

Some universals of grammar with particular reference to coding asymmetries (1)

Martin Haspelmath

1. Some language universals and their proposed explanations

<i>coding length universals:</i>	predictability and coding efficiency
<i>coexpression universals:</i>	semantic extension forces
<i>ordering universals:</i>	domain minimization sequence matching scope matching

examples of
coding length universals:

Table 1: Examples of coding asymmetries

singular	plural	(<i>book – book-s</i>)
nominative (A/S)	accusative (P)	(<i>he – hi-m</i>)
allative	ablative	(<i>to – from</i>)
positive	comparative	(<i>small – small-er</i>)
present	future	(<i>go – will go</i>)
affirmative	negative	(<i>go – don't go</i>)
inanimate patient	animate patient	(Spanish \emptyset <i>la casa – a la mujer</i>)
3rd person	2nd person	(Spanish <i>canta</i> _{3sg} / <i>canta-s</i> _{2sg} 'sing(s)')
2nd person imperative	3rd person imperative	(<i>praise! – let her praise!</i>)
attributive adjective	attributive verb	(<i>small – play-ing</i>)

(cf. Haspelmath 2020a)

Each of these examples represents a universal trend, e.g.

In the world's languages, if the coding is asymmetric, the singular marker is generally shorter than the plural marker (and often zero).

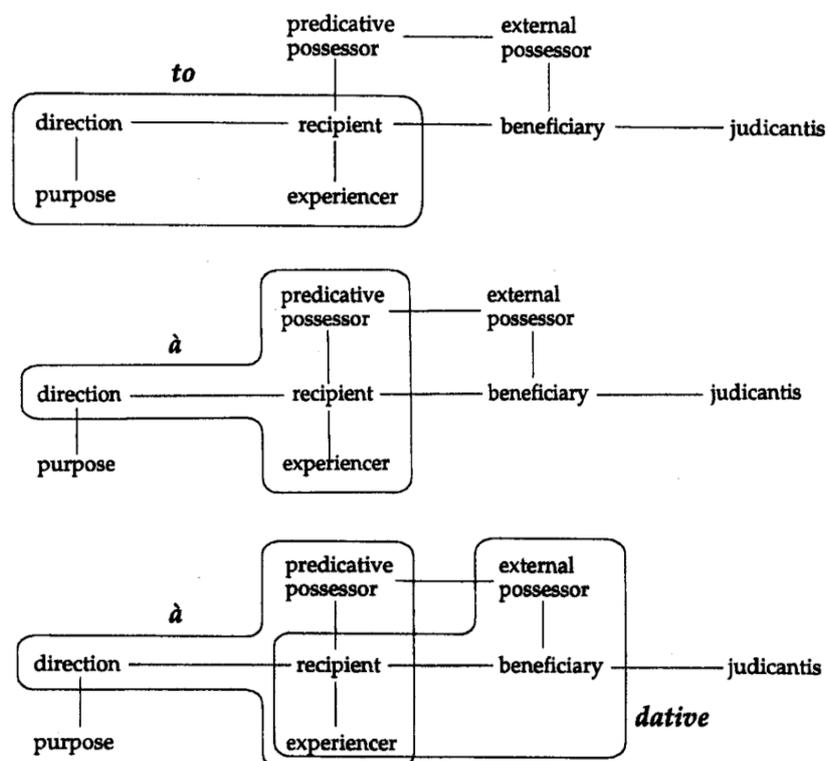
In the world's languages, if the coding is asymmetric, the patient marker is generally shorter for inanimates than for animates (and often zero).

examples of
coexpression universals:

	English <i>to</i>	French <i>à</i>	French dative (<i>lui</i>)
purpose	<i>I'm coming to help you.</i>	<i>Je viens pour t'aider.</i>	* <i>Je lui viens.</i>
direction	<i>I'm going to Leipzig.</i>	<i>Je vais à Leipzig.</i>	* <i>Je lui vais.</i>
pred.possessor	* <i>This book is to me.</i>	<i>Ce livre est à moi.</i>	* <i>Ce livre m'est.</i>
recipient	<i>Give the book to me!</i>	<i>Donne le livre à moi.</i>	<i>Donne lui le livre.</i>
experiencer	<i>This is pleasing to me.</i>	<i>Cela plaît à mon frère.</i>	<i>Cela lui plaît.</i>

external possessor	<i>*My mother is washing the hair to my brother.</i>	<i>*Ma mère lave les cheveux à mon frère.</i>	<i>Ma mère lui lave les cheveux.</i>
beneficiary	<i>*I am buying a book to him.</i>	<i>*J'achète un livre à mon frère.</i>	<i>Je lui achète un livre.</i>
judicantis	<i>*That is too big to me.</i>	<i>*Cela est trop grand à lui.</i>	<i>*Cela lui est trop grand.</i>

(cf. Haspelmath 2003)



An example involving lexical roots (François 2008):

Table 1. Examples of colexification associated with {BREATHE}

	ENGLISH <i>breathe</i>	RUSSIAN <i>du[x]</i>	MWOTLAP <i>mōkheg</i>	NÊLÊMWA <i>horêân</i>	ARAKI <i>soro</i>	FRENCH <i>souffler</i>
BREATHE	+	+	+	+		
take a rest	[+]	[+]	+	+		+
be on vacation		[+]	+			
cease to do				+		
(wind) blow	+	[+]				+
(s.o.) blow	+	[+]	+		+	+
whisper	+				+	+
utter, speak	+			[+]	+	+

The corresponding semantic map:

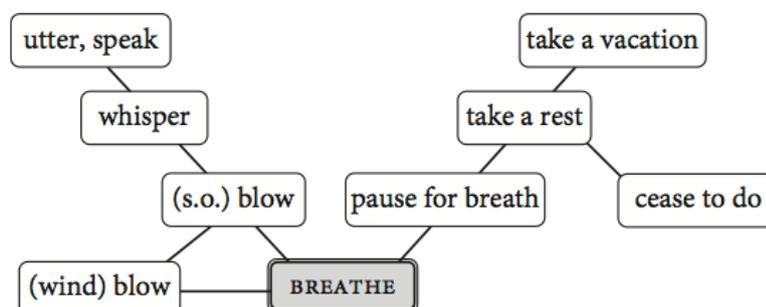


Figure 3. A first semantic map for {BREATHE}.

examples of ordering universals:

Greenberg (1963):

2. In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes.
3. Languages with dominant VSO order are always prepositional.
4. With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional.

Dryer (1992: 108)

VERB PATTERNER
 verb
 verb
 adposition
 copula verb
 'want'
 tense/aspect auxiliary verb
 negative auxiliary
 complementizer
 question particle
 adverbial subordinator
 article
 plural word
 noun
 noun
 adjective
 verb
 verb

OBJECT PATTERNER
 object
 subject
 NP
 predicate
 VP
 VP
 VP
 S
 S
 S
 S
 N'
 N'
 genitive
 relative clause
 standard of comparison
 PP
 manner adverb

most salient:

V-O
 adp-NP
 N-Possr
 comp-S

saw [the girl]
 in [our town]
 tower [of Pisa]
 that [you know]

Greenberg (1963):

14. In conditional statements, the conditional clause precedes the conclusion as the normal order in all languages.

15. In expressions of volition and purpose, a subordinate verbal form always follows the main verb as the normal order except in those languages in which the nominal object always precedes the verb.

Bybee (1985): the order of affixes on verbs tends to be:

V – voice – aspect – tense – mood
mood – tense – aspect – voice – V

e.g. English *They may have been being deceiv-ed.*
mood tense aspect V-voice

Explanations

Coding length universals:

explained by predictability and efficient coding:

efficient coding systems have short coding for predictable information,
and the forms with short coding are predictable because they are
more frequent than the forms with the long coding

Coexpression universals:

explained by “semantic extension forces”:

e.g. English *to* expresses not only direction, but also recipient, because there is
some “force” that makes it extend its meaning in this direction
– these extension forces are often described in terms of
“metaphor” or “metonymy”, and sometimes by “generalization” or
“narrowing”

Ordering universals: three different explanations:

Domain minimization explains the Greenbergian correlations (cf. Hawkins 2014):

<u>saw</u> <u>[the tower</u> <u>[of Pisa]]</u>	domain: 4 words	e.g. English
saw <u>[[of Pisa] the tower]</u>	5 words	
<u>[the tower</u> <u>[Pisa of]]</u> saw	5 words	
<u>[[Pisa of] the tower]</u> saw	4 words	e.g. Turkish

Sequence matching (“iconicity of sequence”) explains why conditional clauses tend to precede, while purpose clauses tend to follow the main clause (other things being equal):

sequence of events:	condition	–	consequence
	action	–	goal of action
sequence of clauses:	conditional	–	matrix clause
	matrix clause	–	purpose clause

Scope matching (Bybee: “relevance”) explains the order of affixes on verbs –
 the scope of aspect includes voice,
 the scope of tense includes aspect,
 the scope of mood includes tense

2. Universals and diversity in an evolutionary perspective

Three types of explanations of universals:

- functional-adaptive: languages are (culturally) adapted to their users’ needs
 e.g. efficient coding, domain minimization
- biocognitive-representational: languages are limited by what can be (biocognitively) represented in speakers’s minds
 e.g. “grammars don’t count”
- mutational: certain changes are impossible or unlikely
 e.g. “direction markers don’t become comitative markers”

Functional-adaptive explanations necessarily have a diachronic/evolutionary component, and mutational explanations are inherently diachronic/evolutionary.

Biocognitive-representational explanations are based on the claim that human biology directly restricts possible mental representations – and human biology is itself subject to biological evolution.

So biological or cultural evolution plays a role in all three types of explanations, and is therefore crucial to understanding language universals, regardless of one’s biases for one or the other explanation type.

Cf. the description of this summer school:

“To the extent that the same structures are recurrently created, propagated and maintained in different languages, an overall pattern will emerge. This is in contrast with the theoretical framework that goes back to Noam Chomsky, where language universals follow from static inbuilt constraints in a speaker’s mind.”

What is perhaps peculiar here is the cultural-evolutionary perspective, not the evolutionary perspective per se. The biological-evolutionary perspective is very much present elsewhere in the field.

cf. Berwick & Chomsky (2015): *Why only us?*
 Ermengildo Bidese: *Evolutionary Linguistics* (new journal)

cf. also Bickel (course description):
 “proposals to explain specific distributional biases in terms of properties of the human brain and the speech/sign apparatus”

Thus, it is very difficult to characterize a particular approach or point of view with a single label. “Evolutionary” is currently fashionable, but none of this is really new.

3. More on universals of asymmetric coding: Simple meaning pairs

3.1. Singular vs. plural vs. dual

	Hebrew	Khanty
SG	<i>yom</i>	<i>xot</i>
PL	<i>yam-im</i>	<i>xot-at</i>
DL	<i>yom-ayim</i>	<i>xot-ɲən</i>
	‘day(s)’	‘house(s)’

3.2. Nominative vs. accusative

	English	German	Quechua
NOM	<i>he</i>	<i>Herr Kim</i>	<i>wasi</i> ‘house’
ACC	<i>hi-m</i>	<i>Herr-n Kim</i>	<i>wasi-ta</i>

In the world’s languages, if the coding is asymmetric, the nominative marker is generally shorter for than the accusative marker (and often zero).

3.3. Allative vs. ablative

	English	Sri Lanka Portuguese	Japanese
ALLATIVE	<i>to Rome</i>	<i>maaket</i> ‘to the market’	<i>Tookyoo e</i> ‘to Tokyo’
ABLATIVE	<i>from Rome</i>	<i>kaaza impa</i> ‘from home’	<i>Tookyoo kara</i> ‘from Tokyo’

In the world’s languages, if the coding is asymmetric, the allative marker is generally shorter for than the ablative marker (and often zero).

3.4. Instrumental vs. comitative

	Welsh	Russian	Hungarian
INSTRUMENTAL	<i>a</i>	<i>myš’-ju</i> ‘with a mouse’	<i>tol-lal</i> ‘with a pen’
COMITATIVE	<i>gyda</i>	<i>s myš’-ju</i>	<i>gyerek-estül</i> ‘with a child’

3.5. Male vs. female occupational terms

	Latin	German	Hungarian
MALE	<i>rex</i>	<i>König</i>	<i>király</i> ‘king’
FEMALE	<i>reg-ina</i>	<i>König-in</i>	<i>király-nő</i> ‘queen’

3.6. Positive vs. comparative vs. superlative

	English	Hungarian	French
POSITIVE	<i>small</i>	<i>kis</i>	<i>petit</i>
COMPARATIVE	<i>small-er</i>	<i>kis-ebb</i>	<i>plus petit</i>
SUPERLATIVE	<i>small-est</i>	<i>leg-kis-ebb</i>	<i>le plus petit</i>

3.7. Cardinal numerals vs. ordinal numerals

	English	Japanese	Lezgian	
CARDINAL	<i>seven</i>	<i>nanatsu</i>	<i>irid</i>	'7'
ORDINAL	<i>seven-th</i>	<i>nanatsu-me</i>	<i>irid lahaj</i>	'7th'

3.8. Present tense vs. future tense

	English	Latin	Kiribati
PRS	<i>they praise</i>	<i>lauda-nt</i>	<i>e taetae</i> 'he speaks'
FUT	<i>they will praise</i>	<i>lauda-b-unt</i>	<i>e na taetae</i> 'he will speak'

3.9. Present tense vs. past tense

	Greek	German	Lezgian	
PRS	<i>ksér-is</i>	<i>weiß-t</i>	<i>či-zwa</i>	'know'
PST	<i>í-kser-es</i>	<i>wuss-te-st</i>	<i>či-zwa-j</i>	'knew'

3.10. Active vs. passive

	Italian	Turkish	Russian	
ACTIVE	<i>vede</i>	<i>gör-üiyor</i>	<i>vidit</i>	'sees'
PASSIVE	<i>è visto</i>	<i>gör-ül-üiyor</i>	<i>vidit-sja</i>	'is seen'

3.11. Basic vs. applicative

	German	Swahili	Dyirbal
BASIC	<i>fahren</i>	<i>andika</i> 'write'	<i>balga-</i> 'beat'
APPLICATIVE	<i>be-fahren</i>	<i>andik-ia</i> 'write to'	<i>balga-ma-</i> 'beat with'

3.12. Affirmative vs. negative

	Hebrew	English	Coptic
AFFIRMATIVE	<i>katavti</i>	<i>I wrote</i>	<i>a-f-sôtm</i> 'he heard'
NEGATIVE	<i>lo katavti</i>	<i>I didn't write</i>	<i>mpe-f-sôtm</i> 'he didn't hear'

3.13. Disjoint anaphoric vs. reflexive

	English	Hebrew	M. Chinese	Japanese
DISJOINT	<i>her</i>	<i>oto</i>	<i>tā</i>	∅
REFLEXIVE	<i>herself</i>	<i>et šacmo</i>	<i>(tā) zijǐ</i>	<i>zibun</i>

In all these cases, the longer forms are less frequent, in all languages. In general, we can say:

(1) **The grammatical form-frequency correspondence hypothesis**

When two grammatical construction types that differ minimally (i.e. that form a semantic opposition) occur with significantly different frequencies, the less frequent construction tends to be overtly coded (or coded with more segments), while the more frequent construction tends to be zero-coded (or coded with fewer segments), if the coding is asymmetric.

The idea of form-frequency correspondences in grammar can be traced back to Greenberg (1966) (see also Croft 2003: Ch. 4; Hawkins 2004: §3.2.2; Haspelmath 2008a; 2008b; Diessel 2019: Ch. 11).

(2) **The form-frequency correspondence universal**

Languages tend to have shorter forms for more frequent meanings.

This is an old insight and is uncontroversial for word length (e.g. Zipf 1935: 23).

4. The big picture

(1) We need to consider universals if we want to **understand Human Language (HL)**.
– many linguists think they can understand HL on the basis of a single language, but this is often illusory

(2) Language universals are not easy to find, because languages are very **diverse and difficult to compare**.

all languages have question pronouns
all languages have negative markers
all languages have demonstratives

(3) Languages are **structurally unique**, and do not conform to the same framework.

(4) Comparison cannot be based on the particular systems of languages, but must make use of **special comparative concepts** (fine-grained variables, or coarse-grained concepts).

(5) Once we have found universals or universal trends, we have new knowledge about Human Language – but we do not have understanding (**explanations**) yet.

(6) Explanations of language universals can come from
– **functional-adaptive** factors: languages are structured in such a way that they can serve their users' needs well
– **biocognitive-representational** factors: languages are structured in such a way that they can be represented by the human mind
– **constraints on possible changes**: some kinds of changes are favoured, while others are disfavoured

5. Incommensurable categories across languages

But categories can also be incommensurable, e.g.

accusative in Japanese, Russian, and ... English?

	Japanese	Russian	English
nominative	<i>Madonna ga</i>	<i>Madonn-a</i>	? <i>Madonna</i>
accusative	<i>Madonna o</i>	<i>Madonn-u</i>	? <i>Madonna</i>

aspect forms in English, Spanish, and Turkish

	English	Spanish	Turkish
habitual	<i>I did</i>	<i>yo hac-ía</i>	<i>yap-ıy-or-du-m/yap-ar-dı-m</i>
progressive	<i>I was doing</i>	<i>yo hac-ía/estaba hac-iendo</i>	<i>yap-ıy-or-du-m</i>
perfective	<i>I did</i>	<i>(yo) hic-e</i>	<i>yap-tı-m</i>

voiced and voiceless stops in English, French, and Ancient Greek

	English	French	Ancient Greek
voiced	<i>better</i>	<i>bibliothèque</i>	<i>beltiōn</i> 'better'
voiceless unaspirated	<i>na[p]kin</i>	<i>nappe, pain</i>	<i>pétron</i> 'rock'
voiceless aspirated	<i>[pʰ]astry</i>	*	<i>pilos</i> 'friend'

Languages can have (apparently) unique categories:

English verbal particles

I looked up the word. *I looked the word up.*
They shrugged off their problems. *They shrugged their problems off.*

German strong and weak adjectives

das dick-e Buch *ein dick-es Buch* 'thick book'
der neu-e Film *ein neu-er Film* 'new movie'

Mandarin Chinese *bǎ* construction

wǒ mài-le wǒ de qìchē
 I sell-PFV I GEN car

wǒ bǎ wǒ de qìchē mài-le
 I ACC I GEN car sell-PFV
 'I sold my car.'

Zuni aspirated velar fricative

[xʰ]

Unique categories make it unlikely that we are dealing with innate categories. But incommensurable categories also seem to make it difficult – we would have to make extra assumptions.

In any event, comparison cannot (always) be naively category-based.

The same applies to **lexical comparison**:

<i>Russian</i>	<i>English</i>	<i>German</i>	<i>French</i>	<i>Danish</i>	<i>Spanish</i>
<i>derevo</i>	<i>tree</i>	<i>Baum</i>	<i>arbre</i>	<i>træ</i>	<i>árbol</i>
	<i>wood</i>	<i>Holz</i>	<i>bois</i>		<i>madera</i>
<i>drova</i>					<i>skov</i>
<i>les</i>	<i>woods</i>	<i>Wald</i>			
	<i>forest</i>		<i>forêt</i>		<i>selva</i>

(Hjelmslev 1943)

6. General solution: Distinguish between language-particular categories and comparative concepts

The proposed solution is to distinguish clearly between **language-particular categories** and concepts designed for comparison: **comparative concepts** (Haspelmath 2010).

Language-particular categories are **defined by language-particular criteria**. They are often defined most easily by **delimiting** them from other categories of the same language.

For example, in phonology, we routinely look for minimal pairs, e.g.

Greek *kronos* 'time' (cf. *chrono-*)
Kronos 'Cronus' (a Titan)

A phoneme can be thought of as a **discrete slice** cut out of the continuous articulatory and acoustic space.

Likewise, a meaning of a word (or of a grammatical marker) can be thought of as a **discrete slice** cut out of the continuous conceptual space.

Cf. Ferdinand de Saussure (1916):

« Abstraction faite de son expression par les mots, notre pensée n'est qu'une masse amorphe et indistincte. [...] Prise en elle-même, la pensée n'est qu'une nébuleuse où rien n'est nécessairement délimité. [...] La substance phonique n'est pas plus fixe ni plus rigide : ce n'est pas un moule dont la pensée doit nécessairement épouser les formes, mais une matière plastique qui se divise à son tour en parties distinctes pour fournir les signifiants dont la pensée a besoin. [...] La langue élabore ses unités en se constituant entre deux masses amorphes. »¹⁶

"If we disregard their expression by words, our thoughts are not more than an amorphous and unclear mass... The phonetic substance is not more fixed or rigid either ..."

Even though typologists have not recognized it clearly until recently, they work with **comparative concepts**, whether in morphosyntax or in typology.

e.g. *WALS*, *PHOIBLE*, *SSWL*, *World Phonotactics Database*, etc.

Comparative concepts need to be defined clearly – one cannot rely on authors of language descriptions to use the same terminology, e.g.

Dahl & Velupillai (2005) (<http://wals.info/chapter/67>)

It is relatively rare for a language to totally lack any grammatical means for marking the future. Most languages have at least one or more weakly grammaticalized devices for doing so. In this chapter, we have therefore decided to map only the inflectionally marked future tenses, inflectional marking being a relatively clear criterion (although there are some borderline cases where it is unclear if one is dealing with a clitic or an affix). ... Many grammars subsume grammatical future-marking devices under the heading “irreal(is)”, especially when their range of use includes negated sentences, counterfactual conditionals, imperatives, etc. With Bybee et al. (1994: 240), we take the view that the distribution of irrealis categories varies too much across languages for them to be acknowledged as a viable cross-linguistic type; such categories are here counted as inflectional futures, if they are expressed inflectionally and cover the same range of uses as other future tenses.

- Greenbergian typologists have always implicitly worked with comparative concepts, separate from descriptive categories, but this distinction has not been explicit.
- Unlike generative grammarians, they have not insisted on “deep analyses” of language-particular phenomena, and have been content with descriptions in grammars, or even with questionnaire data directly from speakers (Dahl 1985).
- But there has been an **implicit widespread idea that descriptive linguists find all the possible categories in typology textbooks** – this is wrong. Descriptive linguists must watch out for surprises.
- Language-particular categories are also known as **emic categories**, while comparative concepts (taking an outsider’s perspective) are known as **etic categories**. (These concepts come from linguistics, but they have become a standard repertoire only in the field of anthropology; many linguists don’t know these terms, as they do not think of languages as part of human cultures.)

Cf. also Bickel’s (2015) “multivariate/distributional typology” programme, which emphasizes the need for fine-grained etic variables:

“the goal is to decompose every single descriptive notion into evermore fine-grained variables ... This goal is best served if the typology is maximally general and contains as few language-particular notions as possible. This is of course a tall order, but it is part and parcel of any attempt to understand the phenomenon of human language in a way that goes beyond cataloguing particulars.”

References

- Berwick, Robert C. & Noam Chomsky. 2015. *Why only us: Language and evolution*. Cambridge MA: MIT Press.
- Bickel, Balthasar. 2015. Distributional Typology. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press. [10.1093/oxfordhb/9780199677078.013.0046](https://doi.org/10.1093/oxfordhb/9780199677078.013.0046).
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.
- Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473(7345). 79–82. doi:10.1038/nature09923.

- François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Martine Vanhove (ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (Studies in Language Companion Series 106), 163–216. 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.
- 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](https://zenodo.org/record/831410).
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. doi:[10.1353/lan.2010.0021](https://doi.org/10.1353/lan.2010.0021).
- Haspelmath, Martin. 2020a. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *to appear*. <https://ling.auf.net/lingbuzz/004531>.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. New York: Oxford University Press.
- Hjelmslev, Louis. 1943. *Omkring sprogteoriens grundlæggelse*. København: Ejnar Munksgaard.
- Jäger, Gerhard. 2018. A Bayesian test of the lineage-specificity of word order correlations, Presented at *The Origins and Evolution of Word Order: A Multidisciplinary Workshop, Part of the 12th International Conference on Language Evolution (Evolang XII)*, Torun, April 2018. (http://www.sfs.uni-tuebingen.de/~gjaeger/talks_complete.shtml)