

On the proper use of phylogenetic information in typology

Gerhard Jäger

Tübingen University

Workshop *Phylogenetic Linguistics and Linguistic Theory*

York, November 15, 2018



WORDS BONES GENES TOOLS
Tracking Linguistic, Cultural, and Biological Trajectories of the Human Past

FRIEDRICH-SCHILLER
UNIVERSITÄT
TÜBINGEN

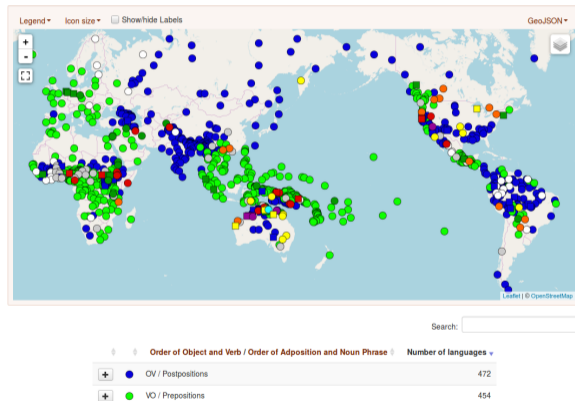


DFG

Introduction

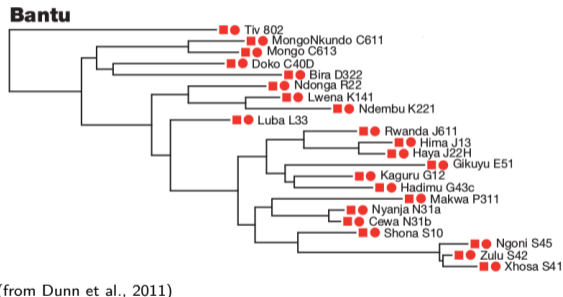
Word order correlations

- Greenberg, Keenan, Lehmann etc.: general tendency for languages to be either consistently head-initial or consistently head-final
- alternative account (Dryer, Hawkins): phrases are consistently left- or consistently right-branching
- can be formalized as collection of implicative universals, such as
With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional. (Greenberg's Universal 4)
- both generativist and functional/historical explanations in the literature



Phylogenetic non-independence

- 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

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.”

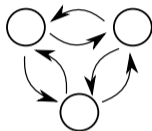
*“In this case, the only way to discover a distributional universal is to **estimate transition probabilities** and as it were to ‘predict’ the stationary distribution on the basis of the equations in (1).”*



The phylogenetic comparative method

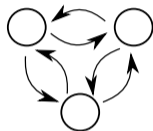
Modeling language change

Markov process



Modeling language change

Markov process

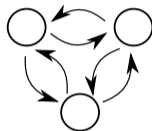


Phylogeny



Modeling language change

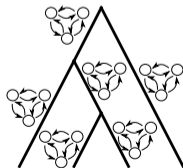
Markov process



Phylogeny

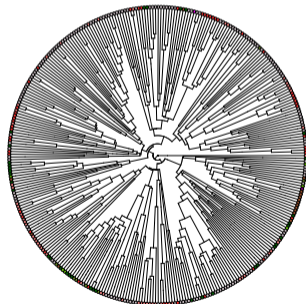


Branching process



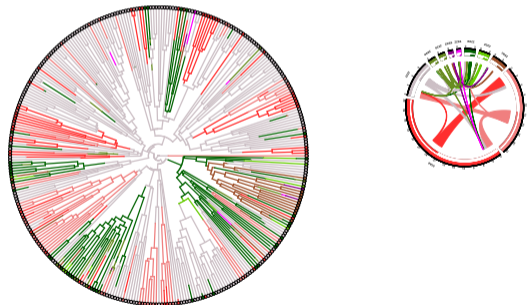
Estimating rates of change

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



Estimating rates of change

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

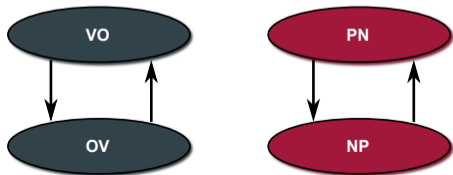


Correlation between features

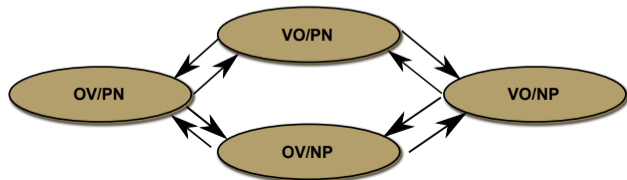
Pagel and Meade (2006)

- construct two types of Markov processes:
 - **independent:** the two features evolve according to independent Markov processes
 - **dependent:** rates of change in one feature depends on state of the other feature
- fit both models to the data
- apply statistical model comparison

Independent model



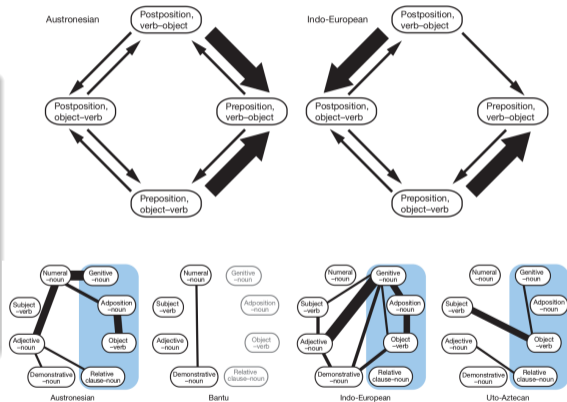
Dependent model



Dunn et al. (2011)

Dunn et al. (2011)

- all 28 pairs of 8 word-order features considered
- 4 language families: Austronesian, Bantu, Indo-European, and Uto-Aztecan
- main finding: wildly different results between families
- conclusion:
word-order correlations are lineage-specific

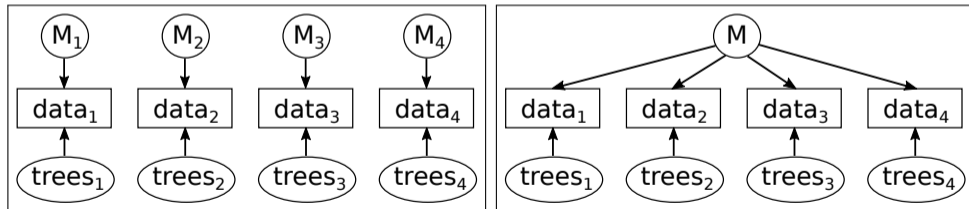


Universal and lineage-specific models

This study

Experiments

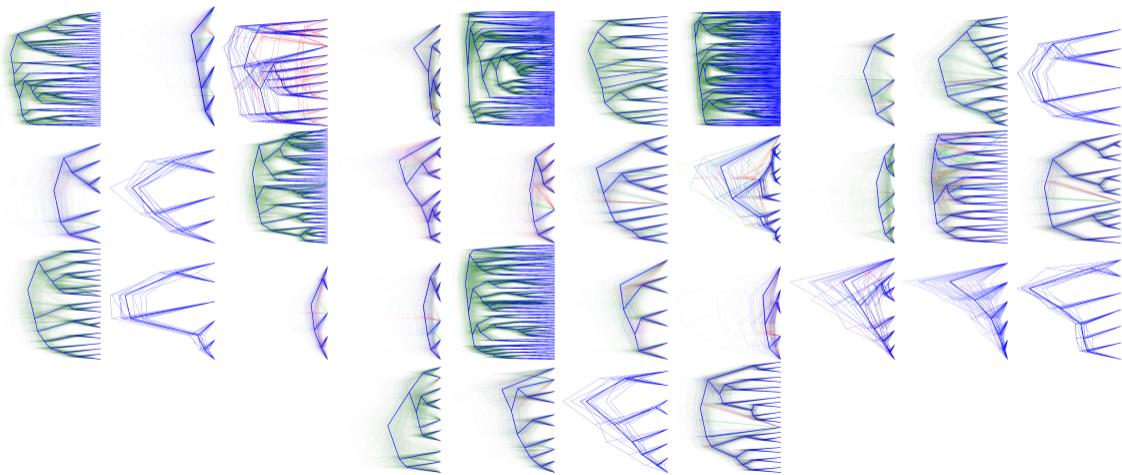
- 1 replication of Dunn et al. (2011) with different data
- 2 model comparison: universal vs. lineage-specific correlations
- 3 word-order correlations across a comprehensive collection of language families



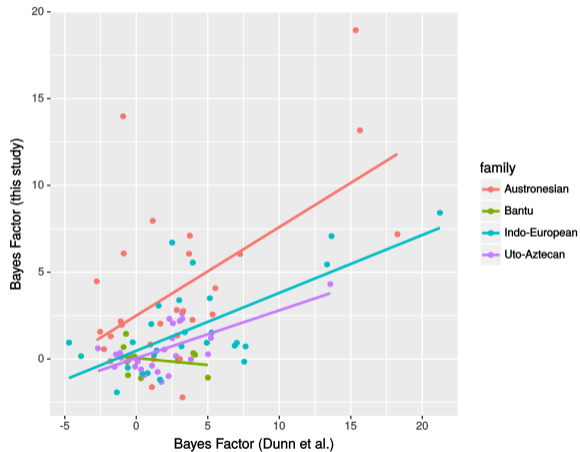
Data

- **word-order data:** WALS
- **phylogeny:**
 - ASJP word lists (Wichmann et al., 2016)
 - feature extraction (automatic cognate detection, *inter alia*) \rightsquigarrow character matrix
 - Maximum-Likelihood phylogenetic inference with Glottolog (Hammarström et al., 2016) tree as backbone
 - advantages over hand-coded Swadesh lists
 - applicable across language families
 - covers more languages than those for which expert cognate judgments are available
 - 1004 languages in total
 - Austronesian: 123; Bantu: 41; Indo-European: 53; Uto-Aztecan: 13
 - 34 families with at least five languages; comprising 768 languages in total

Phylogenetic tree sample



Replication of Dunn et al.



Comparing universal and lineage-specific models

- so far: fitting a separate model for each language family
 - **advantage:** good fit of the lineage-specific data
 - **disadvantage:** many parameters (8 per family for a dependent model)
- statistical model comparison: quantifying to what degree the data support the excess parameters of lineage-specific models
- models to be compared:
 - **universal:** one set of rates (8 parameters), applying to all 4 families
 - **lineage specific:** a separate set of rates for each family
- comparison via **Bayes Factor**
(implementation with RevBayes; Höhna et al. 2016)

Results

universal vs. lineage specific

feature pair	Bayes Factor
Adp-N V-Obj	58.1
Adp-N N-Gen	47.2
N-Adj N-Rel	41.6
N-Gen V-Obj	36.9
Adp-N V-Subj	23.6
N-Gen N-Rel	21.9
N-Dem N-Num	20.6
Adp-N N-Rel	18.7
V-Obj N-Rel	18.1
N-Dem N-Rel	17.4
N-Rel V-Subj	14.5
N-Gen V-Subj	13.7
V-Obj V-Subj	12.1
N-Adj N-Dem	5.4
Adp-N N-Dem	-5.3
N-Dem N-Gen	-5.7
N-Adj N-Num	-5.8
N-Adj V-Subj	-12.3
N-Dem V-Obj	-12.8
N-Num N-Rel	-15.7
N-Adj Adp-N	-17.0
N-Dem V-Subj	-18.6
N-Adj V-Obj	-22.0
N-Adj N-Gen	-23.2
Adp-N N-Num	-28.2
N-Gen N-Num	-34.1
N-Num V-Subj	-37.6
N-Num V-Obj	-45.4

universal

lineage-specific

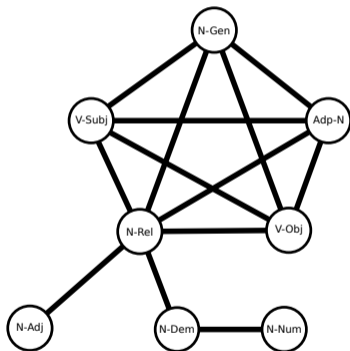
correlated vs. independent

feature pair	Bayes Factor
Adp-N N-Gen	115.7
Adp-N V-Obj	104.8
N-Dem N-Num	99.6
N-Adj N-Num	93.3
N-Gen V-Obj	68.0
N-Adj N-Dem	64.9
N-Adj N-Rel	48.5
N-Gen V-Subj	41.1
V-Obj V-Subj	38.2
V-Obj N-Rel	35.3
N-Dem N-Rel	33.5
Adp-N V-Subj	31.3
N-Gen N-Rel	23.8
N-Dem N-Gen	23.5
Adp-N N-Rel	22.6
N-Gen N-Num	16.5
N-Dem V-Obj	15.4
Adp-N N-Dem	15.0
N-Num V-Subj	14.4
Adp-N N-Num	13.5
N-Adj N-Gen	12.2
N-Num V-Obj	7.6
N-Rel V-Subj	6.8
N-Num N-Rel	5.0
N-Adj Adp-N	0.3
N-Adj V-Subj	-0.3
N-Adj V-Obj	-1.0
N-Dem V-Subj	-1.2

correlated

uncorrelated

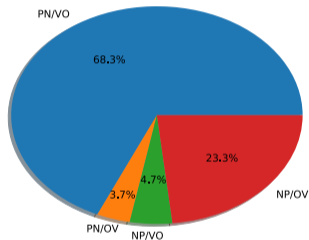
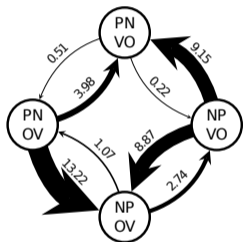
Results



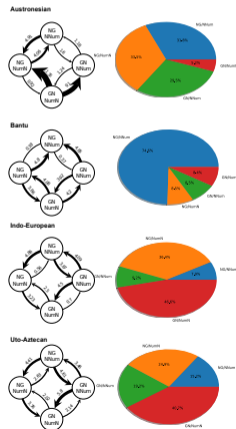
- one tightly connected cluster of mutually universally correlated word order features
- comprises Dryer's (1992) *verb patterners* + V-Subj
- additionally some correlations regarding NP syntax

Results

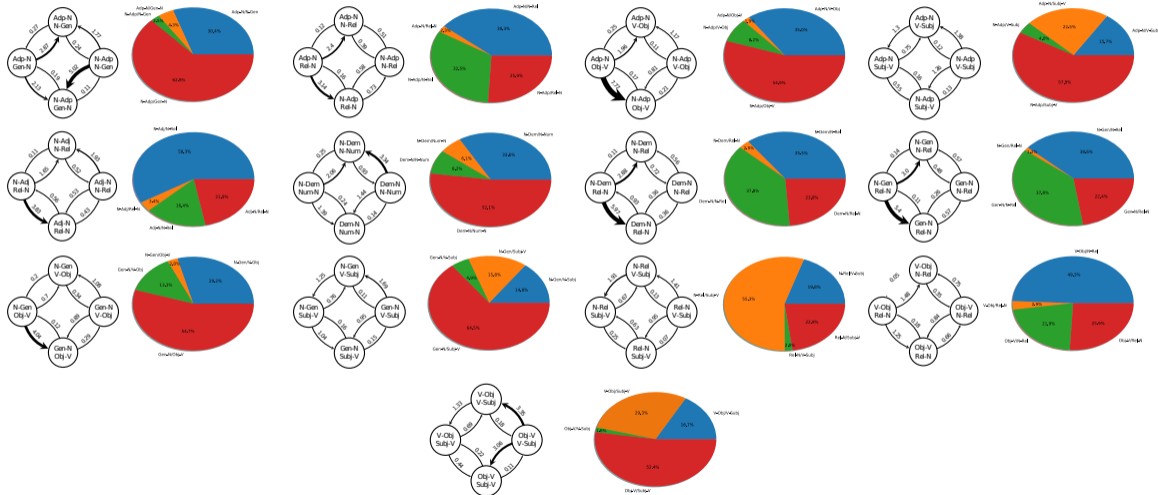
universal (AdvP-N/V-Obj)



lineage-specific (N-Gen/N-Num)

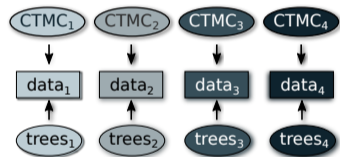


What the universal dependencies look like

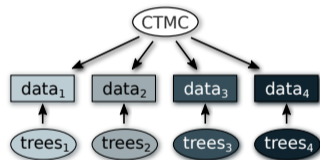


Hierarchical Models

Hierarchical Bayesian models

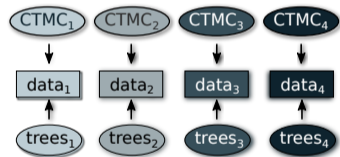


lineage-specific

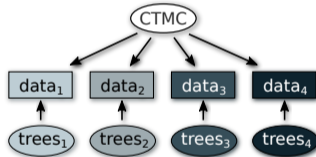


universal

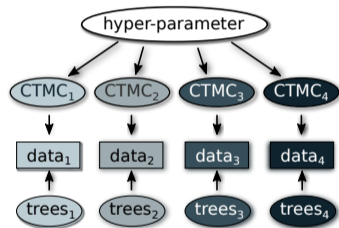
Hierarchical Bayesian models



lineage-specific



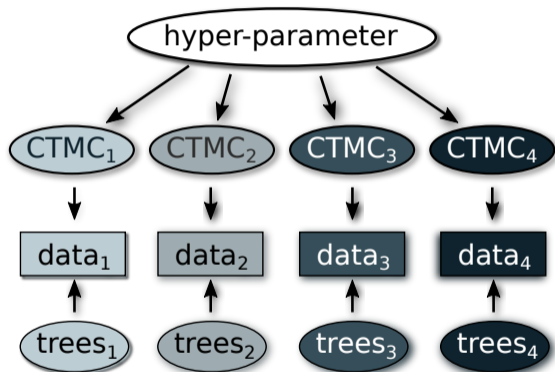
universal



hierarchical

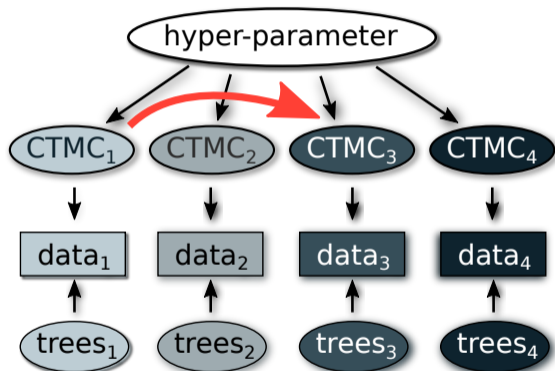
Hierarchical Models

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



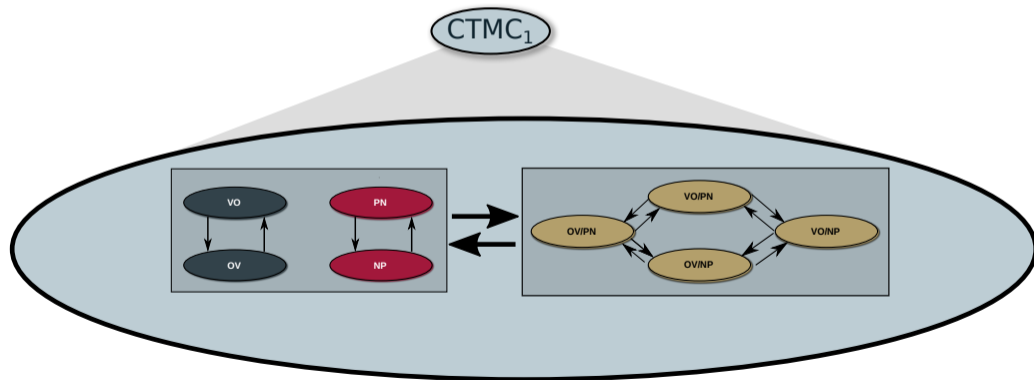
Hierarchical Models

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



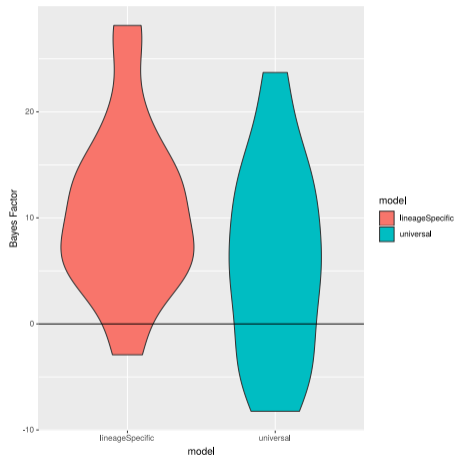
Trans-dimensional parameter estimation

- Which version should we choose for $CTMC_i$ – the dependent or the independent one?
- Choice can be left to the data via **trans-dimensional** parameter estimation
- a.k.a. *Reversible-Jump Markov Chain Monte Carlo*




Model comparison

- overall, hierarchical model outperforms both lineage specific and universal model
- exceptions in extreme cases



Posterior probability of dependent model

feature pair	P(dependent model data)
Adp-N N-Gen	0.94
Adp-N V-Obj	0.93
N-Adj N-Num	0.91
V-Obj V-Subj	0.90
N-Gen V-Obj	0.89
V-Obj N-Rel	0.88
N-Dem N-Num	0.87
N-Dem N-Gen	0.87
N-Adj N-Gen	0.87
N-Rel V-Subj	0.87
N-Adj N-Dem	0.86
Adp-N N-Rel	0.85
Adp-N V-Subj	0.85
N-Dem V-Subj	0.83
N-Adj N-Rel	0.83
N-Gen V-Subj	0.83
N-Gen N-Rel	0.81
N-Adj V-Obj	0.80
Adp-N N-Dem	0.80
N-Adj Adp-N	0.77
N-Gen N-Num	0.76
Adp-N N-Num	0.75
N-Dem V-Obj	0.73
N-Num V-Obj	0.70
N-Dem N-Rel	0.68
N-Num N-Rel	0.59
N-Adj V-Subj	0.58
N-Num V-Subj	0.57



dependent

independent

Intermediate summary

Intermediate summary

- strong signal for universal word-order correlations, e.g.
 - Adp-N / V-Obj
 - Adp-N / N-Gen
 - N-Gen / V-Obj
 - N-Gen / V-Subj
 - N-Dem / N-Num
 - N-Adj / N-Rel
 - V-Obj / V-Subj
 - V-Obj / N-Rel
- signal only becomes apparent if we look at several families **simultaneously**
- Bayesian hierarchical models:
 - allows the model fit for individual families to inform each other
 - lets the data decide to what degree patterns are universal and to what degree lineage-specific

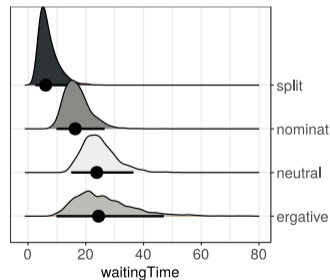
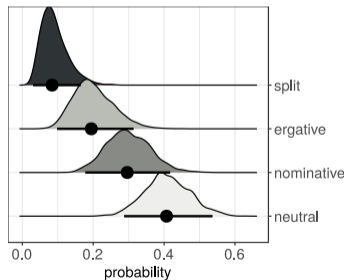
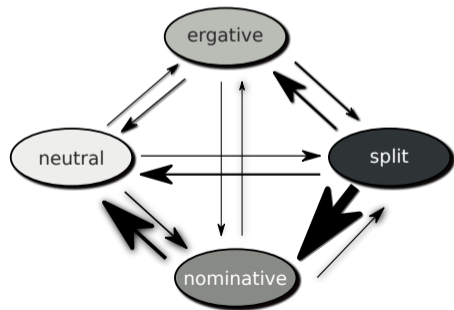
Further applications (work in progress)

Case marking patterns

- Maslova and Nikitina (2007): implementation of Maslova's (2000) program
- rate estimation of CTMC by using two heuristics:
 - how many languages of type A occur in a predominantly B -family
 - how many pairs of closely related languages differ in their type
- no phylogenetic information of intermediate time depths
- no branch length information
- universality is assumed *a priori*
- conclusion: *nominative* is at least three times as likely as *ergative* in the equilibrium distribution of the CTMC

Case marking patterns

- data: from Maslova and Nikitina (2007)
- intersected with (character-transformed) ASJP data
- 260 languages from 23 families



Case marking patterns

- main conclusions
 - with 85% posterior probability, nominative is more likely in equilibrium than ergative
 - with 82% posterior probability, ergative is more likely than nominative
 - very high degree of uncertainty

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

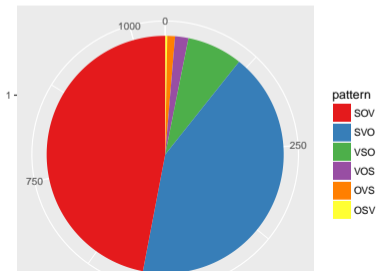
Raw numbers

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

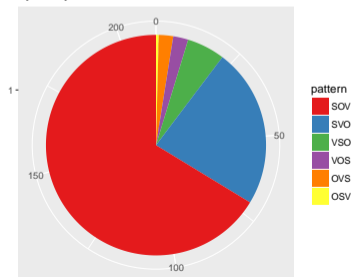
Weighted by lineages

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%

by language



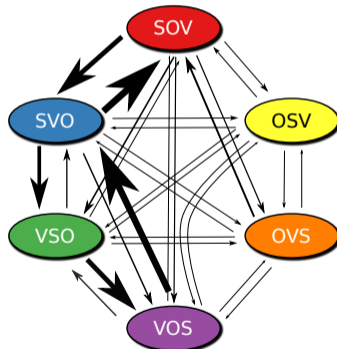
by family



Estimating transition rates

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

expected strength of flow



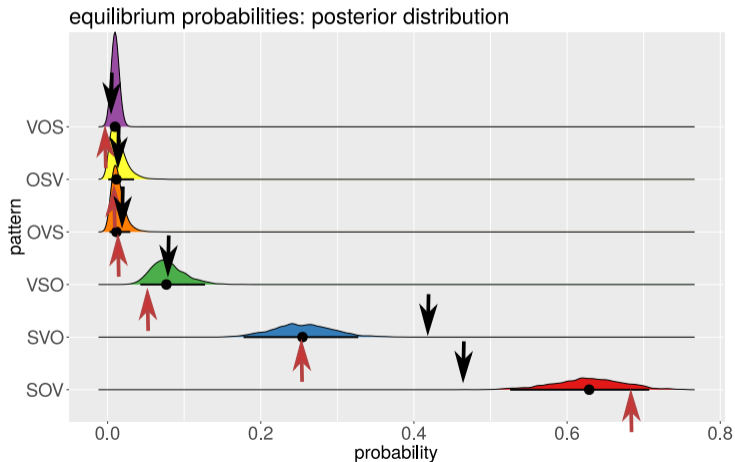
Reconstruction history with SIMMAP

- 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]	–	

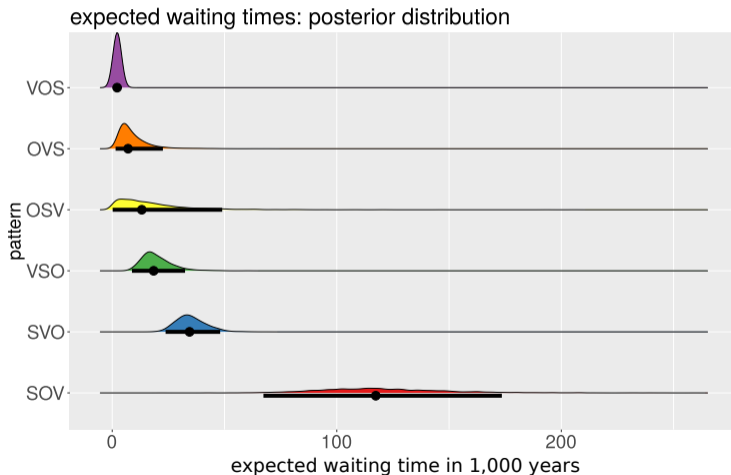
Posterior distributions

Empirical vs. estimated distribution



Posterior distributions

Waiting times



- Matthew S. Dryer. The Greenbergian word order correlations. *Language*, 68(1):81–138, 1992.
- 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.
- Harald Hammarström, Robert Forkel, Martin Haspelmath, and Sebastian Bank. *Glottolog 2.7*. Max Planck Institute for the Science of Human History, Jena, 2016. Available online at <http://glottolog.org>, Accessed on 2017-01-29.
- Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie. The World Atlas of Language Structures online. Max Planck Digital Library, Munich, 2008. <http://wals.info/>.
- 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.
- Elena Maslova. A dynamic approach to the verification of distributional universals. *Linguistic Typology*, 4(3):307–333, 2000.
- Elena Maslova and E. Nikitina. Stochastic universals and dynamics of cross-linguistic distributions: the case of alignment types. unpublished manuscript, Stanford University, 2007.
- 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.
- Søren Wichmann, Eric W. Holman, and Cecil H. Brown. The ASJP database (version 17). <http://asjp.cldd.org/>, 2016.