Some universals of grammar
with particular reference to coding asymmetries (3)

Martin Haspelmath

coding length universals: predictability and coding efficiency

coexpression universals: semantic extension forces

ordering universals: domain minimization
sequence matching
scope matching

1. Greenberg's (1963) coding universals

form-frequency correspondences (singular and NOM/ABS tend to be zero):

> 35. There is no language in which the **plural** does not have some nonzero allomorphs, whereas there are languages in which the **singular** is expressed only by **zero**. The dual and the trial are almost never expressed only by zero.

> 38. Where there is a case system, the only case which ever has only zero allomorphs is the one which includes among its meanings that of the **subject** of the intransitive verb.

more frequently expressed meanings are more likely to be expressed by a short marker:

> 34. No language has a trial number unless it has a dual. No language has a <u>dual</u> unless it has a plural.

> > (cf. "No language has a causative of a transitive unless it has a causative of an intransitive.")

NOT coding length universals:
frequent categories are more likely to have (more) gender distinctions (WHY?):

> 37. A language never has more gender categories in nonsingular numbers than in the **singular**.

> 43. If a language has gender categories in the noun, it has gender categories in the **pronoun**.

> 44. If a language has gender distinctions in the first person, it always has gender distinctions in the second or **third person**, or in both.

> 45. If there are any gender distinctions in the plural of the pronoun, there are some gender distinctions in the **singular** also.

2. Coexpression universals

More examples of semantic maps representing a range of coexpression universals:
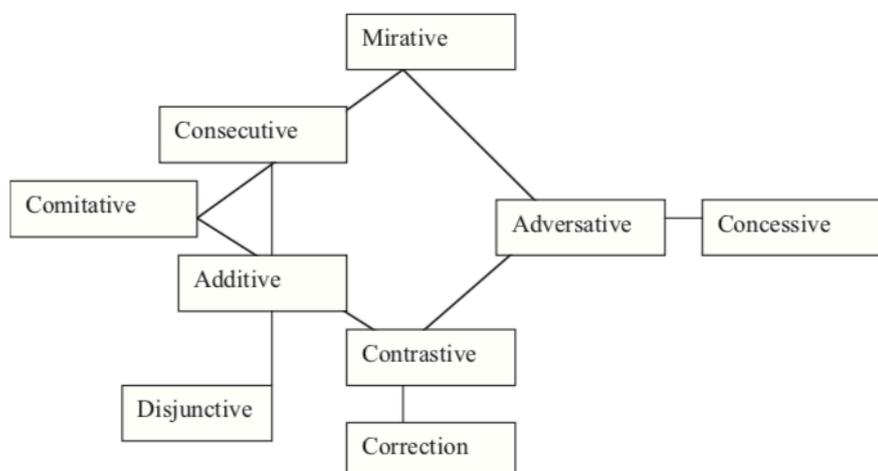
Malchukov (2004):

Figure 1 A semantic map for coordinating connectives.
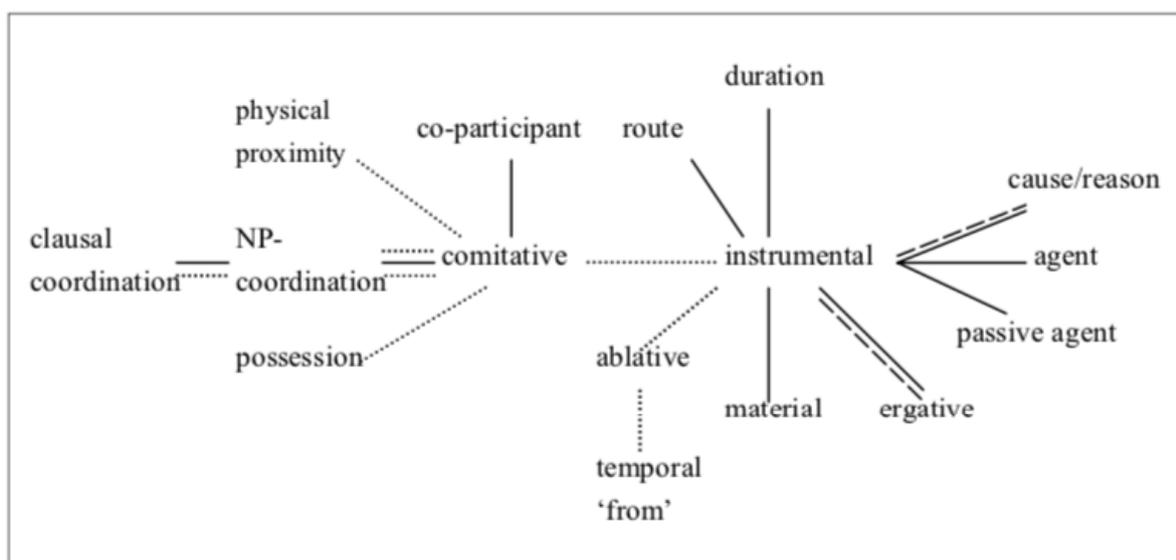
Narrog & Ito (2007):

Figure 8: Our map of the Comitative-Instrumental domain (revised)

CLICS: Colexification database (List et al. 2018, *Linguistic Typology*)

But what explains coexpression patterns? Cognitive representations, such that coexpression always means mental proximity, or at least semantic similarity? William Croft has written about this optimistically:

> The categories defined by constructions in human languages may vary from one language to the next, but they are mapped onto a common conceptual space, which represents a common cognitive heritage, indeed the geography of the human mind (Croft, 2003, p. 139) . . . which can be read in the facts of the world's languages in a way that the most advanced brain scanning techniques cannot ever offer us. (Croft, 2001, p. 364)

But Cristofaro (2010) has pointed out:

Coexpression patterns reflect diachronic changes, and these changes can be due to a variety of processes, often metonymic rather than similarity-based (Croft 2010 replies)

Semantic maps have long been interpreted diachronically, e.g. by Narrog & Ito (2007):
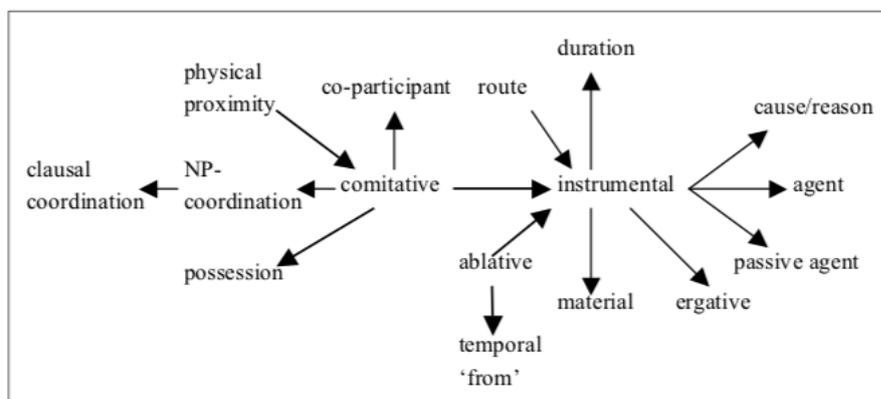

Figure 9: Dynamicized map of the Comitative-Instrumental domain
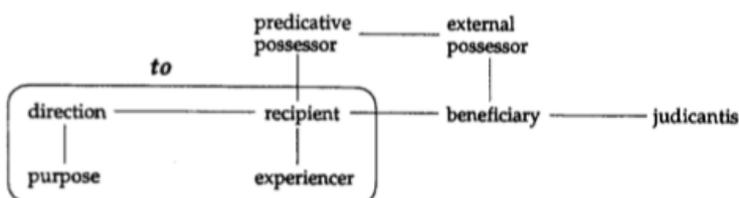
Haspelmath (2003)

FIG. 8.1. A semantic map of typical dative functions/the boundaries of English *to*.
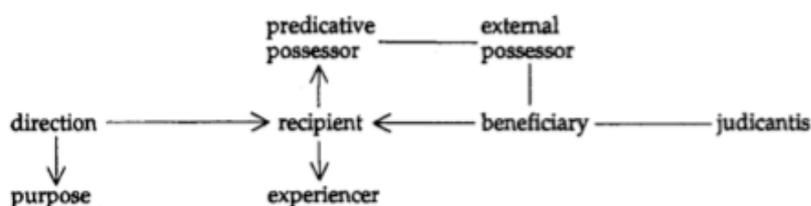

FIG. 8.16. A semantic map of typical dative functions, with directionality.

It may well be that all these patterns are exclusively explained by regularities of semantic change, i.e. that there is no "pull force" that favours coexpression patterns of certain types over patterns of other types.

A warning:
Do not draw synchronic conclusions too rashly from coexpression patterns.

Baunaz & Lander (2018) (adopting a "nanosyntax" approach) consider coexpression patterns among demonstratives, complementizers, relativizers, interrogatives, and indefinites:

Table 1: Syncretism patterns crosslinguistically (neuter/inanimate singular forms).

		DEM_PRO	COMP_FACT	REL_RESTR	WH_PRO	INDET_THING
NGmc	Swedish	det^Pro	att	som^Rvz	vad^Pro	-ting
	Danish	det^Pro	at	som^Rvz	hvad^Pro	-ting
	Icelandic	það^Pro	að	sem^Rvz	hvað^Pro	-hvað-
WGmc	English	that^Pro	that	that^Rvz	what^Pro6	-thing
				% as^Rvz		
	Dutch	dat^Pro	dat	dat^Rvz	wat^Pro	iets
						wat
	German	das^Pro	dass	das^Rel	was^Pro	-was
	Sw. German	das^Pro	dass	wo^Rvz	was^Pro	-is
	Yiddish	jenc^Pro	vos^Fact	vos^Rvz	vos^Pro	-vos
			az	az^Rvz		
Rom	French	ce^Pro	que	que^Rvz	que^Pro	-que
	Italian	quello^Pro	che	che^Rvz	che^Pro	-che
	Spanish	aquél^Pro	que	que^Rvz	qué^Pro	N/A?
	Romanian	acel^Pro	că	ce^Rvz	ce^Pro	ce-
Finno-Ugric	Hungarian	az^Pro	hogy	ami^Rel	mi^Pro	-mi
	Finnish	tä-^Pro 'this'	että	mi-^Rel	mi-^Pro	mi-
Hellenic	Modern Greek	ekíno^Pro	pu^Fact	pu^Rvz	tí^Pro	(-)ti(-)
ESlav	Russian	to^Pro	čto	čto^Rvz	čto^Pro	(-)čto(-)
WSlav	Polish	to^Pro	że	co^Rvz	co^Pro	co-
				% że^Rvz		
	Czech	to^Pro	že	co^Rvz	co^Pro	-co
SSlav	Serbo-Croatian	to^Pro	što^Fact	što^Rvz	što^Pro	-šta/-što
	Bulgarian	tova^Pro 'this'	deto^Fact	deto^Rvz	kakvo^Pro	-shto
	Macedonian	toa^Pro 'this/that'	što^Fact	što^Rvz	što^Pro	-što
			deka	deka^Rvz		

The authors say that this pattern provides evidence for the "functional sequence" in (2):

(2) Dem > Comp > Rel > Wh > Indet

These are meant to be functional heads of the mainstream generative ("cartographic") type, with the crucial extra idea of nanosyntax being that an exponent can "lexicalize" (= express) a syntactic structure if its features are a superset of the structure's features. Thus demonstratives are said to have the features [Dem [Comp [Rel [Wh [Indet]]]]], complementizers are said to have the features [Comp [Rel [Wh [Indet]]]], and so on.

Is there independent evidence for these features? Is there is a deeper difference between nanosyntax and the semantic-map view (which would simply say that (2) is a semantic map (an implicational sequence) and that each exponent must express a contiguous region on the map)?

Fairly clear diachronic tendencies (cf. https://dlc.hypotheses.org/1029):

• indefinite pronouns come from interrogative pronouns (but not vice versa; Haspelmath 1997)

• relativizers in European languages come from interrogatives (and never vice versa; Lehmann 1984).

• For the relation between complementizers and relativizers, I know of no specialized research, but it seems to me that there is a unidirectional tendency as well: Complementizers come from relativizers, not the other way round.

• relativizers occasionally derive from demonstratives (as in Germanic), and never the other way round (Diessel 1999).

What explains these diachronic directionalities is a good question, but there cannot be a biocognitive explanation, as seems to be presupposed by nanosyntax.

3. Biocognitive explanations of universals

(see also https://dlc.hypotheses.org/1012)

Ken Hale (1978): (from an announcement of a course)

```
    This year has seen a number of important publications in language typo-
logy.  Greenberg's Stanford Project on language universals has published its
findings (Greenberg, Joseph H. (ed.) Universals of Human Language, four volumes,
Stanford University Press, 1978) and a book of essays on typology in syntax
has appeared under the editorship of Lehmann at Texas (Lehman, Winfred P.
(ed.) Syntactic Typology, University of Texas Press, 1978).

    Much of this work employs a methodology which gives typology a central
position in linguistic theory:

              data ---> typology ---> theory

That is, typology consists in generalizations deriving from observations of
primary linguistic data, and linguistic theory consists of a model of language
design deriving from a typology of the world's languages.

    In this course, I will review some of this recent work, much of which is
excellent and insightful, and I will argue for a somewhat different approach
to typology:

              data <--> theory ---> typology

Here, typology does not exist as an autonomous entity but, rather, is the product
of linguistic theory.
```

https://twitter.com/themitcho/status/1010423203326582784

Baker (2001), and much other work in the Chomskyan tradition:
  grammatical universals are due to to innate grammatical knowledge ("UG")

  e.g. Smith et al. (2019):
  "the unattested patterns do not arise as they cannot be generated in a manner consistent with Universal Grammar"

But this research programme has largely been given up, as most of the proposals from the 1980s and 1990s have turned out not to be correct (Haspelmath 2008; Baker 2008; Boeckx 2014).

It could of course be that **possible building blocks of grammar are restricted** to a few dozen or a few hundred – just as in chemistry, the number of elements are restricted to about 100, and as in the psychology of emotions, the number of emotions that humans may have are perhaps restructed to six (anger, sadness, fear, disgust, happiness, surprise; argued by psychologist Paul Ekman, and popularized in the Pixar movie "Inside out").

It could also be that there are tens of thousands of natural building blocks of grammar, just as there are tens of thousands of plant species and animal species (as listed in the Encyclopedia of Life). And similarly, genomes may consist of tens of thousands of genes, which may also be thought of as natural kinds.

In other words, it could be that nature provides a set of building blocks, or **natural kinds**, out of which more complex structures are built. Natural kinds are universally presupposed in chemistry (and physics), and widely assumed for biological species.

Linguistics has no clear research programme for finding the natural kinds of grammar

Universal features (as natural kinds) are presupposed

In linguistics, it is widely presupposed that categories and features are natural kinds, i.e. aspects of the innate language faculty. For phonological features, this was explicit in Chomsky & Halle (1968) (*The sound pattern of English*). For syntax, it was explicit even earlier:

> "We require that the grammar of a given language be constituted in accord with a specific theory of linguistic structure in which such terms as "phoneme" and "phrase" are defined independently of any particular language." (Chomsky 1957: 50)

A well-known example from Chomsky (1970):

[±N], [±V]:	noun:	[+N, −V]
	verb:	[−N, +V]
	adjective:	[+N, +V]
	adposition:	[−N, −V]

In other words, universal grammar is thought to provide a "toolbox" of categories that languages may use (Jackendoff 2002).

But while many linguists who work on grammar (probably the majority) share this overall view, it is not so clear what the true natural kinds of language structure are (the „categories of universal grammar", as they are often called, or aprioristic categories, as I have also called them).

For phonology, we have some specific proposals which have found their way into the textbooks (there is even a list on Wikipedia, in the article distinctive features; but all serious phonologists will acknowledged that this list is very controversial, and Mielke (2008) argued against a universal set of distinctive features on the basis of large-scale cross-linguistic evidence).

For morphosyntax, there are no comprehensive proposals, and there are few if any small-scale proposals that have been generally accepted. For example, for the feature „person", there are a variety of feature-value

systems that have been proposed: ±I, ±II, ±III, or ±ego, ±tu, or ±author ±hearer ±participant (cf. Harbour 2017).

For syntactic parameters, Baker (2002) made some far-reaching proposals, but he never published any technical work to justify them, and he does not refer back to his earlier work. Thus, even though many linguists presuppose natural kinds of morphosyntax, they rarely make specific proposals (and textbooks such as Koeneman & Zeijlstra (2017) spend a lot of energy on focus specific language-particular analyses, but say little about the cross-linguistic justification of the categories that are assumed).

But how would we judge whether a proposal is successful?

Linguistics has no good criteria for success

To my mind, a more serious problem than the lack of comprehensive proposals is that linguistics has no clear criteria for assessing whether a feature or category should be assumed to be a natural kind (= part of the innate language faculty).

The typical linguistics paper considers a narrow range of phenomena from a small number of languages (often just a single language) and provides an elegant account of the phenomena, making use of some previously proposed general mechanisms and features.

The typical strucure of generative papers

(i) the data
      = description in generally comprehensible terms

(ii) analysis
      = description making use of "theoretical machinery"
      (highly technical, can be understood only if one is a member
      of the generative in-group)

It could be that this method will eventually lead to **convergent results**, and many linguists apparently have this hope, but I do not see much evidence for this over the last 50 years (convergence seems to come primarily from the impact of fashions and some influential individual scholars).

But: Isn't there convergence in some areas, e.g. the issue of configurationality and lexical categories?

> 1980s: some languages are **nonconfigurational** and lack a VP
> (e.g. Warlpiri, Hungarian, German)
>
> 2000s: **all languages** have a VP

> 1800s through 1990s: some languages lack a **distinction between verbs and nouns**, or at least **between verbs and adjectives**
>
> 2000s: **all languages** have a verb-noun distinction, and a verb-adjective distinction

Is this shift indicative of true convergence of results? I don't think it is, because "having a category" is not a claim that can be falsified (http://dlc.hypotheses.org/879).

Linguists often apply different criteria for different languages

e.g. Does Chamorro "have" a noun-verb-adjective distinction? (Chung 2012)

According to Topping's (1973) pre-generative grammar, Chamorro has two word-classes: Class I (transitive verbs / 'see'-type roots), and Class II (intransitive verbs, nouns, adjectives / 'go'-type, 'person'-type and 'big'-type roots). Class I is defined as combining with preposed subject person forms (cf. preposed *hu* in 1a), while Class II is defined as combining with postposed subject person forms (cf. postposed *yu'* in 1b).

(1) a.　**Hu** *li'i'　i　dǎngkulu　na　tǎotao.*　　　(Class I)
　　　　1SG　see　the　big　　LK　person
　　　　'I saw the big person.' (Chung 2012: 11)

　　b.　*H<um>åhanao* **yu'** *gi　chalan.*　　　(Class II, action-root)
　　　　<AGR>go.PROG　1SG　LOC　road
　　　　'I was going on the road.' (Chung 2012: 11)

Not only action-roots, but also thing-roots and property-roots combine with postposed subject person forms in this way, and they do not require a copula, so by this salient criterion, Chamorro roots fall into two broad classes.

Now Chung (2012) claims that on closer inspection, Chamorro has nouns, verbs and adjectives after all – because if we consider further phenomena, more distinctions emerge. She discusses the six criteria in Table 1.

Table 1. Six features of different kinds of roots in Chamorro

	features	'see'-type roots	'go'-type roots	'big'-type roots	'person'-type roots
1	passive	+	–	–	–
2	postposed subject person form	–	+	+	+
3	incorporatable	–	–	–	+
4	prefixable with *mi-*	–	–	–	+
5	subject-predicate agreement	+	+	+	–
6	specific external argument required	+	+	–	–

Clearly, if all these criteria have the same weight, then five different ways of setting up major classes are possible (see Haspelmath 2012b), and it is not immediately clear which of the major-class divisions, if any, is better than others. But Chung does not even ask which is the most elegant way of describing the language – she merely asks whether there is SOME (however flimsy) evidence for grouping Chamorro words into verbs, nouns and adjectives (as suggested by Baker (2003)). Chung thus asks primarily **whether the noun-verb-adjective classification could be made to work for Chamorro**, following the general thinking of the Uniformity Principle:

　　(6) 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.
　　　(Chomsky 2001: 2)

But of course, Chamorro and English ARE different in their grammar, and these differences must be expressed somehow – saying that Chamorro has nouns/verbs/adjectives just pushes the non-uniformity elsewhere. And more importantly, the Uniformity Principle would also be satisfied if we said that **all languages are like Chamorro** (in having Class I and Class II words), because all languages have SOME difference between transitive verbs and all other words (e.g. that only transitive verbs take objects).

I conclude that there is currently no method for **convergence** of findings concerning the natural kinds of language structure – and there are very few researchers who are making an attempt to pull various findings from around the world together (I think the same could be shown for the configurationality debate – if any diagnostic can be used to argue for a VP, one would not be able to prove that there is no VP).

Since in this approach, languages can only be classified typologically after in-depth study, large-scale comparison is very difficult, and very few natural-kind linguists make claims of world-wide scope. Broadly cross-linguistic works such as Harbour (2016) and Baker (2015) are quite exceptional (for Baker 2015, see Haspelmath 2018).

Thus, linguistics currently has no promising research programme that would give us reason for optimism that we can find the natural kinds of language structure.

Grammatical structures seem to be similar to sound inventories (cf. the 2160 segment types of Phoible.org) and lexical structures in that they allow an open-ended and highly variable range of features and categories, which generally need to be described in language-particular terms.

The biological capacity for language

Nevertheless, it seems clear that language is a biological attribute of humans, like sexuality, sociality, and musicality.

Each language is culturally unique, but the "instinct" to learn and use language seems to be part of human nature, as it is completely uniform across all human groups.

The terms "language faculty" has been used in a more specific (Chomskyan) sense (often identically with "universal grammar"), but the term "linguisticality" could be used in a more general sense, along the lines of sociality and musicality (Fitch 2017; Haspelmath 2020).

The evolution of human linguisticality is an interestig question, but given that we have very little evidence for innate building blocks of grammar, linguists may not be able to contribute very much to its study.

References

Baker, Mark C. 2001. *The atoms of language*. New York: Basic Books.
Baker, Mark C. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge: Cambridge University Press.
Baker, Mark C. 2008. The macroparameter in a microparametric world. In Theresa Biberauer (ed.), *The limits of syntactic variation*. Amsterdam: Benjamins.
Baker, Mark C. 2015. *Case*. Cambridge: Cambridge University Press.

Baunaz, Lena & Eric Lander. 2018. Deconstructing categories syncretic with the nominal complementizer. *Glossa: a journal of general linguistics* 3(1). doi:10.5334/gjgl.349. http://www.glossa-journal.org//articles/10.5334/gjgl.349/.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

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

Chung, Sandra. 2012. Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 38(1–2). 1–56. doi:10.1515/tl-2012-0001.

Cristofaro, Sonia. 2007. Deconstructing categories: Finiteness in a functional-typological perspective. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and empirical foundations*, 91–114. Oxford: Oxford University Press.

Cristofaro, Sonia. 2010. Semantic maps and mental representation. *Linguistic Discovery* 8(1). 35–52. doi:10.1349/PS1.1537-0852.A.345.

Croft, William. 2003. *Typology and universals*. 2nd edition. Cambridge: Cambridge University Press.

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

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

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

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and crosslinguistic comparison. In Michael Tomasello (ed.), *The New Psychology of Language*, vol. 2, 211–243. New York: Lawrence Erlbaum. 10.5281/zenodo.831410.

Haspelmath, Martin. 2008. Parametric versus functional explanations of syntactic universals. In Theresa Biberauer (ed.), *The limits of syntactic variation*. Amsterdam: Benjamins. (27 May, 2016).

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80.

Haspelmath, Martin. 2012. How to compare major word-classes across the world's languages. In Thomas Graf, Denis Paperno, Anna Szabolcsi & Jos Tellings (eds.), *Theories of everything: in honor of Edward Keenan*, 109–130. (UCLA Working Papers in Linguistics 17). Los Angeles: UCLA.

Lehmann, Christian. 1984. *Der Relativsatz: Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik*. Tübingen: G. Narr. http://www.christianlehmann.eu/publ/Lehmann1984_Der_Relativsatz_OCRed.pdf.

Malchukov, Andrej L. 2004. Towards a semantic typology of adversative and contrast marking. *Journal of Semantics* 21(2). 177–198. doi:10.1093/jos/21.2.177.

Narrog, Heiko & Shinya Ito. 2007. Re-constructing semantic maps: The comitative-instrumental area. *STUF–Sprachtypologie und Universalienforschung* 60(4/2007). 273–292. doi:https://doi.org/10.1524/stuf.2007.60.4.273.

Smith, Peter W., Beata Moskal, Ting Xu, Jungmin Kang & Jonathan David Bobaljik. 2019. Case and number suppletion in pronouns. *Natural Language & Linguistic Theory*. doi:10.1007/s11049-018-9425-0.

Topping, Donald. 1973. *Chamorro reference grammar*. Honolulu: University Press of Hawaii.