are based on incorrect assumptions, some are actually diagnostics for something else, and some can be used with confidence only if one is aware of interfering factors that produce misleading results on these tests in some languages. Once we limit ourselves to data from reliable diagnostic tests, we find no evidence …"). But which tests or diagnostics are considered relevant is up to the individual analyst (Croft 2009: 158 describes the method as allowing "protective belts of proliferating auxiliary hypotheses").

evidence that a non-European language like Chinese or Quechua really lacks such phenomena as subject pronouns, complementizers or finiteness (leading to strange analyses such as "pro-drop", zero complementizers, and abstract finiteness).

Thus, it is a much better methodological principle to demand positive evidence both for cross-linguistic differences and for cross-linguistic similarities. And both must be stated in general terms, using comparative concepts, rather than in terms of system-dependent concepts. The English Subject is a category that is defined with respect to properties of English, and the question which of the arguments in Faroese is a subject in the same sense as the English Subject is not answerable because "subject" is a general concept while the English Subject is a p-catgeory, defined only within the system of English grammar.

8. Conclusion: General linguistics need not be based on particular systems

I conclude that some of the traditional generative views on the methodology of comparative grammar are unjsutified: Depth of language-particular description or analysis or explanation (§2) is not a prerequisite for theoretically oriented general grammar, even though this is often thought to be the case (§3). If "depth" means wide-ranging distributional description (§4), this is not true because the wide range of distributional facts cannot be compared simultaneously anyway. If "depth" means framework-bound analysis (§5), this is not true because analysis can (and should be) be framework-free (Haspelmath 2010b). In fact, comparison must be based on system-independent concepts rather than on language-particular descriptions because the latter must be based on ("Boasian") p-categories to be satisfactory (§6). Uniformity cannot be regarded as a default, because this would not rule out cherry-picking analyses that suit one's expectations (§7).

It is true, of course, that the possible existence of a rich set of innate categories and features of universal grammar cannot be ruled out,[10] and that our lack of robust knowledge of these elements may simply be due to the fact that linguists have not converged on the correct set yet. There are thus still many "moving parts" (Peter Jenks, p.c.) of the theory of universal grammar, but the best way to make progress could be the comparative study of languages as advocated by Rizzi (1989) (quoted in §3 above). Just as researchers discovered the chemical elements in the 19th century by in-depth study of a wide range of chemical compounds, it could be that linguists will eventually discover the set of innate categories and features in this way (see Baker 2001 for the analogy between chemistry and linguistics). In an earlier paper, I called this the Mendeleyevian vision (Haspelmath 2019: 127). The argumentation in the current paper does not identify any deep problems with this vision, but it starts out from the observation that so far, "linguists have not been very successful in identifying universal building blocks of grammatical systems" (see §1 above). On the other hand, there have been many large-scale studies in the Greenbergian tradition that have contributed to our knowledge of the world's languages, and these

[10] The traditional view of a rich innate UG was recently reiterated by Ledgeway & Roberts (2017: 581), with reference to the OV/VO parameter: "On the classical Principles and Parameters view, the notion of 'verb' is given by the universal theory of syntactic categories, the notion of 'object' is given by the universal theory of grammatical functions, and the idea that the two combine to form a VP is given by the universal theory of phrase structure. These are all taken to be reflexes of UG principles." (More recently, many generative syntacticians have questioned the innateness of categories and features, as noted in Haspelmath (2021: §5.1), but this makes the framework-bound approach incoherent.)

studies have relied on system-independent comparative concepts rather than universal (Mendeleyev-style) building blocks.

General theoretical approaches that have proposed explanations for cross-linguistic generalizations have often made reference to system-external factors such as parsing efficiency (e.g. Hawkins 2014) or communicative efficiency (e.g. Gibson et al. 2019; Haspelmath 2021; Levshina 2023), and these explanatory theories do not make crucial reference to specific grammatical structures as discovered by (in-depth) distributional description/analysis of particular languages. They invoke universal causal factors to explain universal trends in grammatical patterns, and they have little or nothing to say about specific language-particular classes and patterns that have no cross-linguistic generality. Thus, they can rely on broad cross-linguistic studies of the Greenbergian type that are based on system-independent comparative concepts rather than system-dependent p-categories.

It seems to me that the main reason for the reluctance of generative linguists to embrace framework-free description and system-independent comparison is not their lack of interest in language use or variability, or their commitment to innateness explanations. After all, many generative linguists work in psycholinguistics (where usage events are studied), many work in dialectology and diachrony, and many have become skeptical about a rich innate universal grammar (following Chomsky's shift of interest toward the problem of evolvability). What seems to be lacking in the dominant mode of thinking is the clear distinction between general linguistics and particular linguistics (g-linguistics and p-linguistics, Haspelmath 2021a). While the Chomskyan distinction between I-language (mental representation of linguistic knowledge in the individual) and E-language (language in use and in society) is well-known, linguists rarely acknowledge that what they actually do most of the time is to describe particular languages as sets of social conventions, rather than as individual knowledge systems. We can try to make inferences about mental knowledge systems (the I-languages) from sets of social conventions (p-languages), but these inferences are very uncertain. In practice, the Chomskyan comparative grammar programme tries to infer the innate universal grammar not from I-languages, but from p-languages, but this simply appears to be too ambitious a programme. In Greenbergian comparative grammar, we are much more modest in that we base our comparisons on what we know about the languages as social systems, not on inferred mental knowledge systems, and we are happy to state universals even if we have no explanations for them. This leads us to theories about general grammar that we find interesting in their own right, even if they do not directly address the Chomskyan research goal of discovering the innate universal grammar.

References

Abeillé, Anne & Godard, Danièle (eds.). 2021. *La grande grammaire du français (GGF)*. Arles: Actes Sud.

Adger, David & Svenonius, Peter. 2015. Linguistic explanation and domain specialization: A case study in bound variable anaphora. *Frontiers in Psychology* 6. (doi:10.3389/fpsyg.2015.01421)

Andrews, Avery. 1976. The VP complement analysis in Modern Icelandic. *Northeastern Linguistic Society (NELS)* 6. 1–21.

Baker, Mark C. & McCloskey, Jim. 2007. On the relationship of typology to theoretical syntax. *Linguistic Typology* 11(1). 285–296. (doi:10.1515/LINGTY.2007.023)

Baker, Mark C. 2001. *The atoms of language*. New York: Basic Books.

Baker, Mark. 2021. On Chomsky's legacy in the study of linguistic diversity. In Allott, Nicholas & Lohndal, Terje & Rey, George (eds.), *A companion to Chomsky*, 158–171. Hoboken, NJ: Wiley Blackwell. (doi:10.1002/9781119598732.ch10)

Becker, Laura. 2021. *Articles in the world's languages*. Berlin: De Gruyter Mouton.

Belletti, Adriana & Rizzi, Luigi. 2002. Editors'introduction: Some concepts and issues in linguistic theory. In Chomsky, Noam (ed.), *On nature and language*, 1–44. Cambridge: Cambridge University Press.

Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11(1). 239–251.

Boas, Franz. 1911. Introduction. In Boas, Franz (ed.), *Handbook of American Indian languages*, 1–83. Washington, DC: Bureau of American Ethnology.

Bošković, Željko. 2022. Formalism and, not vs, functionalism. *Lingbuzz*. (https://ling.auf.net/lingbuzz/006337)

Carnie, Andrew & Guilfoyle, Eithne (eds.). 2000. *The syntax of the verb initial languages*. New York: Oxford University Press.

Chomsky, Noam A. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam A. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam A. 2001. Derivation by phase. In Kenstowicz, Michael (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic approach*. New York: Oxford University Press.

Comrie, Bernard. 1978. Ergativity. In Lehmann, Winfred P. (ed.), *Syntactic typology: Studies in the phenomenology of language*, 329–394. Austin: University of Texas Press.

Comrie, Bernard. 1984. Language universals and linguistic argumentation: A reply to Coopmans. *Journal of Linguistics* 20(1). 155–163. (doi:10.1017/S0022226700009889)

Coopmans, Peter. 1984. Surface word-order typology and universal grammar. *Language* 60(1). 55–69. (doi:10.2307/414190)

Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, William. 2009. Methods for finding universals in syntax. In Scalise, Sergio & Magni, Elisabetta & Bisetto, Antonietta (eds.), *Universals of language today*, 145–164. Dordrecht: Springer.

Croft, William. 2010. Ten unwarranted assumptions in syntactic argumentation. In Boye, Kasper & Engberg-Pedersen, Elisabeth (eds.), *Language usage and language structure*, 313–350. Berlin: De Gruyter Mouton.

Davis, Henry & Gillon, Carrie & Matthewson, Lisa. 2014. How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language* 90(4). e180–e226.

Deal, Amy Rose. 2015. Ergativity. In Kiss, Tibor & Alexiadou, Artemis (eds.), *Syntax: Theory and analysis (Volume 1)*, 654–708. Berlin: De Gruyter Mouton. (doi:10.1515/9783110377408.654)

Diessel, Holger. 1999. *Demonstratives: Form, function, and grammaticalization*. Amsterdam: Benjamins.

Dryer, Matthew S. & Haspelmath, Martin (eds.). 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (https://wals.info/)

Dryer, Matthew S. 1997. Are grammatical relations universal? In Bybee, Joan L. & Haiman, John & Thompson, Sandra A. (eds.), *Essays on language function and language type: Dedicated to T. Givón*, 115–143. Amsterdam: Benjamins.

Dryer, Matthew S. 2005. Order of subject, object and verb. In Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard (eds.), *The World Atlas of Language Structures*, 330–333. Oxford: Oxford University Press. (https://wals.info/chapter/81)

Eythórsson, Thórhallur & Barðdal, Jóhanna. 2005. Oblique subjects: A Common Germanic inheritance. *Language* 81(4). 824–881.

Gibson, Edward & Futrell, Richard & Piandadosi, Steven T. & Dautriche, Isabelle & Mahowald, Kyle & Bergen, Leon & Levy, Roger. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5). 389–407. (doi:10.1016/j.tics.2019.02.003)

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph H. (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Haspelmath, Martin & Dryer, Matthew & Gil, David & Comrie, Bernard (eds.). 2005. *The world atlas of language structures*. Oxford: Oxford University Press.

Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11(1). 119–132.

Haspelmath, Martin. 2010a. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. (doi:10.1353/lan.2010.0021)

Haspelmath, Martin. 2010b. Framework-free grammatical theory. In Heine, Bernd & Narrog, Heiko (eds.), *The Oxford Handbook of Linguistic Analysis*, 341–365. Oxford: Oxford University Press.

Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15(3). 535–567.

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. 2019. 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. 2019. 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. (doi:10.1075/alal.20032.has)

Haspelmath, Martin. 2021a. General linguistics must be based on universals (or nonconventional aspects of language). *Theoretical Linguistics* 47(1–2). 1–31. (doi:10.1515/tl-2021-2002)

Haspelmath, Martin. 2021b. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. (doi:10.1017/S0022226720000535)

Haspelmath, Martin. 2022. Reassessing the legacy of Aspects. *Inference: International Review of Science* 6(4). (https://inference-review.com/letter/reassessing-the-legacy-of-aspects)

Hawkins, John A. 1985. Complementary methods in universal grammar: A reply to Coopmans. *Language* 61(3). 569–587. (doi:10.2307/414386)

Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75. 244–285.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. New York: Oxford University Press.

Himmelmann, Nikolaus P. 2022. Against trivializing language description (and comparison). *Studies in Language* 46(1). 133–160. (doi:10.1075/sl.19090.him)

Holmberg, Anders. 2017. Linguistic typology. In Roberts, Ian G. (ed.), *The Oxford handbook of Universal Grammar*, 355–376. Oxford: Oxford University Press.

Joseph, John E. 2022. Saussure's universal grammar, Chomsky's structuralism. In Cosenza, Giuseppe & Forel, Claire A. & Puskas, Genoveva (eds.), *Saussure and Chomsky: Converging and diverging*, 19–36. Berne: Peter Lang.

Julien, Marit. 2002. *Syntactic heads and word formation.* New York: Oxford University Press.

Lasnik, Howard & Lidz, Jeffrey L. 2017. The argument from the poverty of the stimulus. In Roberts, Ian G. (ed.), *The Oxford handbook of universal grammar*, 221–248. Oxford: Oxford University Press.

Levinson, Stephen C. & Evans, Nicholas. 2010. Time for a sea-change in linguistics: Response to comments on 'The myth of language universals.' *Lingua* 120(12). 2733–2758.

Levshina, Natalia. 2023. *Communicative efficiency: Language structure and use*. Cambridge: Cambridge University Press.

Pullum, Geoffrey K. & Scholz, Barbara C. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19(1/2). 9–50.
Reinhart, Tanya & Reuland, Eric. 1993. Reflexivity. *Linguistic Inquiry* 24(4). 657–720.
Rizzi, Luigi. 1989. The new comparative syntax: Principles and Parameters of Universal Grammar. *Cahiers Ferdinand de Saussure* 43. 65–78.
Roberts, Ian G. 2017. Introduction. In Roberts, Ian G. (ed.), *The Oxford handbook of Universal Grammar*, 1–34. Oxford: Oxford University Press.
Schmidt, Wilhelm. 1926. *Die Sprachfamilien und Sprachenkreise der Erde*. Heidelberg: Winter.
Skirgård, Hedvig et al. 2023. Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances* 9(16). eadg6175. (doi:10.1126/sciadv.adg6175)
Song, Jae Jung. 2018. *Linguistic typology*. Oxford: Oxford University Press.
Stassen, Leon. 1997. *Intransitive predication*. Oxford: Oxford University Press.
Ting, Connie. 2023. Malagasy N-bonding: A licensing approach. *Glossa* 8(1). (doi:10.16995/glossa.6386)
Truswell, Robert. 2014. Binding theory. In Carnie, Andrew & Sato, Yosuke & Siddiqi, Dan (eds.), *The Routledge handbook of syntax*, 214–238. London: Routledge.
Ultan, Russell. 1978. Toward a typology of substantival possession. In Greenberg, Joseph H. (ed.), *Universals of human language*, vol. 4: Syntax, 11–49. Stanford: Stanford University Press.
Whitney, William D. 1875. *The life and growth of language*. London: Henry S. King & Co.
Woolford, Ellen. 2006. Lexical case, Inherent case, and argument structure. *Linguistic Inquiry* 37(1). 111–130. (doi:10.1162/002438906775321175)
Zaenen, Annie & Maling, Joan & Thrainsson, H. 1985. Case and grammatical functions: The Icelandic passive. *Natural Language and Linguistic Theory* 3(4). 441–483.