Invited by University of Manchester, 2021 December 14

Method and theory in comparative grammar: Measurement uniformity vs. building block uniformity

Martin Haspelmath
Max Planck Institute for Evolutionary Anthropology (Leipzig)

1. Introduction

I will contrast two approaches to worldwide grammatical comparisons:

• an approach based on "measurement uniformity" and a clear separation between an **autonomous comparative method** and explanatory theories
(e.g. Dryer 2006; Haspelmath 2010; 2014)

• an approach based on the idea that languages are made from the same building blocks, so that the explanatory theory (i.e. the hypothesized uniform building blocks) is **not autonomous from the method for comparison**
(e.g. Huang & Roberts 2016; Holmberg 2017)

similar in other sciences:

comparative biology:
comparison is autonomous from explanation
(adaptation and common descent are independent theories)

comparative chemistry:
comparison involves uniform building blocks
(the Periodic Table of Elements)

2. Measurement uniformity as a basis for testing universals

To compare languages, we need **uniform yardsticks for comparison**.

Consider the Greenbergian universal:

> (1) If the adnominal possessor precedes the noun, the object tends to precede the verb; and if the adnominal possessor follows the noun, the object tends to follow the verb.

In order to test this claim, we need to determine the order of **adpossessor** and **noun**, and the order of **object** and **verb** in a representative set of the world's languages.

– how do we measure "order"?
(Dryer 2005: dominant order = more than 67% of occurrences in texts)
– how do we "measure" "possessor", "noun", "object" and "verb"?
(Greenberg 1963: semantically)

NOTE: Comparison is not based on ***the rules*** of the languages
– because the rules do not make reference to text frequencies and to semantic notions

Compare also economics:

> We measure economic indicators (like inflation) by uniform yardsticks, while ignoring culture-specific rules about money and buying (let alone mental representations of money).

Comparison of phonological systems:
by means of **phonetic** properties, not phonological values

In general: comparison requires **comparative concepts**
not **descriptive categories** (Haspelmath 2010)

Measurement uniformity allows large-scale quantitative testing of universals.

(e.g. Dryer & Haspelmath 2013)

3. Building-block uniformity as a basis for testing universals

Many generative linguists say that one needs to have "deeper", non-surface descriptions as the basis for comparison:

> Bobaljik (2015: 318)
> "one of the hurdles to seeing more fruitful interaction between typological studies and formal generative approaches lies in the granularity of the questions being asked, and the degree to which we are ready to **look beyond the surface descriptions**, and to ask questions about patterns at a higher level of abstraction."

> Holmberg (2016: 363)
> "as linguistic theory progresses..., the more confident we can be that the observations are accurate, and **the more abstract the properties** can be that are subject to typological research"

> Roberts (2019: 12):
> "From the perspective of generative grammar, much typological analysis seems excessively **surface-oriented**."

The idea is that the right typological generalizations must be found at the level of abstract analyses (***the rules of the languages***), of the sort typically offered by generativists (in terms of transformations, the X-bar schema, and other abstract elements).

Abstractness is a feature of all science, but de facto, "deep analysis" here primarily means analysis in terms of innate categories (Haspelmath 2019) –
what is of interest is **uniformity of building blocks**, and this is ensured by assuming that the building blocks are innate.

(If the innateness claim is dropped, the approach becomes incoherent; see Haspelmath 2021c).

But we do not know what the building blocks are – this is subject to **constant reevaluation**. Each new language may lead a researcher to make a new proposal about the innate building blocks.

Building-block uniformity does not (yet) allow large-scale quantitative testing of universals.

4. Argument coding splits and a theory that explains them

Four of the most famous argument coding splits are

– differential object marking (DOM)
– dative alternations
– person-split ergativity
– "person case constraint" (PCC)

I claim that they are all special instances of the universal in (1) (Haspelmath 2021a).

(1) **The role-reference association universal** (Universal 1)
Deviations from usual associations of role rank and referential prominence tend to be coded by longer grammatical forms if the coding is asymmetric.

4.1. Differential object marking

(2) Hindi

a. मैंने इस लड़के को देखा
*Maĩ-ne is laṛke **ko** dekhaa.*
I-ERG this boy ACC saw
'I saw this boy.' (special marking of animate object)

b. मैंने यह फिल्म देखी
Maĩ-ne yah philm dekhii.
I-ERG this film saw
'I saw this film.'

(3) **The DOM universal** (Universal 2)
If a language has an asymmetric split in object flagging depending on some prominence scale, then the special flag is used on the prominent object.

4.2. Scales of referential prominence

(4) a. inherent prominence

person scale:	1st/2nd > 3rd person
(nominality scale:	person form (independent or index) > full nominal
animacy scale:	human (> animal) > inanimate

b. discourse prominence

definiteness scale:	definite (> specific indefinite) > indefinite nonspecific
givenness scale:	discourse-given > discourse-new
focus scale:	background > focus

4.3. Dative alternations

In Wolof, a dative preposition *ci* is required on the R when it is indefinite.

(5) Wolof (Atlantic)

a. *Jox naa xalebujigéén ji benn velo.*
give 1SG girl DEF INDF bicycle
'I gave the girl a bicycle.'

b. **Jox naa benn xalebujigéén velo bi.*
give 1SG INDF girl bicycle DEF
'I gave a girl the bicycle.'

c. *Jox naa velo bi* ***ci*** *benn xalebujigéen.*
give 1SG bicycle DEF to INDF girl
'I gave the bicycle to a girl.'

(6) **The dative alternation universal** (Universal 3)
If a language has an asymmetric split in recipient flagging depending on some prominence scale, then the special flag is used on the **non-prominent** recipient.

(7) English

a. *I gave him it.*
b. **I gave the boy it.*
c. *I gave it* ***to*** *the boy.* (special marking of full-nominal recipient)

4.4. Person-split ergativity

(8) **The split-ergaivity universal** (Universal 4)
If a language has an asymmetric split in ergative flagging depending on some prominence scale, then the special flag is used on the **non-prominent** subject.

(9) Warrgamay (Pama-Nyungan)

a. *ngana-Ø gaga-ma*
we-NOM go-FUT
'We will go.' (no flag on S-argument)

b. *ngana-Ø* *ngulmburu-Ø* *ngunda-lma*
we-ERG woman-ACC see-FUT
'We will see the woman.' (no flag on person-form A-argument)

c. *maal-**du*** *ngulmburu-Ø* *ngunda-lma*
man-ERG woman-ACC see-FUT
'The man will see the woman.' (ergative flag on full nominal A-argument)

4.5. Person-role interactions ("PCC", cf. Haspelmath 2004)

(10) French ("*me-lui* constraint")

a. (1>3) *Agnès* *me* *la* *présentera.*
Agnès 1SG.REC 3SG.F.THM present.FUT.3SG
'Agnès will introduce her to me.'

b. (3>1) **Agnès* *me* *lui* *présentera.*
Agnès 1SG.THM 3SG.F.REC present.FUT.3SG
'Agnès will introduce me to her.'

c. *Agnès* *me* *présentera* *à* *elle.*
Agnès 1SG.THM present.FUT.3SG to her
'Agnès will introduce me to her.'

(11) **Ditransitive person-role universal** (Universal 5)
If T is 1st/2nd and R is 3rd (i.e. if T is higher on the person scale than R), a language may require a longer construction (not involving person indexes), while (short) person indexes are always allowed when the R is 1st/2nd and the T is 3rd person.

4.6. Universals 2-5 as special cases of Universal 1

Universal 1 subsumes 2-5:

(1) **The role-reference association universal** (Universal 1)
Deviations from usual associations of role rank and referential prominence tend to be coded by longer grammatical forms if the coding is asymmetric.

Role rank:	A (subject) outranks P (object) R (recipient) outranks T (theme)
	A/R tend to be referentially prominent, P/T tend to be non-prominent.
DOM:	prominent objects get special coding
dative alternations:	non-prominent recipients get special coding
split ergativity:	non-prominent subjects get special coding
"PCC":	special coding when T is prominent and R is non-prominent (on the person scale)

4.7. What explains Universal 1?

Universal 1 can be explained as a special case of a still more general universal:

(12) **The differential coding universal** (Universal 6)
In asymmetric differential coding situations, deviations from the frequent or usual associations between conditions and meanings tend to be coded by longer grammatical forms.

For example, adnominal possessive coding tends to be longer under the condition of alienable nouns, and it is often short or zero when the noun in inalienable (Haspelmath 2017).

(13) Mandarin Chinese

a. 我妈妈
wǒ māmā
'my mom'

b. 我的房子
wǒ ***de*** *fángzi*
'my house'

The fact that we get special coding when in unusual or infrequent can be explained by the **efficiency theory asymmetric coding** (Haspelmath 2021b: unexpected meanings need more coding, and languages tend to adapt to their users' needs).

5. Uniform yardsticks vs. uniform building blocks again

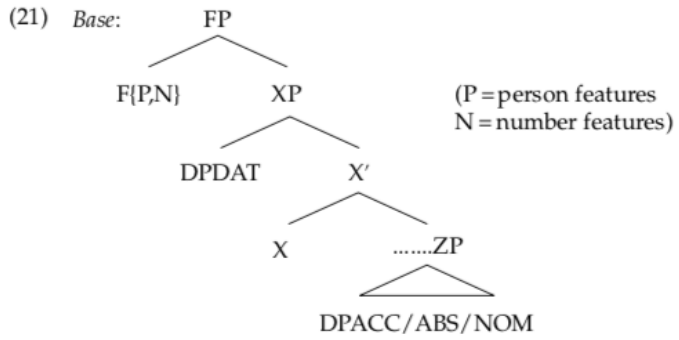
The universals 1-6 are all formulated in terms of **comparative concepts** –
these concepts are defined in the same way for all languages, and they are independent of language-particular rules.

This allows us to test these claims **objectively** – there is no need to first establish the "correct analyses" for particular languages.

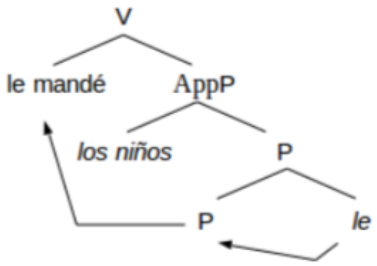
If building-block uniformity is assumed, such objective tests are impossible, because there is a lot of subjectiveness in abstract analyses
(authors "adopt a framework", "argue for their points", "build on assumptions")

e.g. Anagnostopoulou (2017: 11)

> "Anagnostopoulou argues that the PCC and the restriction on nominative objects
> arise in such "two arguments against one head" contexts. Whenever a dative argu- ment enters Move/Agree with a functional head F checking its person features, as in Step I of (21), the lower argument enters Move/Agree with F second and checks the remaining number features, as in Step II of (21):"



Ormazabal & Romero (2019: (44)):



Anagnostopoulou criticizes my earlier proposal
(see blogpost interview: https://dlc.hypotheses.org/2205)

and so do Ormazabal & Romero
(see my blogpost reply: https://dlc.hypotheses.org/2454).

But their proposals fall short of making readily testable claims – what they do (appear to) achieve is to **use the same building blocks** both for description and for explanation.

(Is this parsimonious? But efficiency of coding is a feature of most communication systems and appears to come fro free.)

6. Beyond "functionalism vs. formalism"

For a long time, the contrast between syntacticians who work in the Greenbergian tradition and those working in the Chomskyan tradition has been framed as one between "functionalists" and "formalists"
(e.g. Newmeyer 1998; Thomas 2020)

Bošković (2021):
– functionalist and formalist work is more mutually compatible than is often thought
– generativists are not as "rigidly universalist" as it may appear
(e.g. not all languages have a DP, cf. Bošković (2012) on the NP/DP parameter)
– Greenbergians are not as rigidly "particularist" as it may appear
they keep repeating the "mantra" that "a language must be described in its own terms", but they still compare languages and thus contribute to UG

It is good to see a generativist reaching out and trying to emphasize the commonalities, rather than the differences –

> but Bošković does not talk about building block uniformity, and his "NP/DP parameter" also assumes that "NP" and "DP" are universally available (and thus innate) categories

I keep asking: "Why can't we talk to each other", and my current answer is:

- there is no difference in terms of
  - – ultimate goals
  - – diversity of languages considered
  - – philosophical commitments (cf. Lakoff's 1991 "cognitive commitment")
  - – range of facts considered (social variation, corpus data, etc.)

- the difference is a methdological choice:
  - – do we pursue a "Darwinian vision", separating comparison and explanation?
  - – do we pursue a "Mendeleyevian vision", looking for general building blocks?

Both approaches seem reasonable, but the "Darwinian" approach offers a better hope of explaining argument coding universals.

References

Anagnostopoulou, Elena. 2017. The Person Case Constraint. *The Wiley Blackwell companion to syntax*, 1–47. Wiley Blackwell. (doi:10.1002/9781118358733.wbsyncom101)

Bobaljik, Jonathan. 2015. Distributed Morphology: An interview with Jonathan Bobaljik. *Revista Virtual de Estudos da Linguagem* 13. 314–321.

Bošković, Željko. 2021. Formalism and, not vs, Functionalism. *LingBuzz*. (https://ling.auf.net/lingbuzz/006337)

Dryer, Matthew S. 2005. Determining dominant word order. In Haspelmath, Martin & Dryer, Matthew S & Gil, David & Comrie, Bernard (eds.), *The world atlas of language structures*, 371. Oxford: Oxford University Press. ((http://wals.info/chapter/s6))

Dryer, Matthew S. 2006. Descriptive theories, explanatory theories, and basic linguistic theory. In Ameka, Felix K. & Dench, Alan & Evans, Nicholas (eds.), *Catching language: The standing challenge of grammar writing*, 207–234. Berlin: Mouton de Gruyter. (http://www.acsu.buffalo.edu/~dryer/desc.expl.theories.pdf)

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

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. 2004. Explaining the Ditransitive Person-Role Constraint: A usage-based approach. *Constructions* 2. (doi:10.5281/zenodo.831408)

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

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. 2017. Explaining alienability contrasts in adpossessive constructions: Predictability vs. iconicity. *Zeitschrift für Sprachwissenschaft* 36(2). 193–231. (doi:10.1515/zfs-2017-0009)

Haspelmath, Martin. 2019. Ergativity and depth of analysis. *Rhema* 2019(4). 108–130. (doi:10.31862/2500-2953-2019-4-108-130)

Haspelmath, Martin. 2021a. Role-reference associations and the explanation of argument coding splits. *Linguistics* 59(1). 123–174. (doi:10.1515/ling-2020-0252)

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. 2021c. General linguistics must be based on universals (or nonconventional aspects of language). *Theoretical Linguistics*. (https://ling.auf.net/lingbuzz/005158)

Holmberg, Anders. 2016. Linguistic typology. In Roberts, Ian (ed.), *The Oxford handbook of Universal Grammar*, 355–376. Oxford: Oxford University Press.

Holmberg, Anders. 2017. Universal grammar. In Ledgeway, Adam & Roberts, Ian G. (eds.), *The Cambridge handbook of historical syntax*, 275–300. Cambridge: Cambridge University Press.

Huang, C.-T. James & Roberts, Ian. 2016. Principles and parameters of Universal Grammar. In Roberts, Ian (ed.), *The Oxford handbook of Universal Grammar*. Oxford: Oxford University Press.

Lakoff, George. 1991. Cognitive versus generative linguistics: How commitments influence results. *Language and Communication* 11(1/2). 53–62.

Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, MA: MIT Press.

Ormazabal, Javier & Romero, Juan. 2019. Prolegomena to the study of object relations. *Lingvisticæ Investigationes* 42(1). 102–131. (doi:10.1075/li.00031.orm)

Roberts, Ian. 2019. *Parameter hierarchies and universal grammar*. Oxford: Oxford University Press.

Thomas, Margaret. 2020. *Formalism and functionalism in linguistics: The engineer and the collector*. New York: Routledge.