Bayesian Typology

Gerhard Jäger

Tübingen University

RAILS, Universität des Saarlandes

October 24, 2019

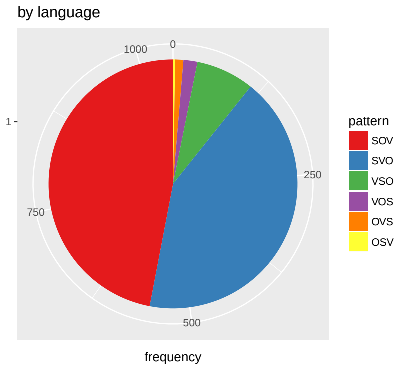
Major word orders

Statistics of major word order distribution

- data: WALS intersected with ASJP
- 1,055 languages, 201 lineages, 71 families with at least 3 languages

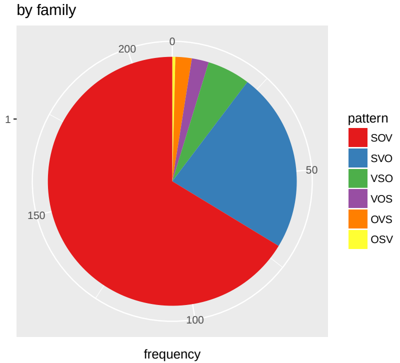
Raw numbers

SOV	SVO	VSO	VOS	OVS	OSV
497	447	78	20	10	3
47.1%	42.4%	7.4%	1.9%	0.9%	0.3%

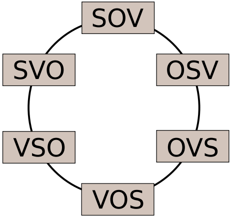


Weighted by lineages

SOV	SVO	VSO	VOS	OVS	OSV
135.1	46.9	10.5	4.0	3.7	0.8
67.2%	23.3%	5.2%	2.0%	1.8%	0.4%



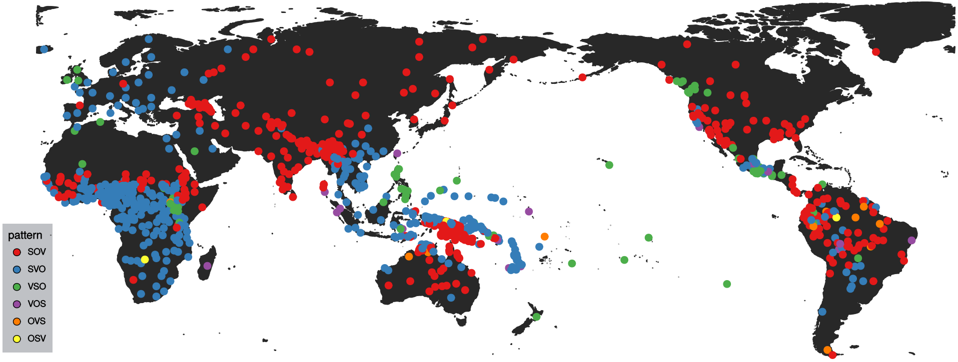
- Gell-Mann and Ruhlen (2011):
  - Proto-world was SOV
  - general pathway: SOV $\rightarrow$ SVO $\leftrightarrow$ VSO/VOS
  - minor pathway: SOV $\rightarrow$ OVS/OSV
  - exceptions due to diffusion
- Ferrer-i-Cancho (2015):



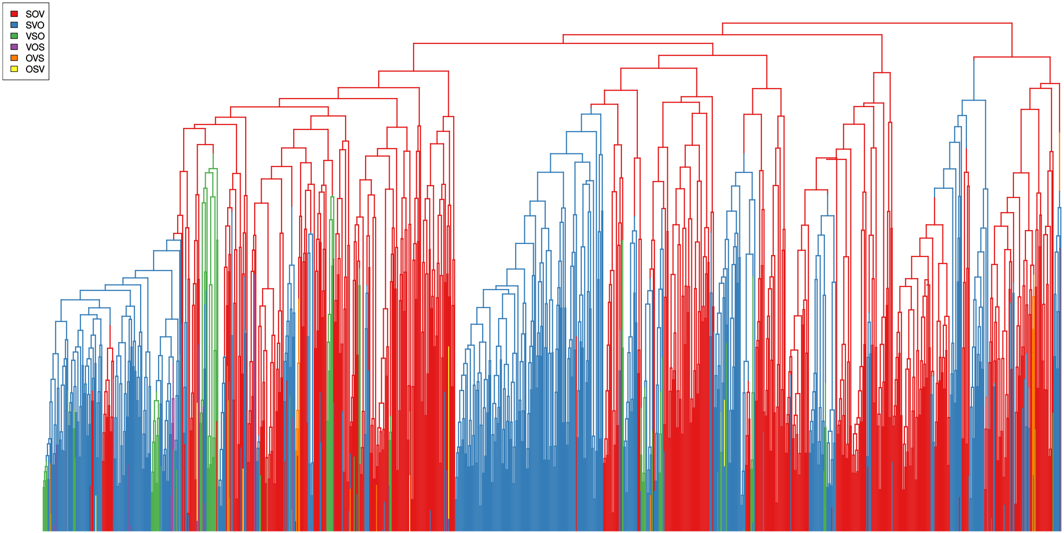
  - permutation circle
  - transition probability inversely related to path length

- languages are phylogenetically structured
- if two closely related languages display the same pattern, these are not two independent data points

⇒ we need to control for phylogenetic dependencies



Phylogenetic non-independence



Typological distributions

- common practice since Greenberg (1963):
  - collect a sample of languages
  - classify them according to some typological feature
  - $\Rightarrow$ skewed distribution indicates something interesting going on

- Problem: languages are not independent samples
- skewed distribution may reflect
  - skewed diversification rate across families
  - properties of an ancestral bottleneck
- balanced sampling mitigates the first, but not the second problem

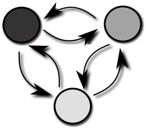
Maslova (2000):

> *"If the A-distribution for a given typology cannot be assumed to be stationary, a distributional universal cannot be discovered on the basis of purely synchronic statistical data."*

> *"In this case, the only way to discover a distributional universal is to **estimate transition probabilities** and as it were to 'predict' the stationary distribution on the basis of the equations in (1)."*

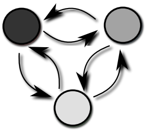
The phylogenetic comparative method

Markov process



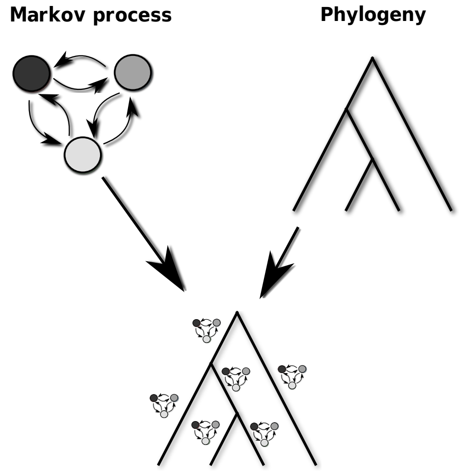
cf. Dunn et al. (2011); Levinson and Gray (2012), *inter alia*

Markov process

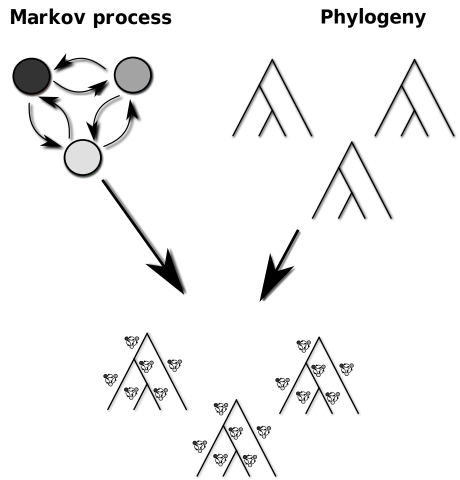


Phylogeny

cf. Dunn et al. (2011); Levinson and Gray (2012), *inter alia*

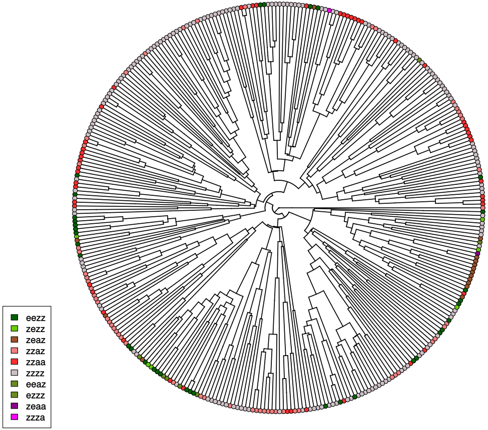


cf. Dunn et al. (2011); Levinson and Gray (2012), *inter alia*

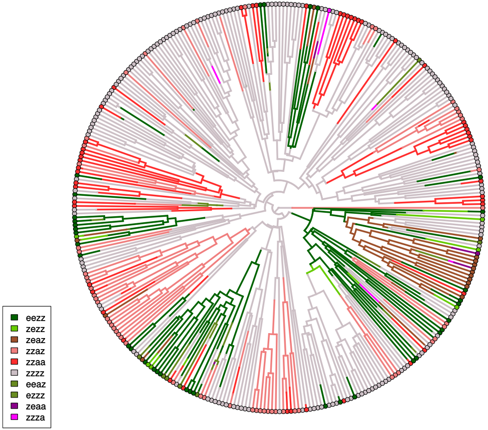


cf. Dunn et al. (2011); Levinson and Gray (2012), *inter alia*

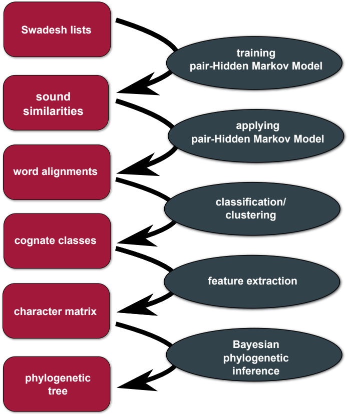
- if phylogeny and states of extant languages are known...

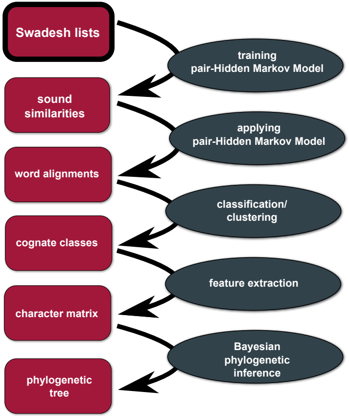


- if phylogeny and states of extant languages are known...
- ... transition rates and ancestral states can be estimated based on Markov model

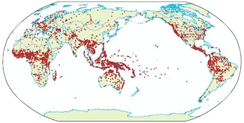


Inferring trees across many families

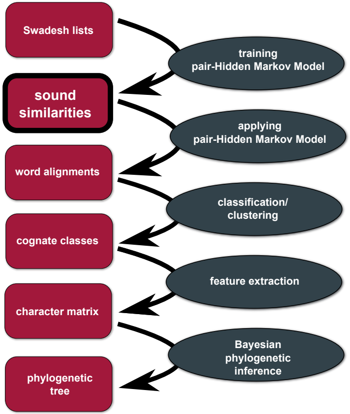

Swadesh lists
training
pair-Hidden Markov Model
sound
similarities
applying
pair-Hidden Markov Model
word alignments
classification/
clustering
cognate classes
feature extraction
character matrix
Bayesian
phylogenetic
inference
phylogenetic
tree

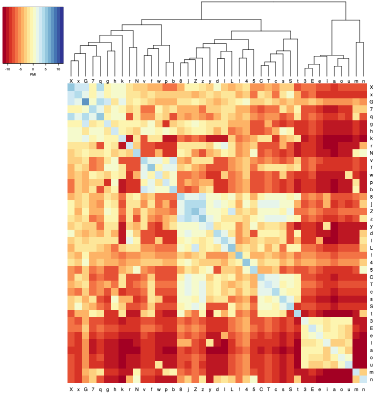


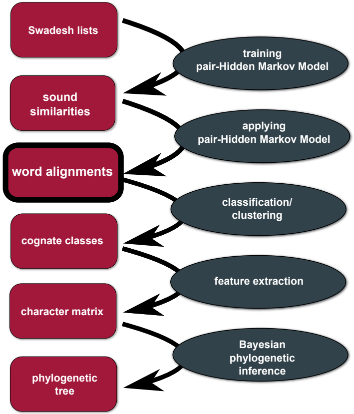
concept	Latin	English
I	ego	Ei
you	tu	yu
we	nos	wi
one	unus	w3n
two	duo	tu
person	persona, homo	pers3n
fish	piskis	fiS
dog	kanis	dag
louse	pedikulus	laus
tree	arbor	tri
leaf	foly∼u*	lif
skin	kutis	skin
blood	saNgw∼is	bl3d
bone	os	bon
horn	kornu	horn
ear	auris	ir
eye	okulus	Ei



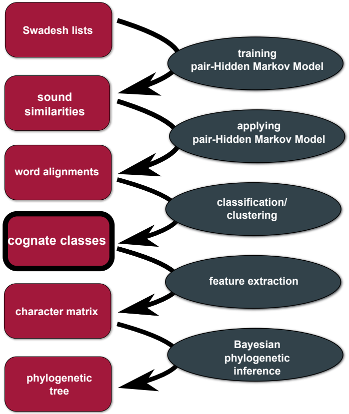
From words to trees

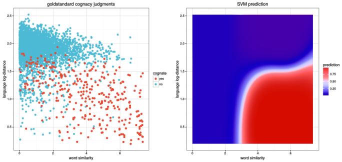




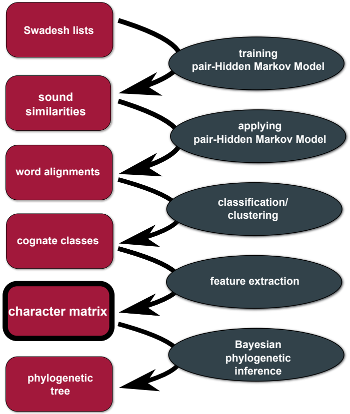


Language	*fish:z*	*tongue:1*	*smoke:1*
Abui-Atangmelang	-af-u		
Abui-Fuimelang	-af-u	tal-i-fi--	
Adang	aab--	tal-E-b---	awai--b-a-n-o-7o-
Blagar-Bakalang	-ab--	--j-e-bur-	--ad--b-a-n-aNka-
Blagar-Bama	aab--	teg-e-bur-	------b-e-n-a-xa-
Blagar-Kulijahi	-ab--	tej-e-bur-	------b-e-n-aNka-
Blagar-Nule	aab--	tej-e-bur-	--ad--b-e-n-aNka-
Blagar-Tuntuli	aab--	tej-e-bur-	a-adgeb-a-n-a-q--
Blagar-Warsalelang	-ab--	tel-e-bur-	a-ad--b-a-n-a-x--
Bunaq			------b-o-t-o-h--
Deing	haf--		------buu-n------
Hamap	7ab--	nar-ø-buN-	------b-a-n-o-7--
Kabola	hab--	tal-e-b---	awal--b-e-n-e-7o-
Kaera-Padangsul	-ab--	talee-b---	a-ad--b-e-naa-x--
Kafoa	-afUi	tal-i-p---	------f-o-n-a----
Kamang	-ap-i	nal---pu--	------p-u-n----a-
Kiraman	-Eb--	nal-i-bar-	--ar--b-a-n-o-kan
Klon	-eb-i	gel-E-b---	--ed-ab-o-n------
Kui	-eb--	tal-i-ber-	--ar--b-o-n-o-k--
Kula	-ap-i	-il-I-p---	------p---n-ekka-
Nedebang	aaf-i	gel-e-fu--	--ar-ab-u-n------
Reta	aab--	nal-e-bul-	a-ad--b-o-n-a----
Sar-Adiabang	haf--	--p-e-fal-	--ar--buu-n------
Sar-Nule	haf--	nal-e-faj-	
Sawila	-ap-i	gal-impuru	------p-u-n-a-ka-
Teiwa-Madar	xaf--	gel-i-vi--	------buu-n------
Wersing	-ap-i	nej-e-bur-	--ad-ap-u-n-a-k--
Wpantar	hap--	nal-e-bu--	------b-unn-a----

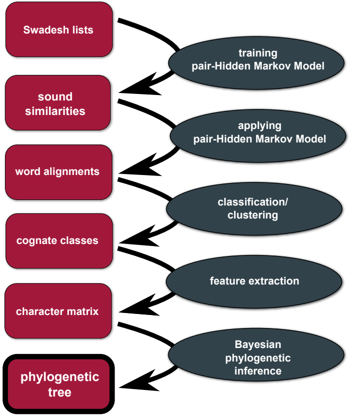


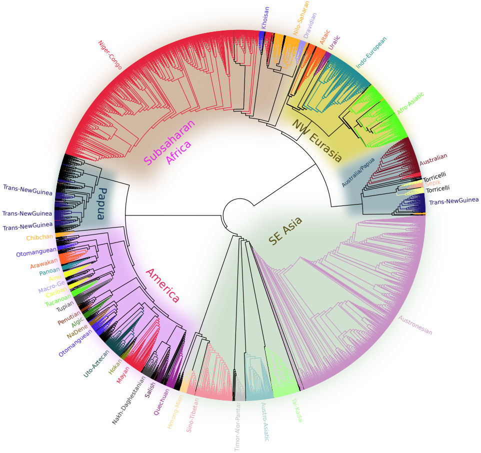


	English	Spanish	Modern Greek	Standard German
I	Ei:A	yo:B	exo:C	iX:D
you	yu:A	ustet:B, tu:C	esi:D	du:E
we	wi:A	nosotros:B	emis:C	vir:A
one	w3n:A	uno:B	enas:C, ena:C	ains:D
two	tu:A	dos:B	8y∼o:C, 8io:D	cvai:E
person	pers3n:A	persona:A	an8∼ropos:B	mEnS:C
fish	fiS:A	peskado:A, pes:A	psari:B	fiS:A
dog	dag:A	pero:B	sTili:C, sTilos:C	hunt:D
come	k3m:A	veni:B	erx∼o:C	kh∼om3n:A
sun	s3n:A	sol:B	ily∼os:C, iLos:C	zon3:A
star	star:A	estreya:A	asteri:A, astro:A	StErn:A
water	wat3r:A	agw∼a:B	nero:C	vas3r:A
stone	ston:A	piedra:B	petra:B	Stain:A
fire	fEir:A	fuego:B	foty∼a:C	foia:D
path	pE8:A	senda:B	8romos:C	pf∼at:A, vek:D
mountain	maunt3n:A	sero:B, monta5a:A	vuno:C, oros:D	bErk:E
full	ful:A	yeno:B	yematos:C, pliris:D	fol:A
new	nu:A	nuevo:A	neos:A, Tenury∼os:B	noi:A
name	nem:A	nombre:A	onoma:A	nam3:A




```
TNG.ENGAN.MAIBI              10000000000000000000000000000000000000000000
TNG.ENGAN.POLE               00000000000000000000000000000000000010000000
TNG.ENGAN.SAU                00000000000000000000000000000000000000000000
TNG.ENGAN.YARIBA             10000000000000000000000000000000000000000000
TNG.FASU.FASU                00000000000000000000000000000000000010000000
TNG.FASU.NAMUMI              00000000000000000000000000000000000000001000
TNG.FINISTERRE-HUON.AWARA    00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.BORONG   00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.BURUM    00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.BURUM_MIND 00000000000000000000000000000000001010000000
TNG.FINISTERRE-HUON.DEDUA    00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.HUBE     00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.KATE     00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.KOMBA    00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.KOSORONG 00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.MAPE     00000000000000000000000000000000000100000000
TNG.FINISTERRE-HUON.MAPE_2   00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.MIGABAC  00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.MINDIK   00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.MOMOLILI 00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.NABAK    00000000000000000000000000000000000100000000
TNG.FINISTERRE-HUON.NANKINA  00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.NEK      00000000000000000000000000000000000001000000
TNG.FINISTERRE-HUON.NUKNA    00000000000000000000000000000000000001000000
TNG.FINISTERRE-HUON.ONO      00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.SELEPET  00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.TIMBE    00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.TOBO     00000000000000000000000000000000000010000000
TNG.FINISTERRE-HUON.WANTOAT  00000000000000000000000000000000000110000000
TNG.FINISTERRE-HUON.YOPNO    00000000000000000000000000000000000001000000
TNG.GOILALAN.AFOA            00000000000000000000000000000000000110000000
TNG.GOILALAN.KUNIMAIPA       00000000000000000000000000000000000010000000
TNG.GOILALAN.MAFULU          00000000000000000000000000000000000010000000
```


Swadesh lists
training
pair-Hidden Markov Model
sound
similarities
applying
pair-Hidden Markov Model
word alignments
classification/
clustering
cognate classes
feature extraction
character matrix
Bayesian
phylogenetic
inference
phylogenetic
tree


Subsaharan
Africa
NW Eurasia
Papua
SE Asia
America
Niger-Congo
Khoisan
Nilo-Saharan
Dravidian
Altaic
Uralic
Indo-European
Afro-Asiatic
Australia/Papua
Australian
Torricelli
Sepik
Torricelli
Trans-NewGuinea
Austronesian
Tai-Kadai
Austro-Asiatic
Timor-Alor-Pantar
Sino-Tibetan
Hmong-Mien
Quechuan
Salish
Nakh-Daghestanian
Mayan
Hokan
Uto-Aztecan
Otomanguean
NaDene
Algic
Penutian
Tupian
Tucanoan
Cariban
Macro-Ge
Ainu
Panoan
Arawakan
Otomanguean
Chibchan
Trans-NewGuinea
Trans-NewGuinea
Trans-NewGuinea

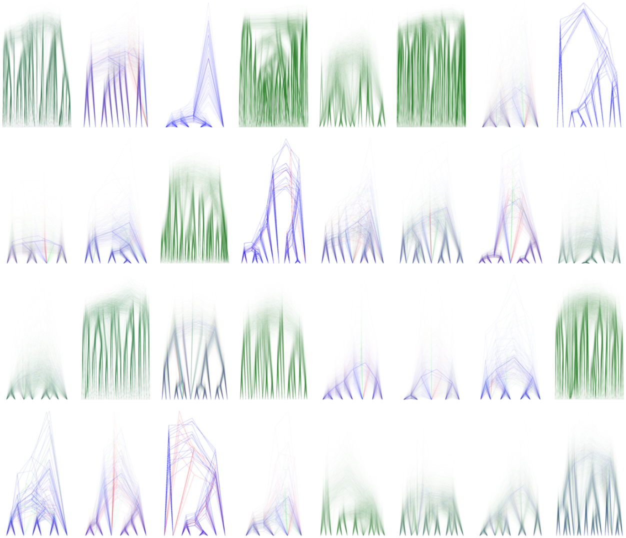
Estimating word-order transition patterns

(data from all 77 families with $\geq$ 3 languages in data base; 924 languages in total)

- estimate posterior tree distributions with MrBayes for each family, using Glottolog as constraint tree
- estimate transition rates
- estimate stationary distribution of major word order categories
- apply *stochastic character mapping* (SIMMAP; Bollback 2006)
- estimate expected number of mutations for each transition type

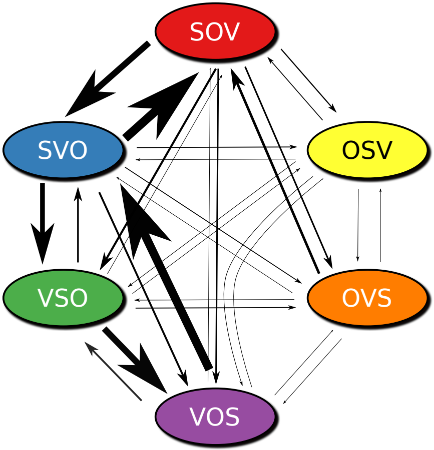
Estimating posterior tree distributions

- using characters extracted from ASJP data (Jäger 2018)
- Glottolog as constraint tree
- Γ-distributed rates
- ascertainment bias correction
- relaxed molecular clock (IGR)
- uniform tree prior
- stop rule: 0.01, samplefreq=1000
- if convergence later than after 1,000,000 steps, sample 1,000 trees from posterior



- totally unrestricted model, all 30 transition rates are estimed independently
- implementation using RevBayes (Höhna et al., 2016)

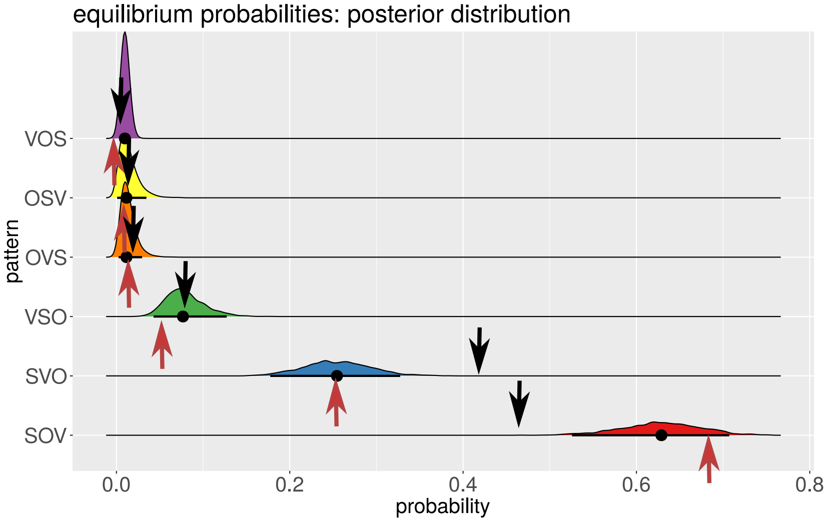
expected strength of flow



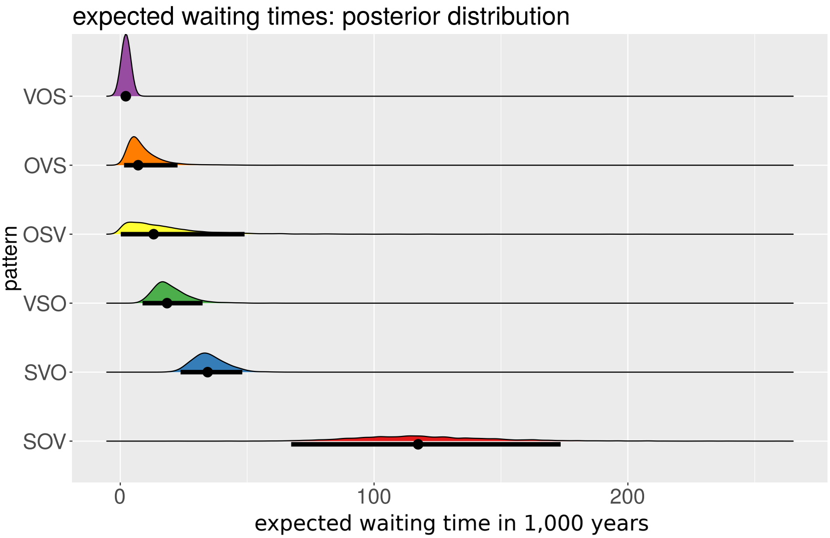
- estimated frequency of mutations within the 77 families under consideration (posterior mean and 95% HPD, 100 simulations

	SOV		SVO		VSO		VOS		OVS		OSV	
SOV	–		51.5	[19; 82]	10.2	[1; 19]	7.5	[0; 29]	5.8	[0; 14]	4.2	[0; 13]
SVO	83.8	[31; 131]	–		22.3	[2; 42]	10.4	[0; 30]	2.8	[0; 8]	3.9	[0; 12]
VSO	1.4	[0; 5]	8.3	[0; 24]	–		29.0	[5; 45]	3.0	[0; 9]	1.1	[0; 5]
VOS	4.3	[0; 15]	141.9	[115; 188]	30.9	[17; 47]	–		2.1	[0; 9]	1.0	[0; 3]
OVS	11.1	[0; 28]	0.8	[0; 4]	1.8	[0; 8]	0.4	[0; 3]	–		0.8	[0; 5]
OSV	4.2	[0; 15]	0.4	[0; 3]	1.9	[0; 11]	1.1	[0; 7]	1.1	[0; 9]	–	

Empirical vs. estimated distribution



Waiting times



Differential case marking

- three universal core roles
  - **S:** intransitive subject
  - **A:** transitive subject
  - **O:** transitive object

German

Der Junge	ist	dreckig.
the boy.NOM	is	dirty

'The boy is dirty.'

Der Junge	wirft	einen Stein.
DEF boy.NOM	throw	a.ACC stone

'The boy is throwing a stone.'

Kalkatungu (Australia)

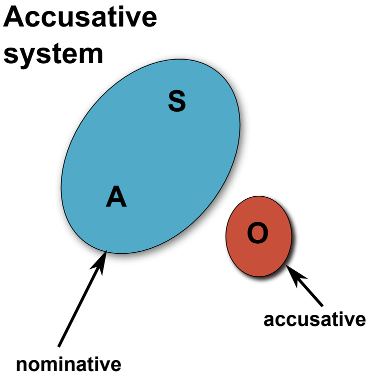
Kaun	muu-yan-ati
dress.ABS	dirt-PROP-INCH

'The dress is dirty.'

Kuntu	wampa-ngku	kaun	muu-yan-puni-mi.
not	girl-ERG	dress.ABS	dirty-PROP-CAUS-FUT

'The girl will not dirty the dress.'

S
A
O

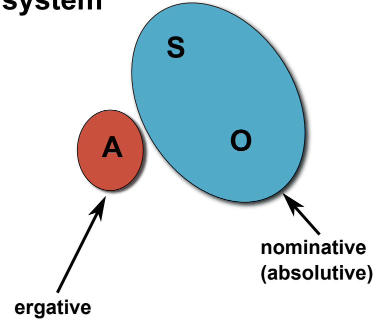


Latin

Puer puellam vidit.
boy.NOM girl.ACC saw *'The boy saw the girl.'*

Puer venit.
boy.NOM came *'The boy came.'*

Ergative system

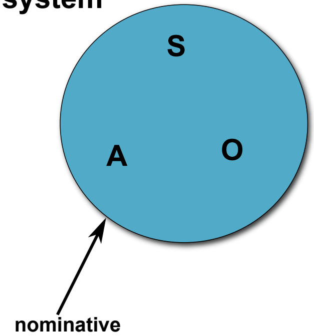


Dyirbal

ŋuma yabu-ŋgu bura-n.
father mother.ERG see-NONFUT
'The mother saw the father.'

ŋuma banaga-nu.
boy.NOM came *'The boy came.'*

Neutral system



Mandarin

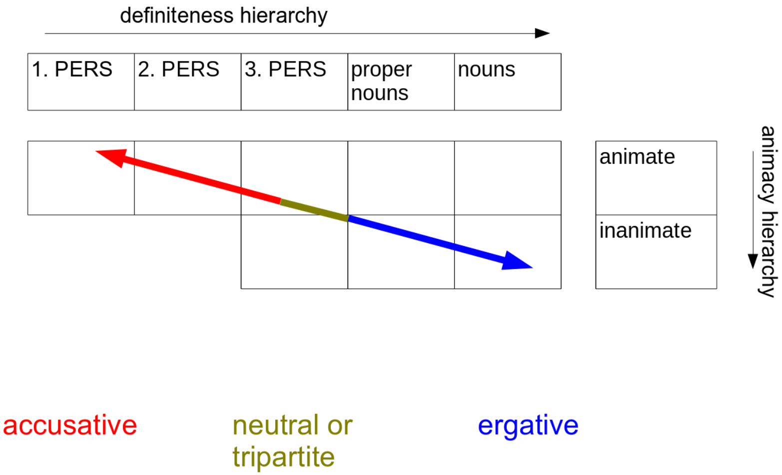
rén lái le.
person come CRS
'The person has come.'

zhāngsān mà lǐsì le ma.
Zhangsan scold Lisi CRS Q
'Did Zhangsan scold Lisi?'

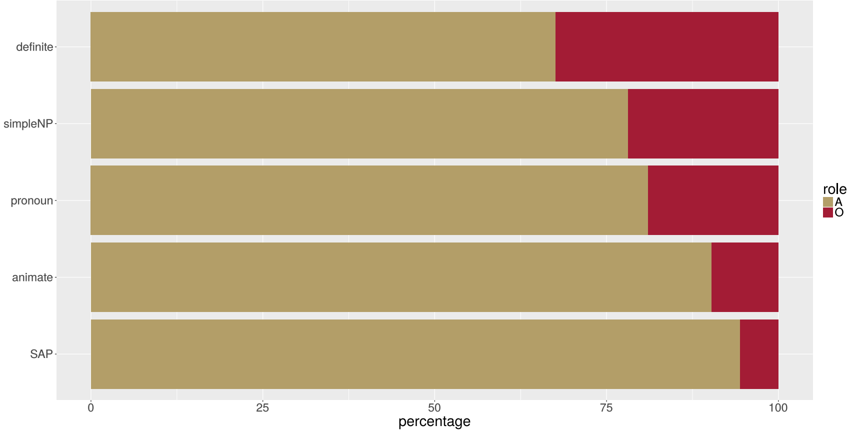
- many languages have mixed systems
- e.g., some NPs have accusative and some have neutral paradigm, such as Hebrew

(1) Ha-seret her?a ?et-ha-milxama
the-movie showed acc-the-war
'The movie showed the war.'

(2) Ha-seret her?a (*?et-)milxama
the-movie showed (*acc-)war
'The movie showed a war'
(from Aissen, 2003)


definiteness hierarchy
1. PERS
2. PERS
3. PERS
proper nouns
nouns
animate
inanimate
animacy hierarchy
accusative
neutral or tripartite
ergative

probability P(syntactic role|prominence of NP)



A is prominent	A is non-prominent	O is prominent	O is non-prominent
e(rgative)	e(rgative)	a(ccusative)	a(ccusative)
e	e	a	z(ero)
e	e	z	a
e	e	z	z
e	z	a	a
...	...	...	...
z	e	z	z
z	z	a	a
z	z	a	z
z	z	z	a
z	z	z	z

actually attested:

1. **zzzz**: no case marking
2. **zzaa**: non-differential object marking
3. **zzaz**: harmonic differential object marking
4. **ezzz**: non-differential subject marking
5. **zeaz**: split ergative
6. **eeaz**: non-differential subject marking plus differential object marking
7. **ezzz**: dis-harmonic differential subject marking
8. **zezz**: harmonic differential subject marking
9. **zeaa**: harmonic differential subject marking plus non-differential object marking
10. **zzza**: dis-harmonic differential object marking

Differential case marking and referential scales

- received wisdom (Silverstein, 1976; Comrie, 1981; Aissen, 2003, , *inter alia*):
  - if object-marking is differential, upper segments of a referential hierarchy receive accusative marking
  - if object-marking is differential, lower segments of a referential hierarchy receive accusative marking
- Bickel et al. (2015):
  - large differences between macro-areas
  - no universal effects of referential scales on differential case marking

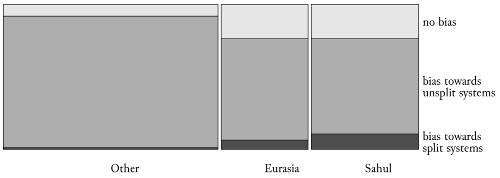


Figure 1: Estimated biases of families having split case markig for A across macro-areas. (The sizes of the individual tiles in the plot are proportional to frequencies, using the 'mosaic' plot technique provided by Meyer et al. 2006)

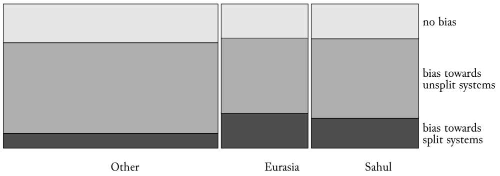
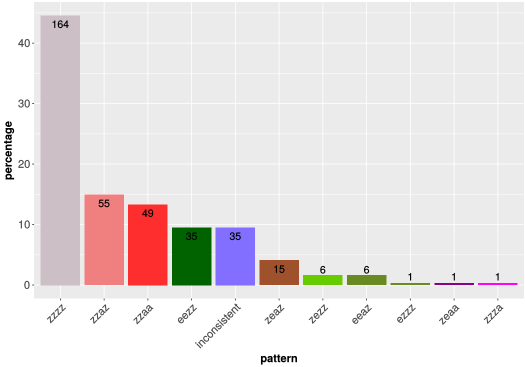


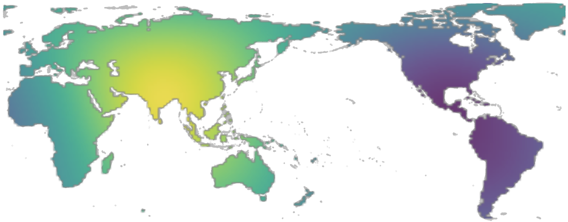
Figure 2: Estimated biases of families having split case markig for P across macro-areas (using the same mosaic plot techniques as in Figure 1)

- genetically diverse sample of 460 case marking systems
- used here: 368 systems
  - one system per language
  - only languages with ISO code
  - only languages present in ASJP
- 2 out of 333 systems (99.4%) are obey the Silverstein hierarchy (not counting inconsistent states)

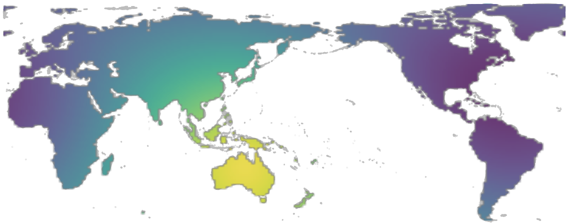


- differential object marking concentrated in Eurasia
- diffential subject marking concentrated in Sahul
- only cases of anti-DOM and anti-DSM (one instance of each) in North America

Differential object marking



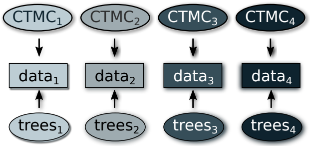
Differential subject marking



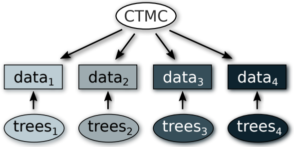
Phylogenetic trees for the case data

- 39 families and 63 isolates in the intersection of the Autotyp data and ASJP (Wichmann et al., 2018)
- for each of these families, I inferred a posterior distribution of 1,000 trees (using lexical data from ASJP) to reflect uncertainty in tree structure and branch length
- Glottolog tree was used as constraint tree

Hierarchical Bayesian models

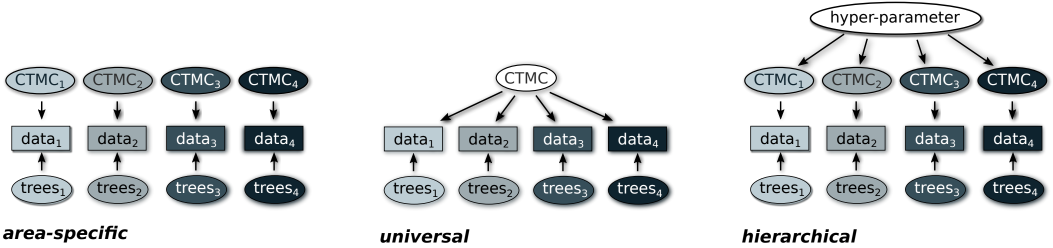


area-specific



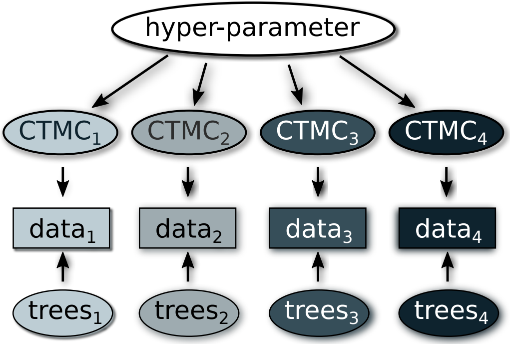
universal

Hierarchical Bayesian models

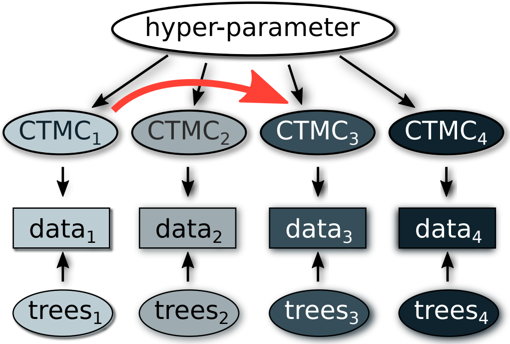


Hierarchical Models to capture areal effects

- each macro-area has its own parameters
- parameters are all drawn from the same distribution f
- shape of f is learned from the data
- prior assumption that there is little cross-area variation $\rightarrow$ can be overwritten by the data

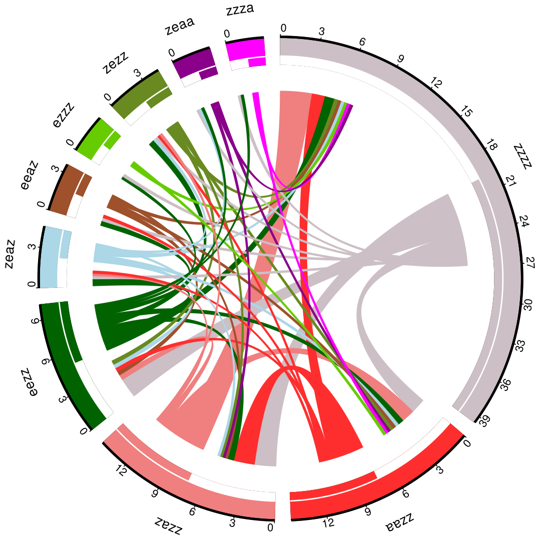


- each macro-area has its own parameters
- parameters are all drawn from the same distribution f
- shape of f is learned from the data
- prior assumption that there is little cross-area variation $\rightarrow$ can be overwritten by the data
- enables **information flow across areas**



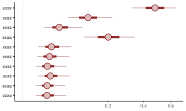
What about isolates?

- Continuous Time Markov Chain defines a unique **equilibrium distribution**
- hierarchical model assumes a different CTMC, and thus a different equilibrium distribution for each lineage
- by modeling assumption, root state of a lineage is drawn from this distribution (Uniformity Principle)
- isolates are treated as families of size 1, i.e., they are drawn from their equilibrium distribution

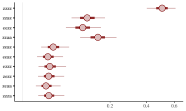

zzzz
zzaa
zeaz
eeaz
eezz
ezzz
zezz
zeaa
zzza
zeaz

Estimated equilibrium distributions

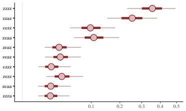
Africa



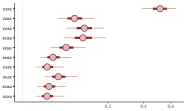
Americas



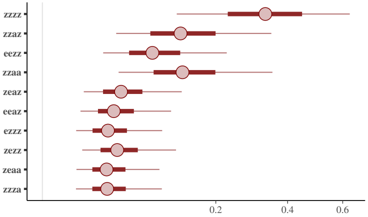
Eurasia



Sahul

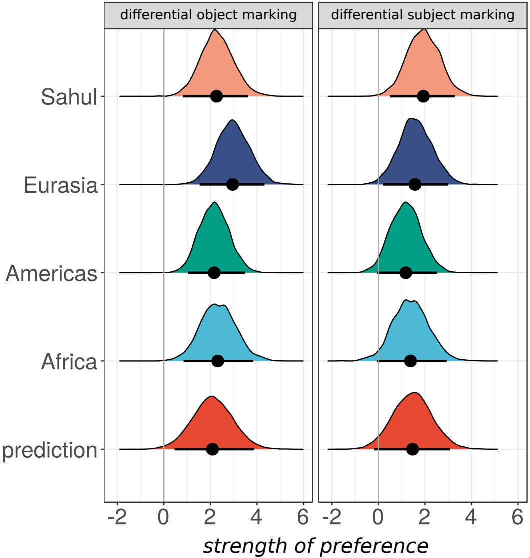


posterior prediction



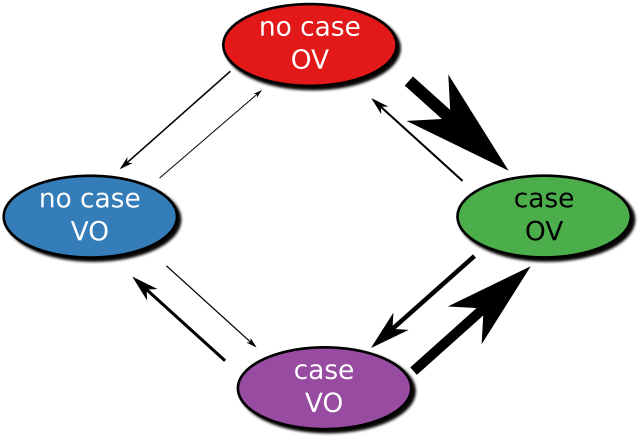
Preference for scale-respecting differential case marking

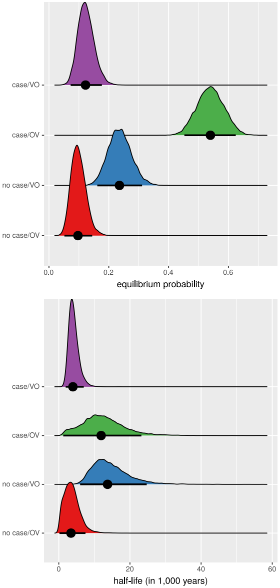
- **strength of preference** of DOM over anti-DOM:
$$\log \frac{P(..\mathrm{az})}{P(..\mathrm{za})}$$
- DSM over anti-DSM:
$$\log \frac{P(\mathrm{ze}..)}{P(\mathrm{ez}..)}$$



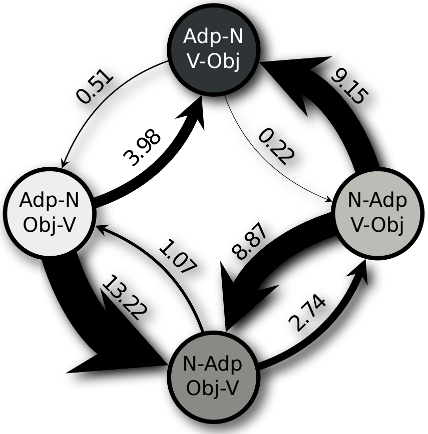
Further variables

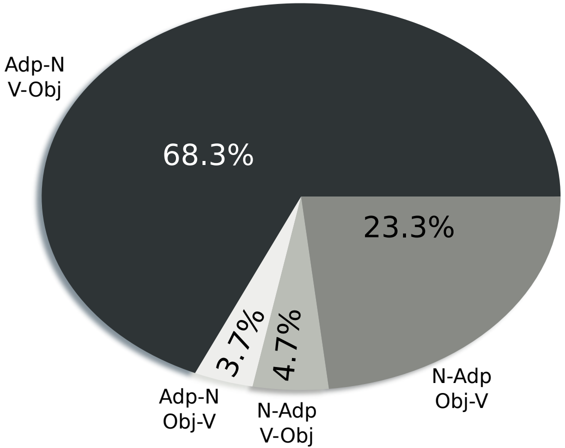
Word order and case





Word order correlations





- Maslova's program can be carried out with phylogenetic comparative method
- future research:
    - equilibrium distributions generally resemble family-wise weighted distributions — bug or feature?
    - hierarchical models instead of one Markov process for all lineages?
    - more data!!! (but there are never enough of them)
    - better methods for feature selection?

References

Judith Aissen. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21(3):435–483, 2003.

Balthasar Bickel, Alena Witzlack-Makarevich, and Taras Zakharko. Typological evidence against universal effects of referential scales on case alignment. In Ina Bornkessel-Schlesewsky, Andrej L. Malchukov, and Marc D. Richards, editors, *Scales and hierarchies: A cross-disciplinary perspective*, pages 7–43. de Gruyter, Berlin/Munich/Boston, 2015.

Jonathan P. Bollback. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*, 7(1):88, 2006.

Bernard Comrie. *Language Universals and Linguistic Typology*. Basil Blackwell, Oxford, 1981.

Michael Dunn, Simon J. Greenhill, Stephen Levinson, and Russell D. Gray. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473(7345): 79–82, 2011.

Ramon Ferrer-i-Cancho. Kauffman's adjacent possible in word order evolution. arXiv preprint arXiv:1512.05582, 2015.

Murray Gell-Mann and Merritt Ruhlen. The origin and evolution of word order. *Proceedings of the National Academy of Sciences*, 108(42):17290–17295, 2011.

Joseph Greenberg. Some universals of grammar with special reference to the order of meaningful elements. In *Universals of Language*, pages 73–113. MIT Press, Cambridge, MA, 1963.

Sebastian Höhna, Michael J. Landis, Tracy A. Heath, Bastien Boussau, Nicolas Lartillot, Brian R. Moore, John P. Huelsenbeck, and Frederik Ronquist. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic biology*, 65(4):726–736, 2016.

Gerhard Jäger. Global-scale phylogenetic linguistic inference from lexical resources. *Scientific Reports*, 5, 2018. https://www.nature.com/articles/sdata2018189.

Stephen C. Levinson and Russell D. Gray. Tools from evolutionary biology shed new light on the diversification of languages. *Trends in Cognitive Sciences*, 16(3):167–173, 2012.

Elena Maslova. A dynamic approach to the verification of distributional universals. *Linguistic Typology*, 4(3):307–333, 2000.

Mark Pagel and Andrew Meade. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist*, 167(6): 808–825, 2006.

Mark Pagel and Andrew Meade. BayesTraits 2.0. software distributed by the authors, November 2014.

Frederik Ronquist and John P. Huelsenbeck. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572–1574, 2003.

Michael Silverstein. Hierarchy of features and ergativity. In R. M. W. Dixon, editor, *Grammatical Categories in Australian Languages*, pages 112–171. Australian Institute of Aboriginal Studies, Canberra, 1976.

Søren Wichmann, Eric W. Holman, and Cecil H. Brown. The ASJP database (version 18). http://asjp.clld.org/, 2018.