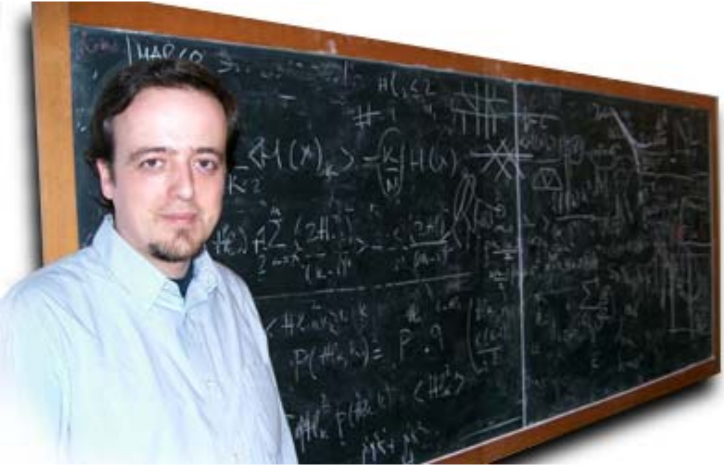
Harnessing Bayesian phylogenetics to test a Greenbergian universal

Gerhard Jäger[1] Ramon Ferrer-i-Cancho[2]

Tübingen University[1]

Universitat Politècnica de Catalunya[2]

52nd Annual Meeting of the Societas Linguistica Europaea
Leipzig, August 21, 2019

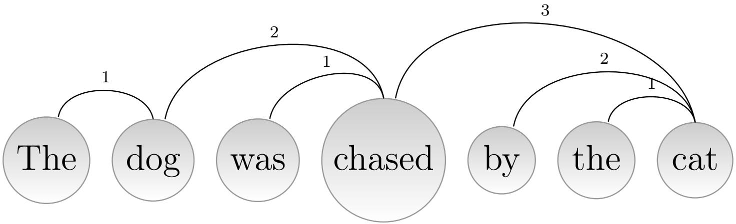


Greenberg's Universal 17

With overwhelmingly more than chance frequency, languages with dominant order VSO have the adjective after the noun. (Greenberg, 1963)

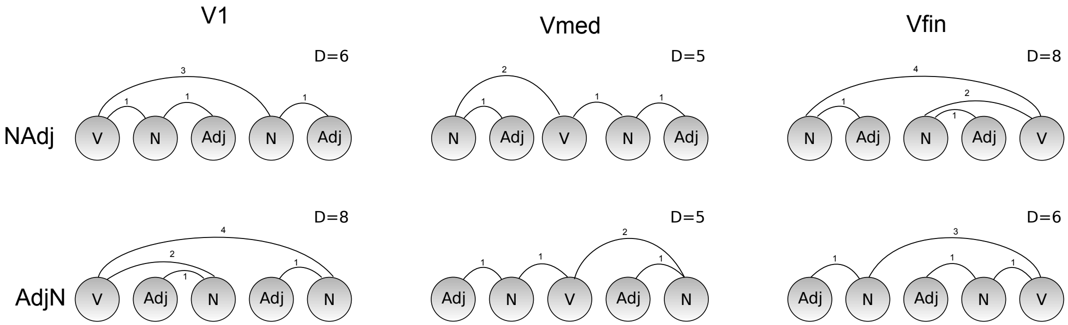
Mirror image: Verb-final languages prefer adjective-noun order.

But: Dryer (1992)



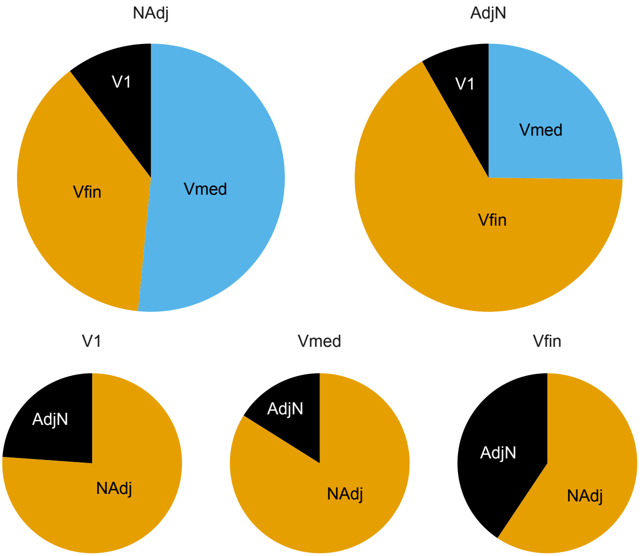
$n = 7,\ D = 10$

- Dependency distances.
- DDm: dependency distance minimization principle (Liu et al., 2017).
- Cognitive origins of DDm: interference and decay (Liu et al., 2017).
- The challenge of aggregating D over heterogeneous data: sentences of different lengths, multiple authors, ... (Ferrer-i-Cancho and Liu, 2014)

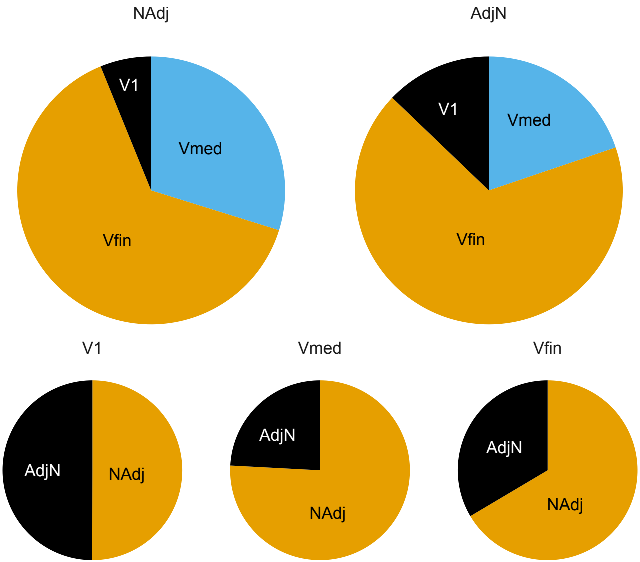


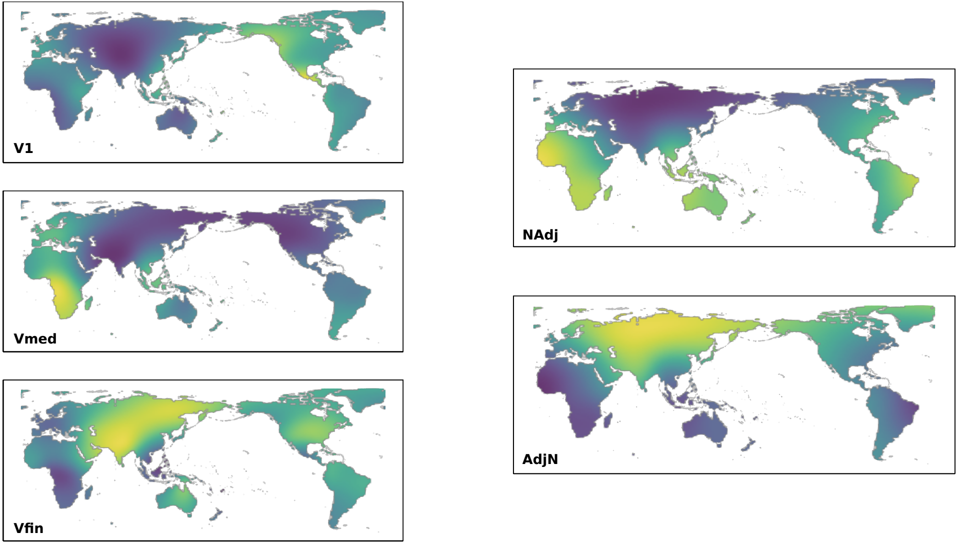
DDm provides functional motivation for Universal 17 and its mirror image.

Frequency distribution (WALS)



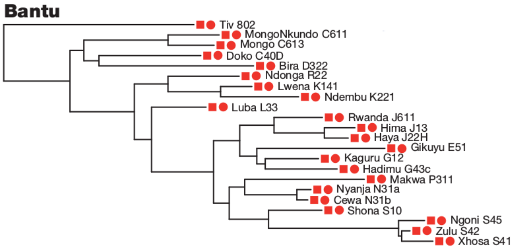
Frequency distribution, weighted by lineage




V1
Vmed
Vfin
NAdj
AdjN

- 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



(from Dunn et al., 2011)

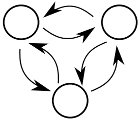
Maslova (2000):

> *"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."*

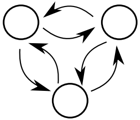
> *"In this case, the only way to discover a distributional universal is to* ***estimate transition probabilities*** *and as it were to 'predict' the stationary distribution on the basis of the equations in (1)."*

The phylogenetic comparative method

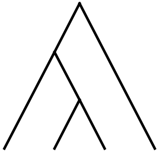
Markov process



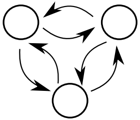
Markov process



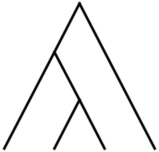
Phylogeny



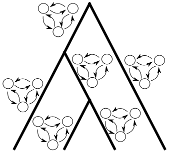
Markov process



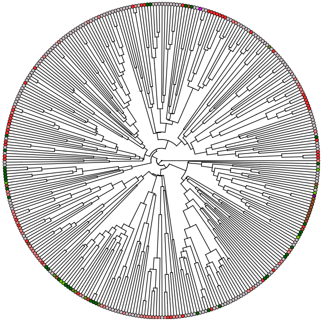
Phylogeny



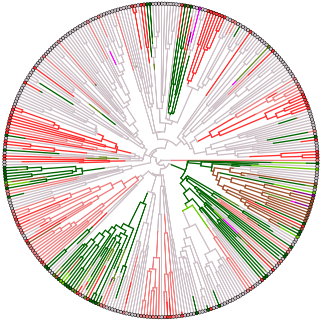
Branching process

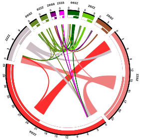


- if phylogeny and states of extant languages are known...



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



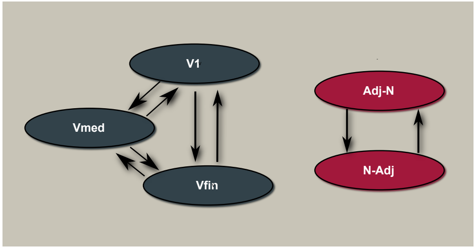


Correlation between features

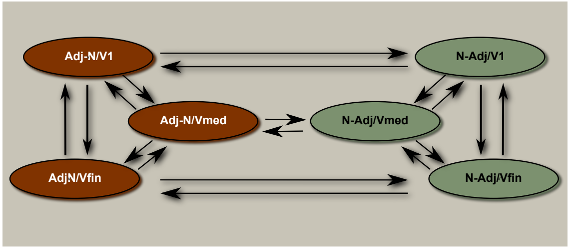
Pagel and Meade (2006)

- construct two types of Markov processes:
  - **independent:** the two features evolve according to independend 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

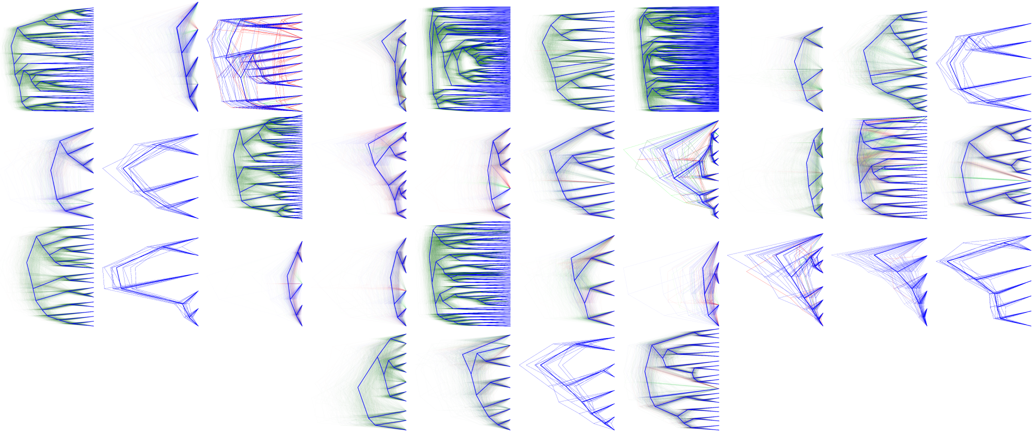
Independent model

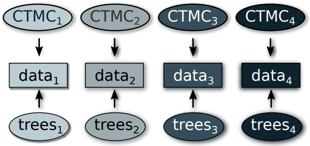


Dependent model

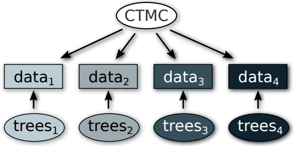


- **word-order data:** WALS
- **phylogeny:**
  - ASJP word lists (Wichmann et al., 2016)
  - feature extraction (automatic cognate detection, *inter alia*) $\rightsquigarrow$ character matrix
  - Bayesian phylogenetic inference with Glottolog (Hammarström et al., 2016) tree as backbone
  - advantages over hand-coded Swadesh lists
    - applicable across language familes
    - covers more languages than those for which expert cognate judgments are available
  - 902 languages in total
  - 76 families and 105 isolates

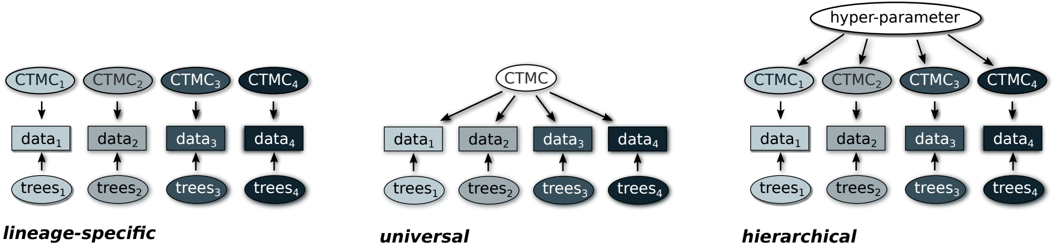




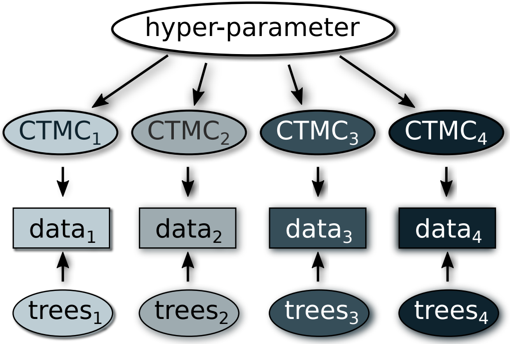
lineage-specific



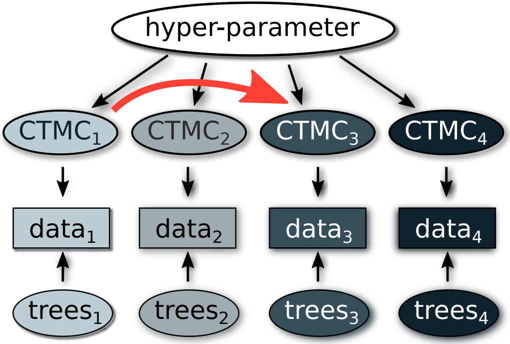
universal


CTMC1
CTMC2
CTMC3
CTMC4
data1
data2
data3
data4
trees1
trees2
trees3
trees4
lineage-specific
CTMC
data1
data2
data3
data4
trees1
trees2
trees3
trees4
universal
hyper-parameter
CTMC1
CTMC2
CTMC3
CTMC4
data1
data2
data3
data4
trees1
trees2
trees3
trees4
hierarchical

- each family 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-family variation $\rightarrow$ can be overwritten by the data



- each family 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-family variation $\rightarrow$ can be overwritten by the data
- enables **information flow across families**

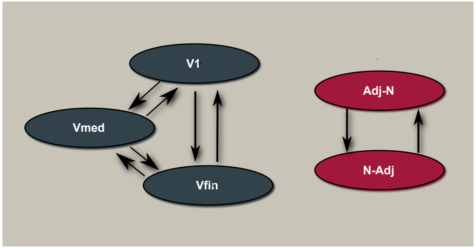


What about isolates?

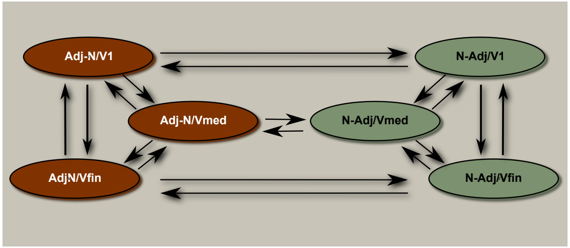
- Continuous Time Markov Chain defines a unique **equilibrium distribution**
- hierarchical model assumes a different CTMC, and thus a different equilibrium distribution for each lineage
- by modeling assumption, root state of a lineage is drawn from this distribution (Uniformity Principle)
- isolates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

Results

Independent model



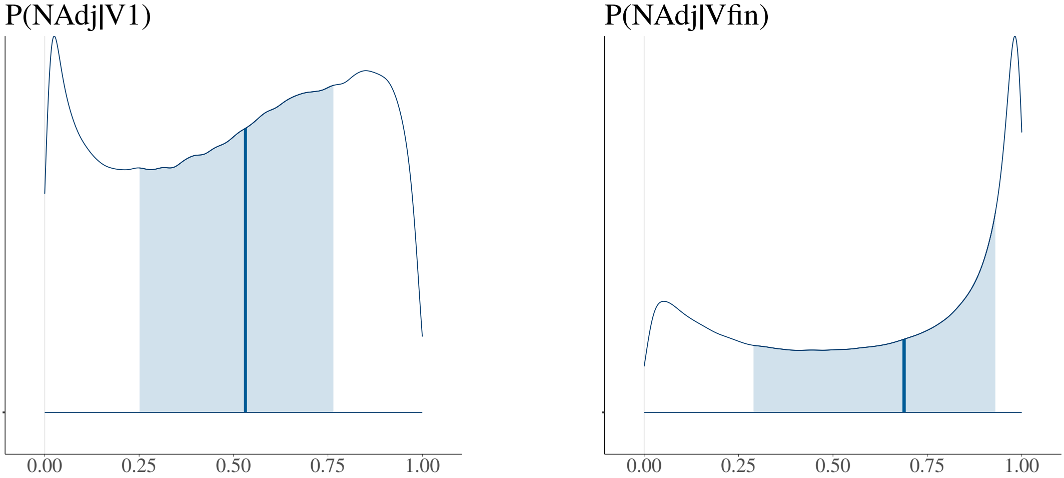
Dependent model



- **Bayes Factor:** 260 in favor of dependent model[1]

[1] In the abstract we reported the opposite conclusion, but there we used a non-hierarchical universal model.

17



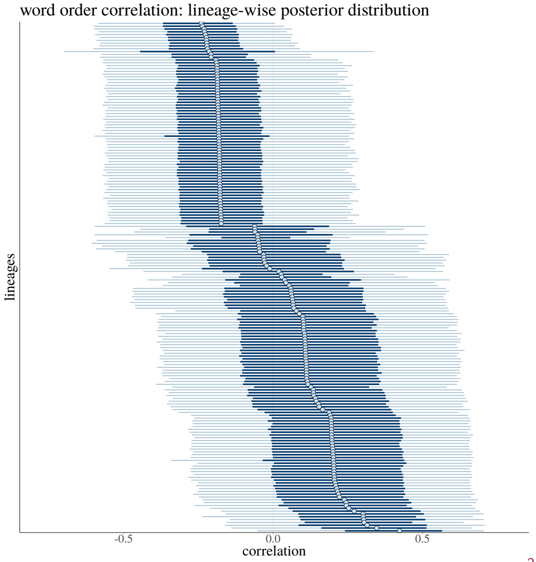
Correlation between verb order and adjective order

- lineages fall into two, about equally sized, groups:
  1. negative or no correlation

     Nuclear Macro-Je, Mande, Siouan, Pama-Nyungan, Austronesian, ...
  2. positive correlation

     Uto-Aztecan, Afro-Asiatic, Indo-Euroean, Dravidian, Austroasiatic,

     Otomanguean, ...

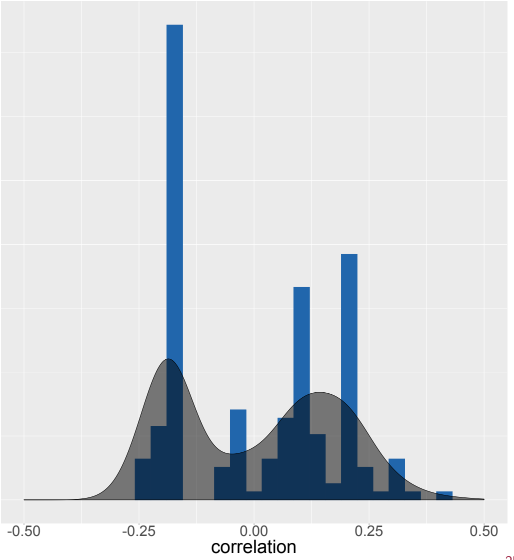


Correlation between verb order and adjective order

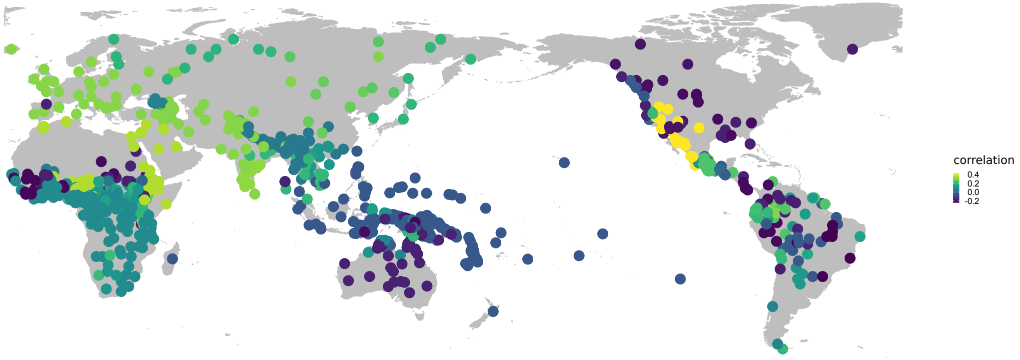
- lineages fall into two, about equally sized, groups:
  1. negative or no correlation

     Nuclear Macro-Je, Mande, Siouan, Pama-Nyungan, Austronesian, ...
  2. positive correlation

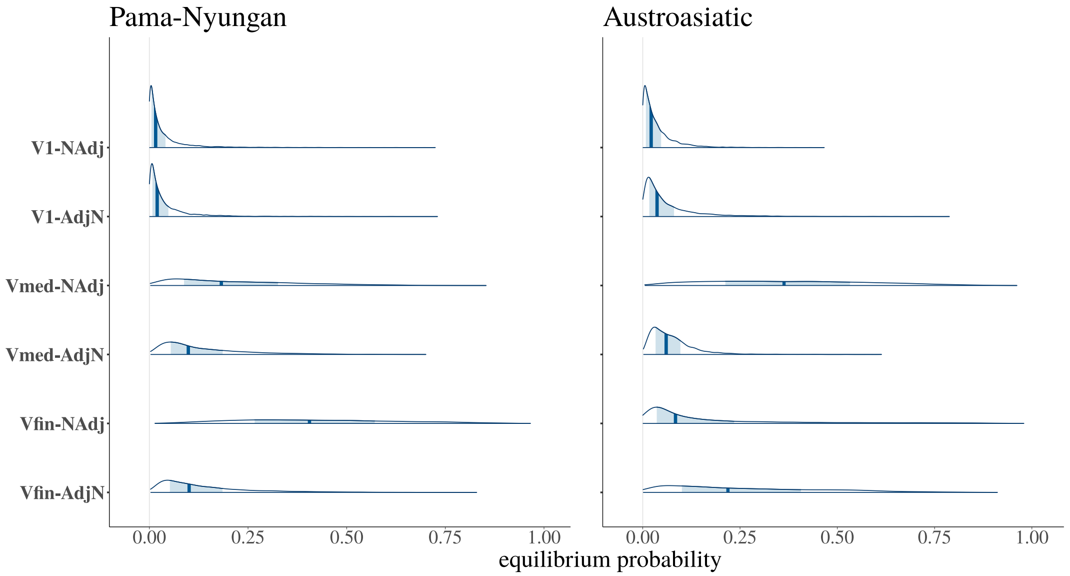
     Uto-Aztecan, Afro-Asiatic, Indo-Euroean, Dravidian, Austroasiatic, Otomanguean, ...



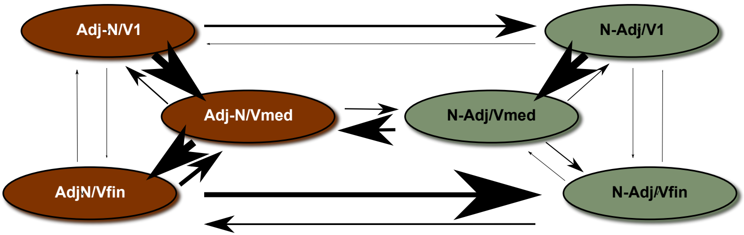
Correlation between verb order and adjective order



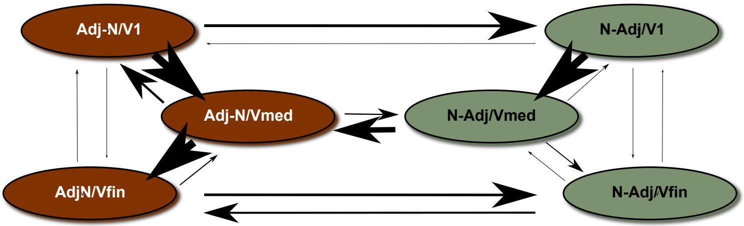
A representative family for each type



Pama-Nyungan



Austroasiatic



Conclusion

- no empirical support for Universal 17
- more nuanced picture for its mirror image:
  - two different possible dynamics governing relationship between verb-object and noun-adjective order
  - Dependency Length Minimization is operative in one dynamic, but not the other
  - reminds of an OT style pattern, with two competing constraints

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.

Ramon Ferrer-i-Cancho and H. Liu. The risks of mixing dependency lengths from sequences of different length. *Glottotheory*, (5):143–155, 2014.

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.

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/.

H. Liu, C. Xu, and J. Liang. Dependency distance: a new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21:171–193, 2017.

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.

Søren Wichmann, Eric W. Holman, and Cecil H. Brown. The ASJP database (version 17). http://asjp.clld.org/, 2016.