A phylogenetic perspective on semantic typology

Michael Dunn, Uppsala University

BAULT Symposium 2016

Semantic variation

Meaning varies through time

http://commons.wikimedia.org/wiki/File:Gervase_Babington.jpg

weak
ignorant

foolish

In pride we speak it, or at least inwardlie thinke it, wee are not as those seely Idiotes are. C16th

http://commons.wikimedia.org/wiki/File:Magpie_chasing_Brown_Goshawk_(Immature).jpg

weak
ignorant

Here we see that a small sillie Bird knoweth how to match with so great a Beast

by C16th

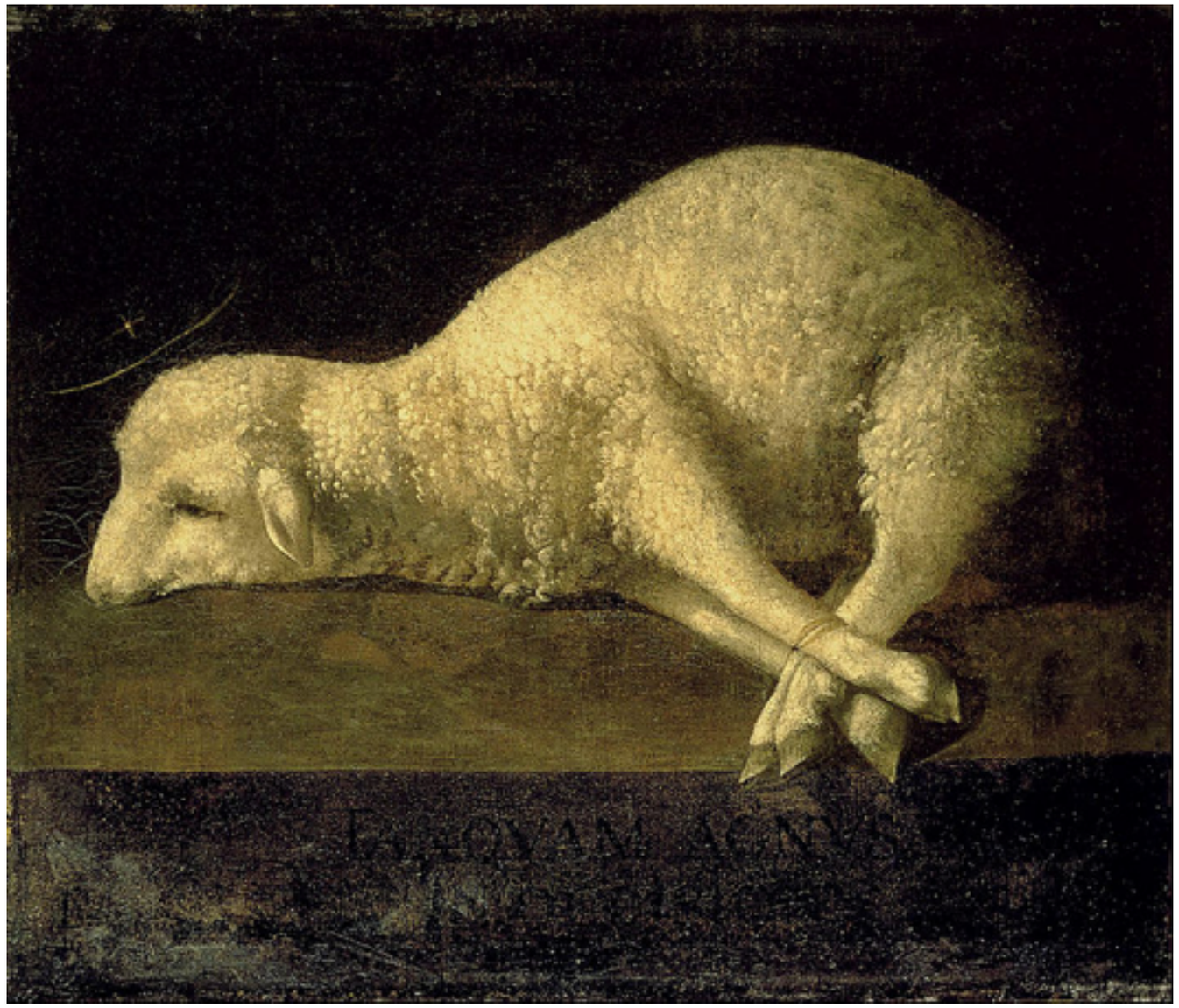
pitiable

Sely Scotland, that of helpe has gret neide

Later C14th

http://commons.wikimedia.org/wiki/File:Wallace,_as_depicted_in_a_children%27s_history_book_from_1906.jpg

innocent
harmless



‘Alas’, he seide, ‘þis seli best: þat no-þing ne doth a-mis!’

Late C14th

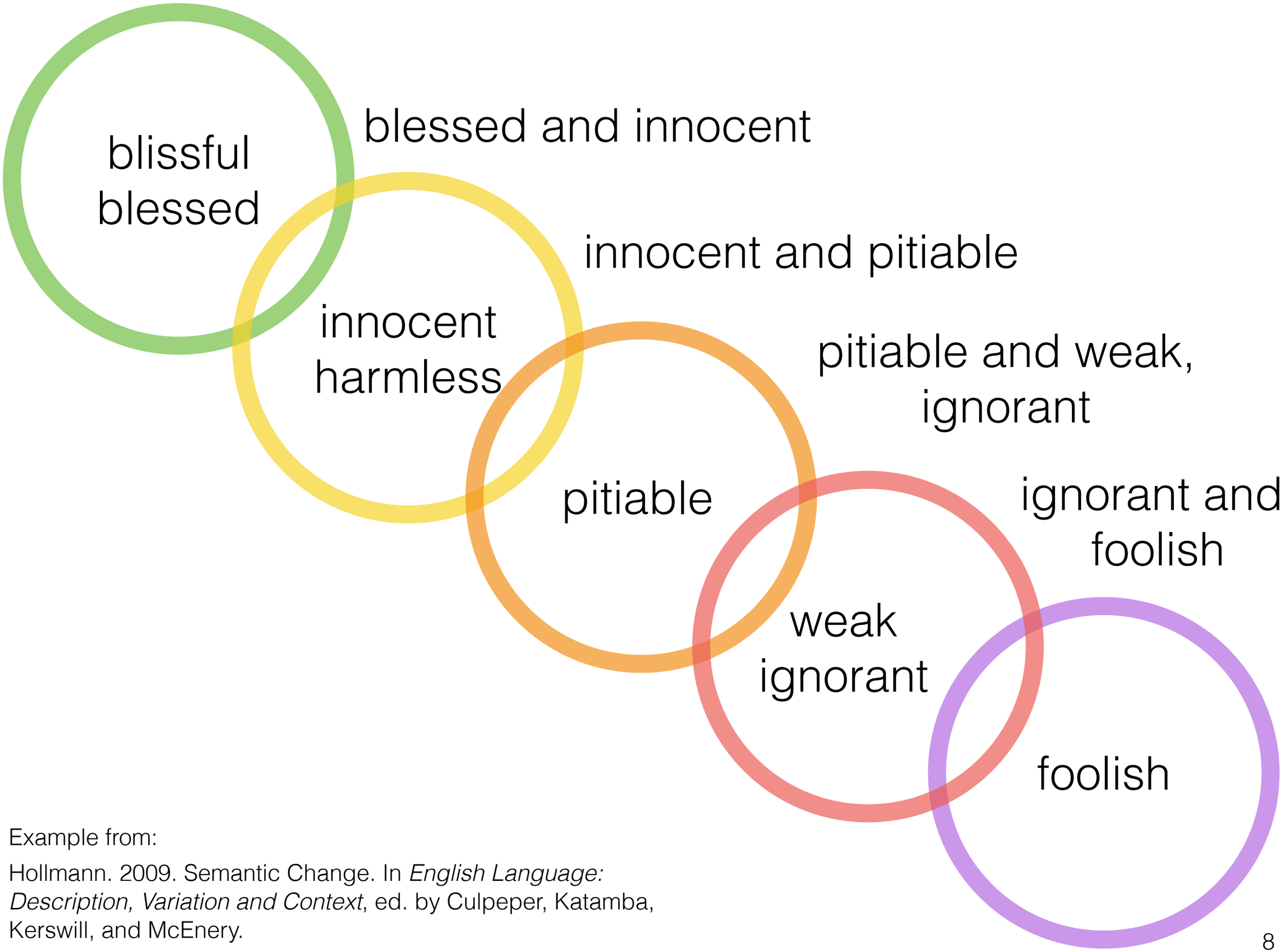
http://commons.wikimedia.org/wiki/File:Francisco_de_Zurbar%C3%A1n_-_Agnus_Dei_-_Google_Art_Project.jpg

blissful
blessed

A Jhesu,
blyssede [es] þat
abbaye and cely
is þat religione

http://commons.wikimedia.org/wiki/File:Godric-Finchale.jpg

1350 (examples from 1200)



Example from:

Hollmann. 2009. Semantic Change. In *English Language: Description, Variation and Context*, ed. by Culpeper, Katamba, Kerswill, and McEnery.

Meaning also varies through space

gas
first floor
corn
entree

https://commons.wikimedia.org/wiki/File:Queen_Elizabeth_II_1959.jpg

https://commons.wikimedia.org/wiki/File:Uncle_Sam_(pointing_finger).png

General patterns

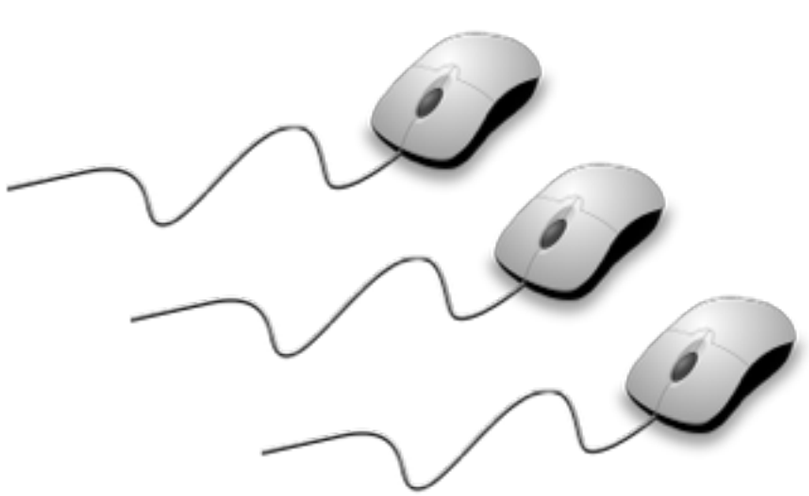
- *rhetorical devaluation* (Dahl 2001)
  terms expressing the extreme of positive evaluation (*excellent, wonderful, awesome*) tend to be overused → loss of informational strength → rapid turnover

- *collateral taboo*
  if one of the meanings of a polysemous/ homophonous word is obscene or otherwise fraught then the other meanings tend to drop out (*gay*)

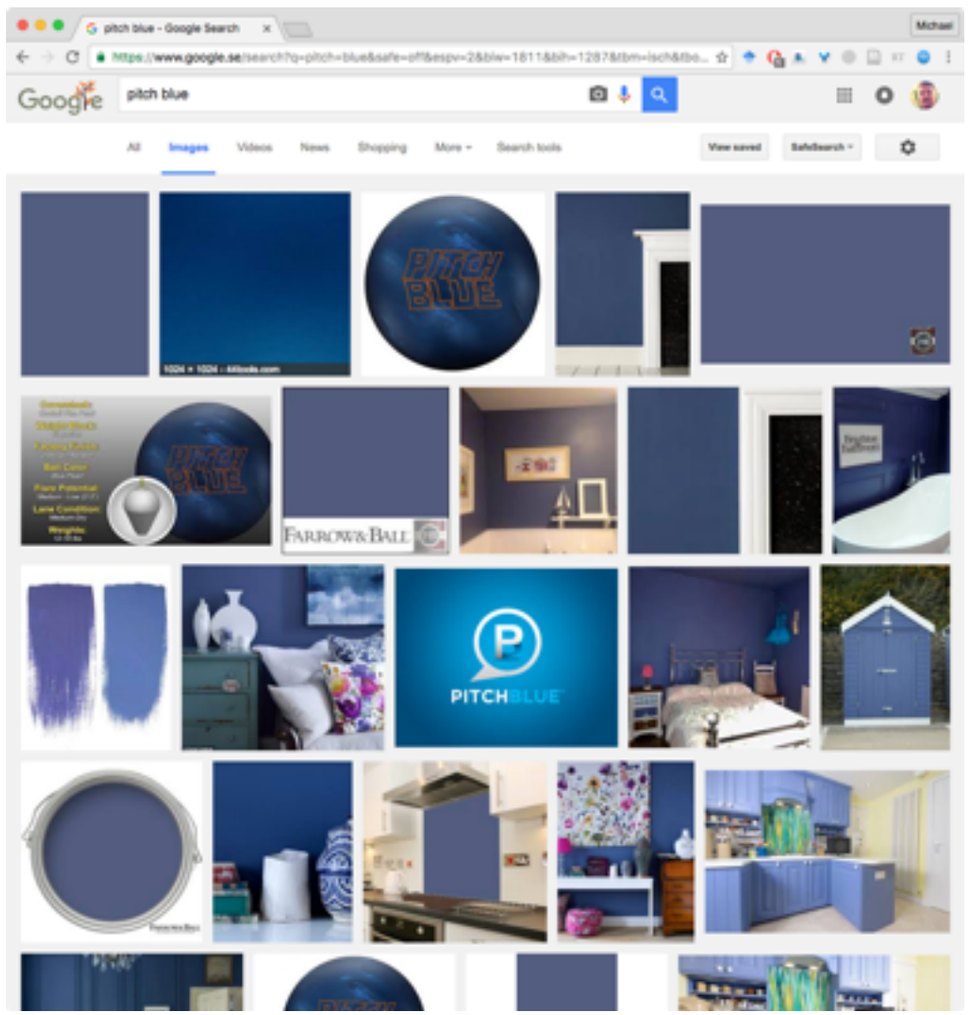
General patterns

- *bifurcation*
  a word acquires second meanings through reanalysis of some special use of the first meaning

mouse (sg.) mice (pl.)
mouse (sg.) mouses (pl.)



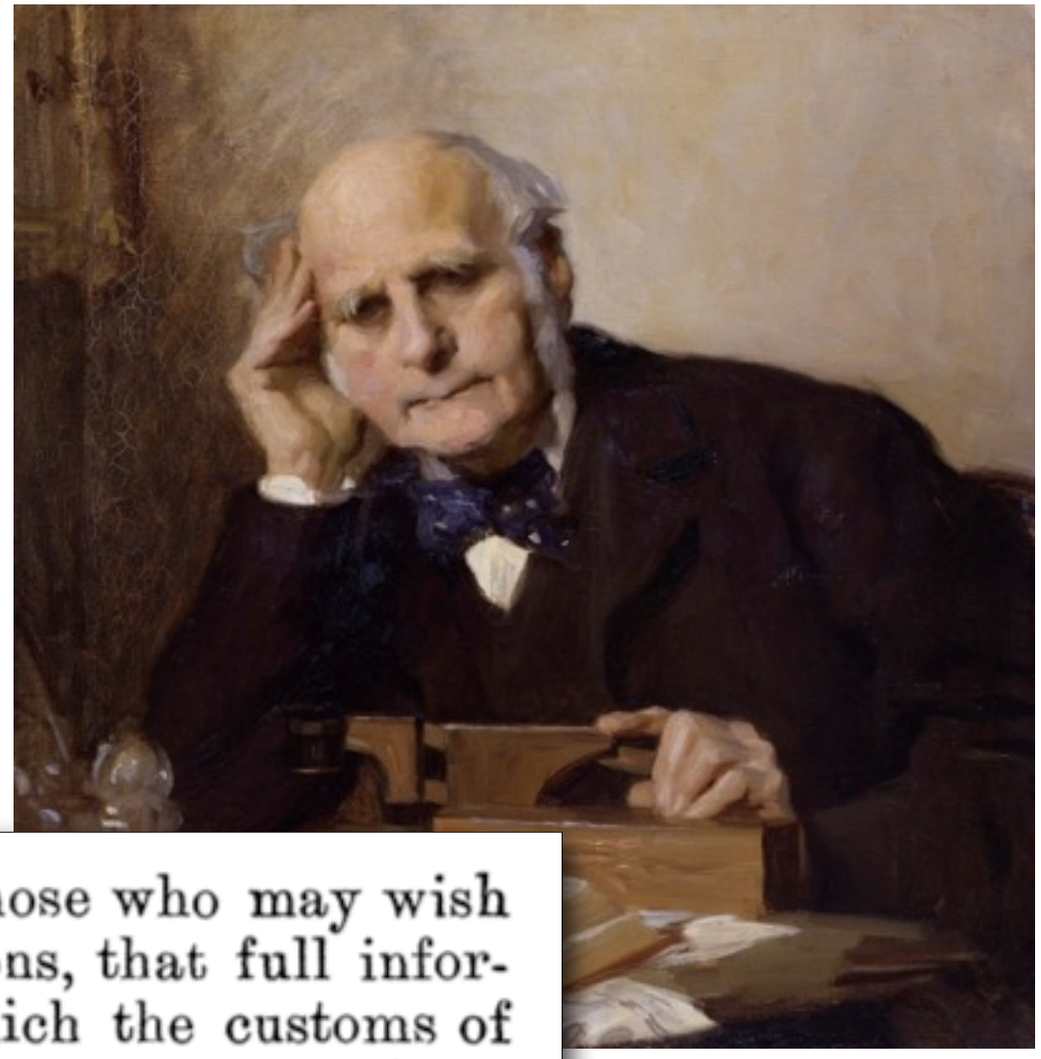
pitch black → pitch blue



Phylogenetic perspective on variation

Patterns are the result of processes

Galton's problem



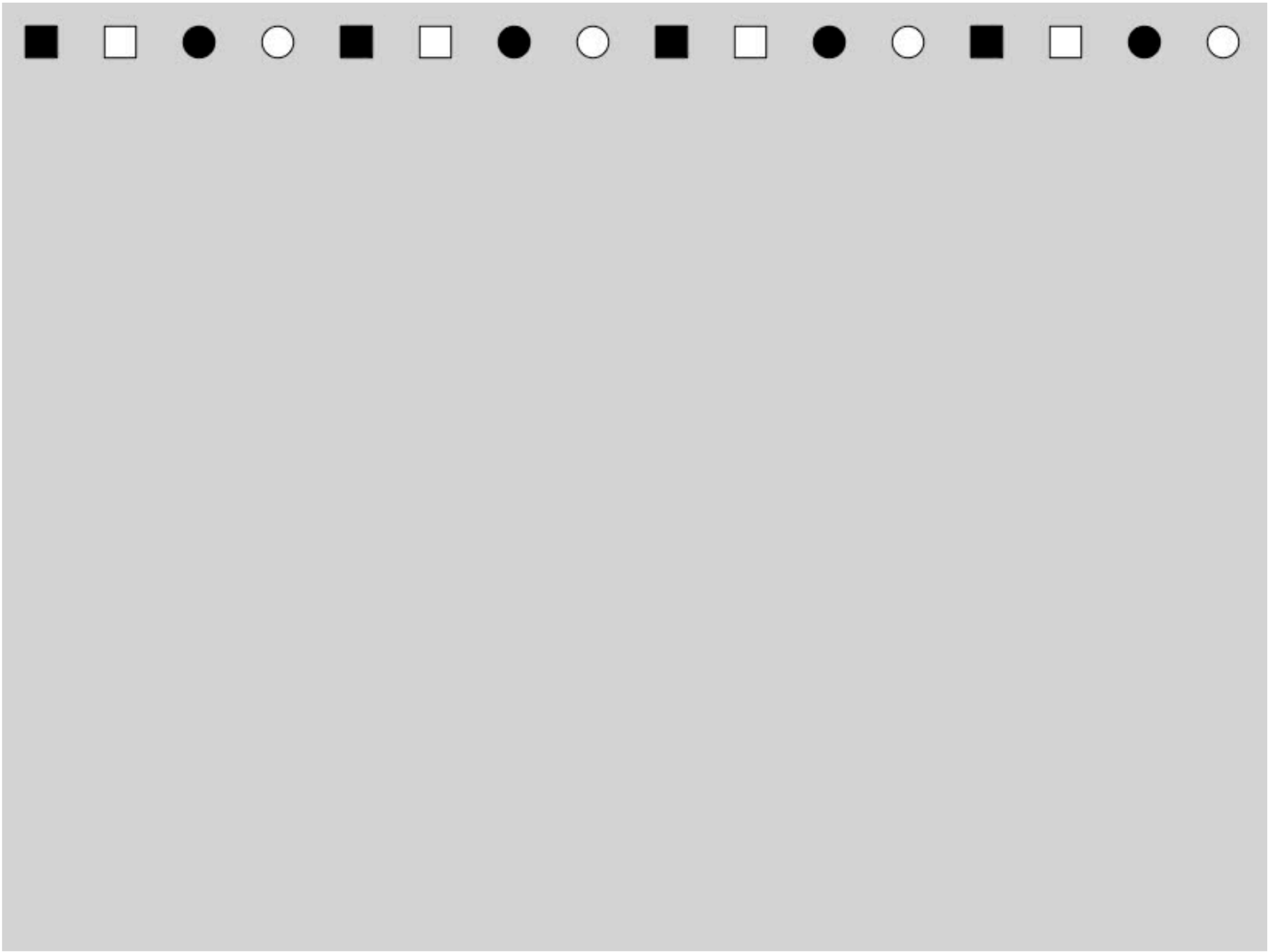
and failure was tabulated.

It was extremely desirable for the sake of those who may wish to study the evidence for Dr. Tylor's conclusions, that full information should be given as to the degree in which the customs of the tribes and races which are compared together are independent. It might be, that some of the tribes had derived them from a common source, so that they were duplicate copies of the same original. Certainly, in such an investigation as this, each of the observations ought, in the language of statisticians, to be carefully "weighted." It would give a useful idea of the distribution of the several customs and of their relative prevalence in the world, if a

Drift model

- Each figure (colour + shape, representing e.g. a language with a particular typological configuration) has 0-3 descendants
- Number of descendants is completely random

↓ generations



↓ generations

LETTER

doi:10.1038/nature09923

Evolved structure of language shows lineage-specific trends in word-order universals

Michael Dunn[1,2], Simon J. Greenhill[3,4], Stephen C. Levinson[1,2] & Russell D. Gray[3]

Languages vary widely but not without limit. The central goal of linguistics is to describe the diversity of human languages and explain the constraints on that diversity. Generative linguists following Chomsky have claimed that linguistic diversity must be constrained by innate parameters that are set as a child learns a language[1,2]. In contrast, other linguists following Greenberg have claimed that there are statistical tendencies for co-occurrence of traits reflecting universal systems biases[3–5], rather than absolute constraints or parametric variation. Here we use computational phylogenetic methods to address the nature of constraints on linguistic diversity in an evolutionary framework[6]. First, contrary to the generative acc
evolution of only a
strongly correlated. S
izations, we show th
between traits are line
These findings suppo
order—cultural evolu
linguistic structure,
shaping and constrai

Human language is
tems not only for its s
every level of structur
languages, some with j
than 100, some with c
simple words only, son
some in the middle, ar
and the systematic con
generative approach
diversity can be expla
these parameters cont
example, the setting 'l
verbs before objects (
('into the goal')[1,7]. Acc
when child learners si
tings other than those
tions such changes
language change acro
should therefore be re
must co-vary[8].

In contrast, the statis
samples languages to
occurring traits are ex
universal cognitive or
tendencies are the so-c
of elements in a claus
worldwide sample of 6
expected linkages betw
lation of the word-or
correlates with prepositions, as well as relative clauses and genitives

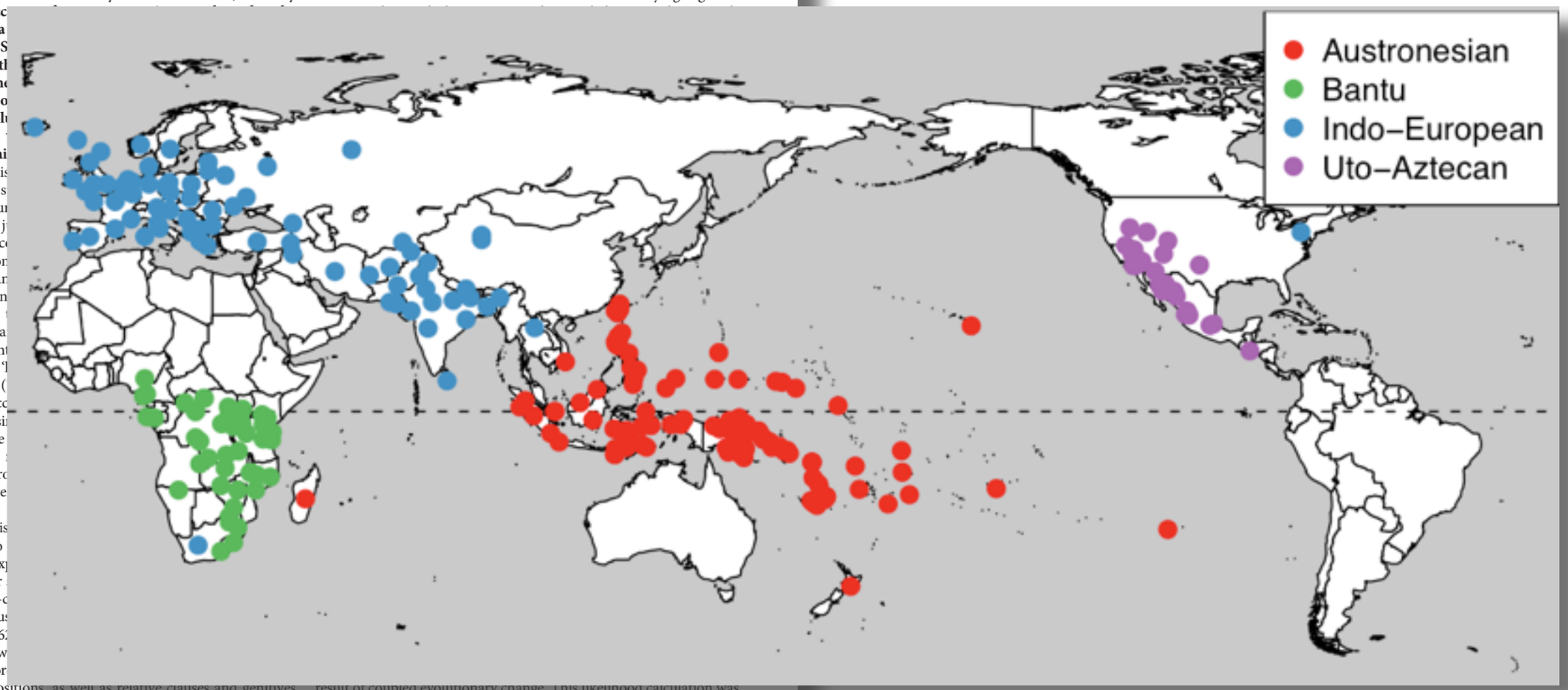
after the noun, whereas dominant object–verb ordering predicts post-positions, relative clauses and genitives before the noun[4]. One general explanation for these observations is that languages tend to be consistent ('harmonic') in their order of the most important element or 'head' of a phrase relative to its 'complement' or 'modifier'[3], and so if the verb is first before its object, the adposition (here preposition) precedes the noun, while if the verb is last after its object, the adposition follows the noun (a 'postposition'). Other functionally motivated explanations emphasize consistent direction of branching within the syntactic structure of a sentence[9] or information structure and processing efficiency[5].

To demonstrate that these correlations reflect underlying cognitive

result of coupled evolutionary change. This likelihood calculation was

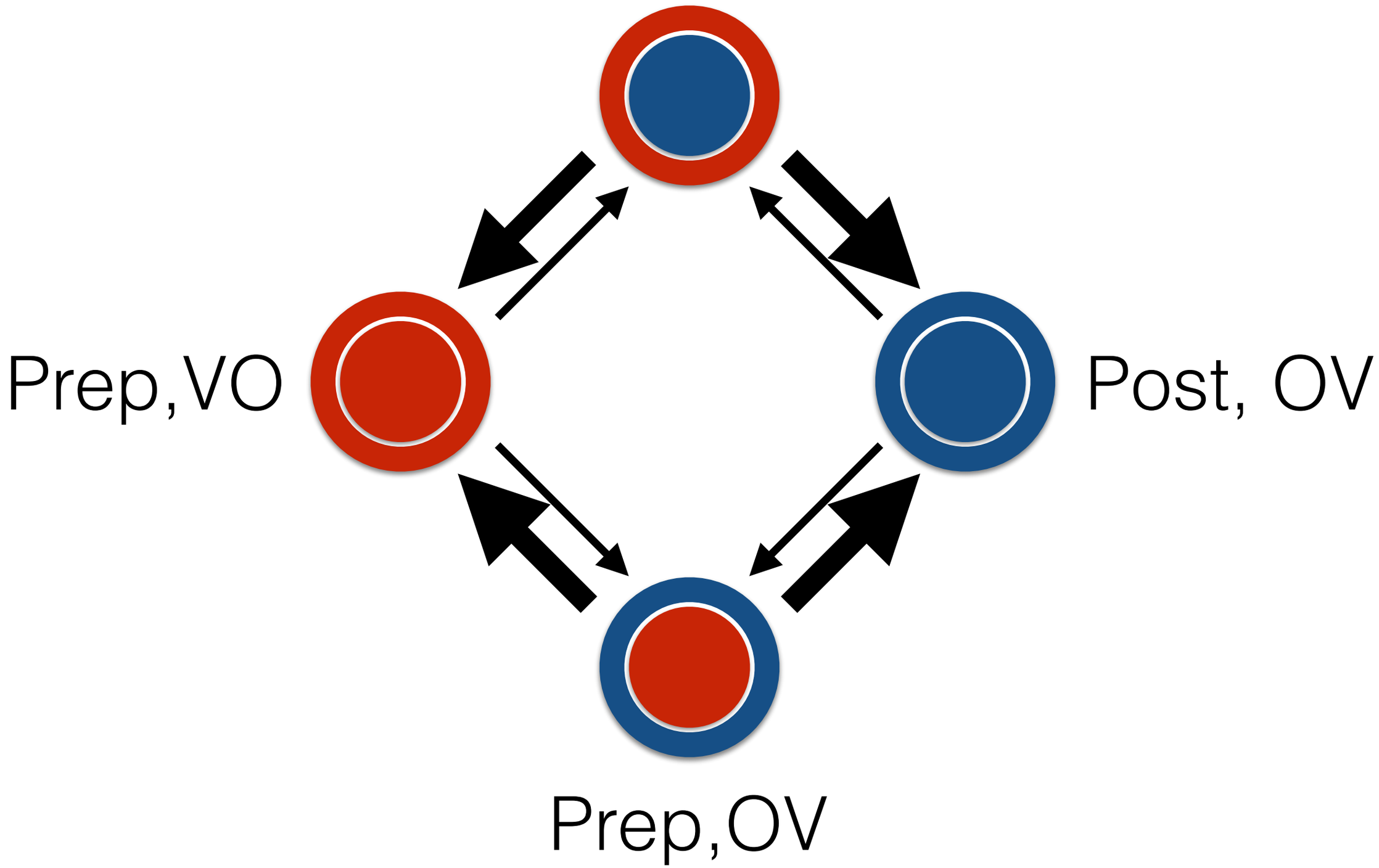
[1]Max Planck Institute for Psycholinguistics, Post Office Box 310, 6500 AH Nijmegen, The Netherlands. [2]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands. [3]Department of Psychology, University of Auckland, Auckland 1142, New Zealand. [4]Computational Evolution Group, University of Auckland, Auckland 1142, New Zealand.

©2011 Macmillan Publishers Limited. All rights reserved



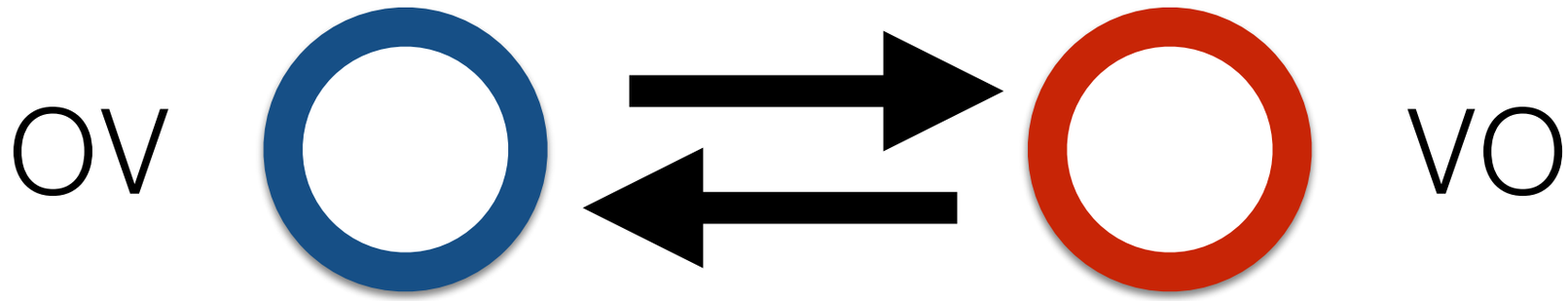
301 language sample,
4 families

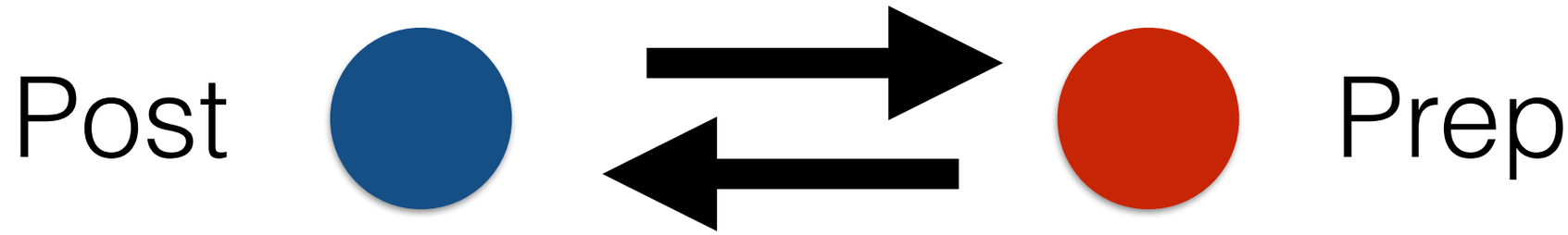
Dependent model



Pagel, Mark. 1994. “Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters.” Proceedings of the Royal Society of London B: Biological Sciences 255(1342): 37–45.

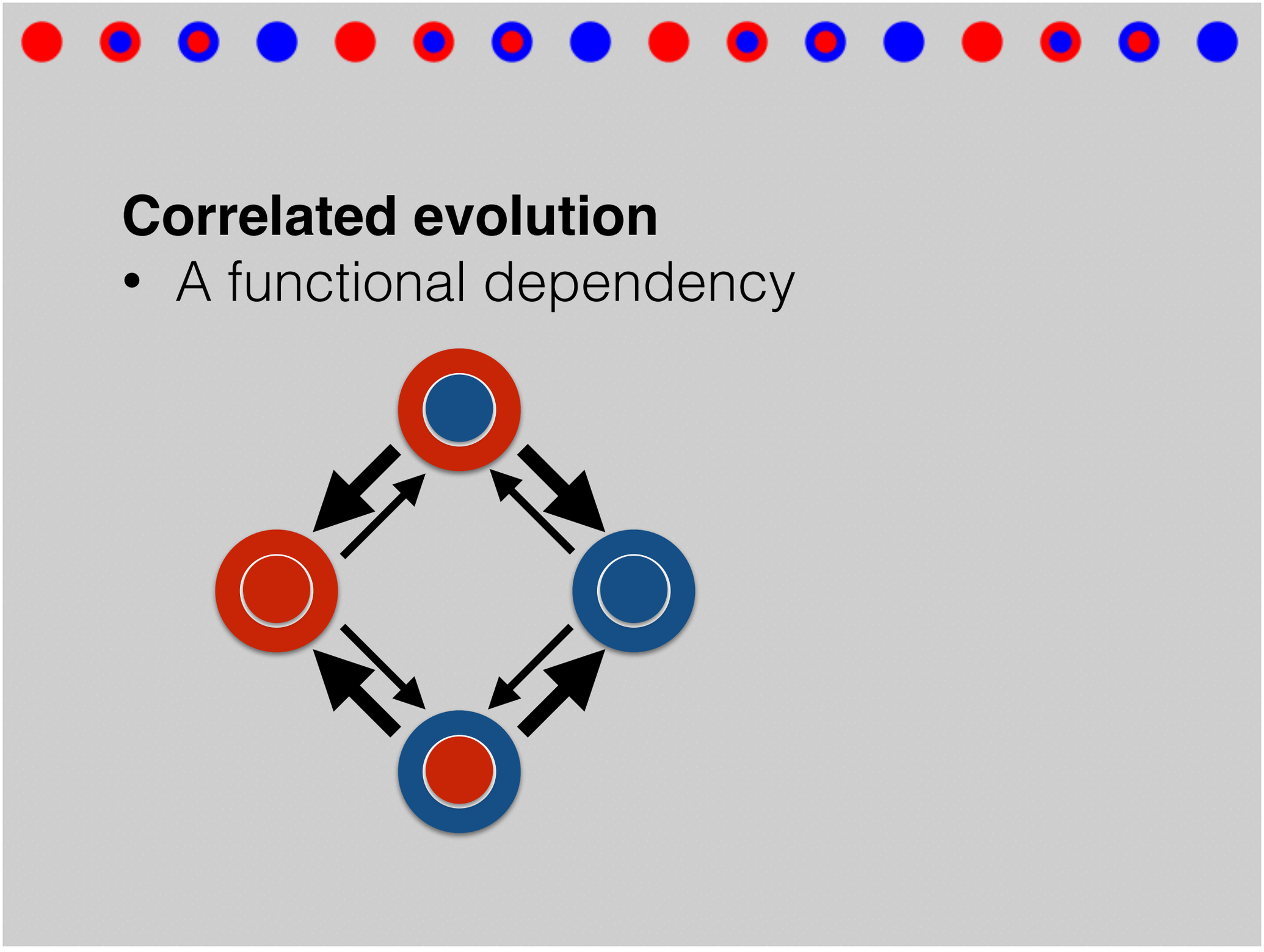
Independent model

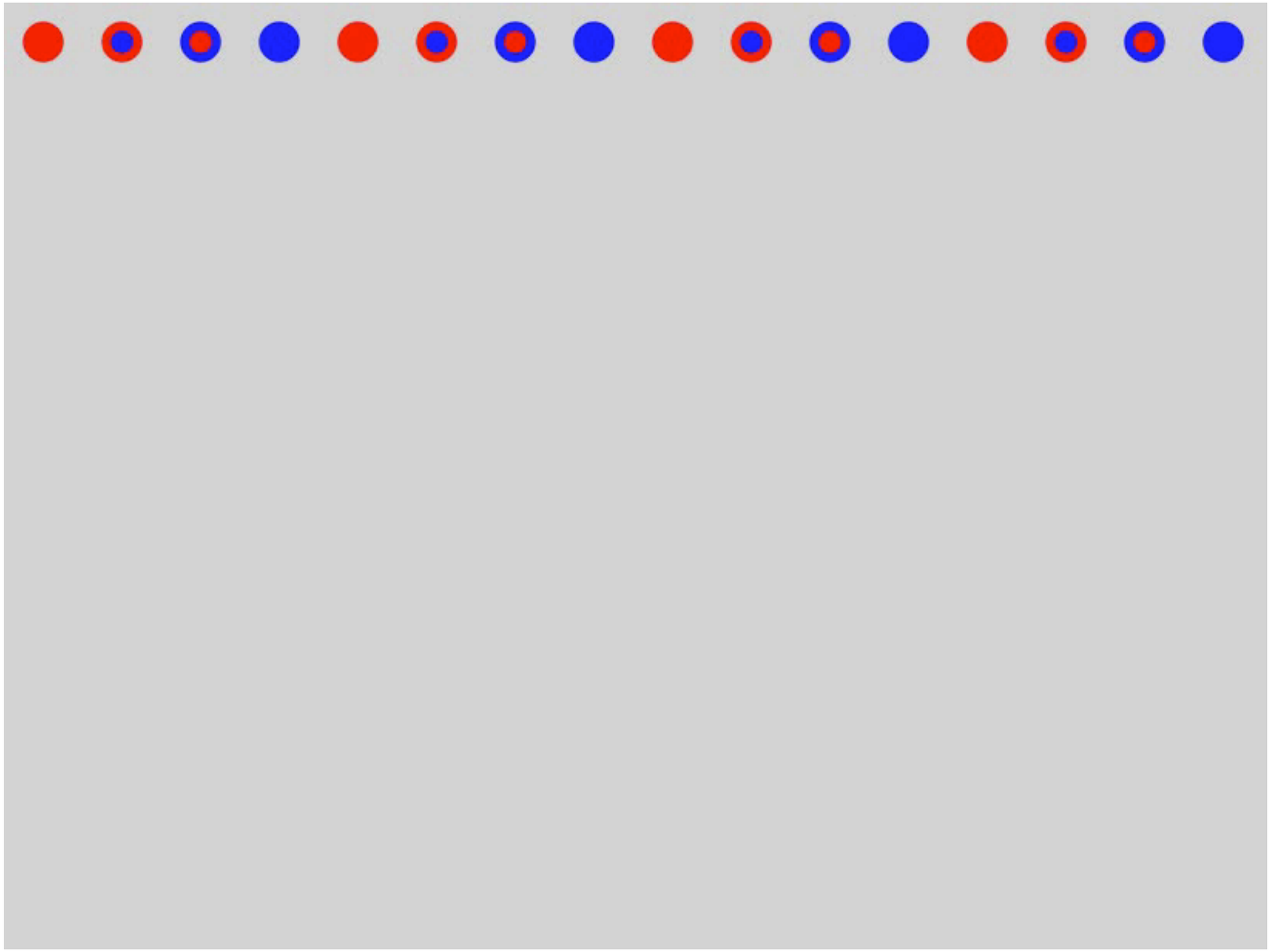


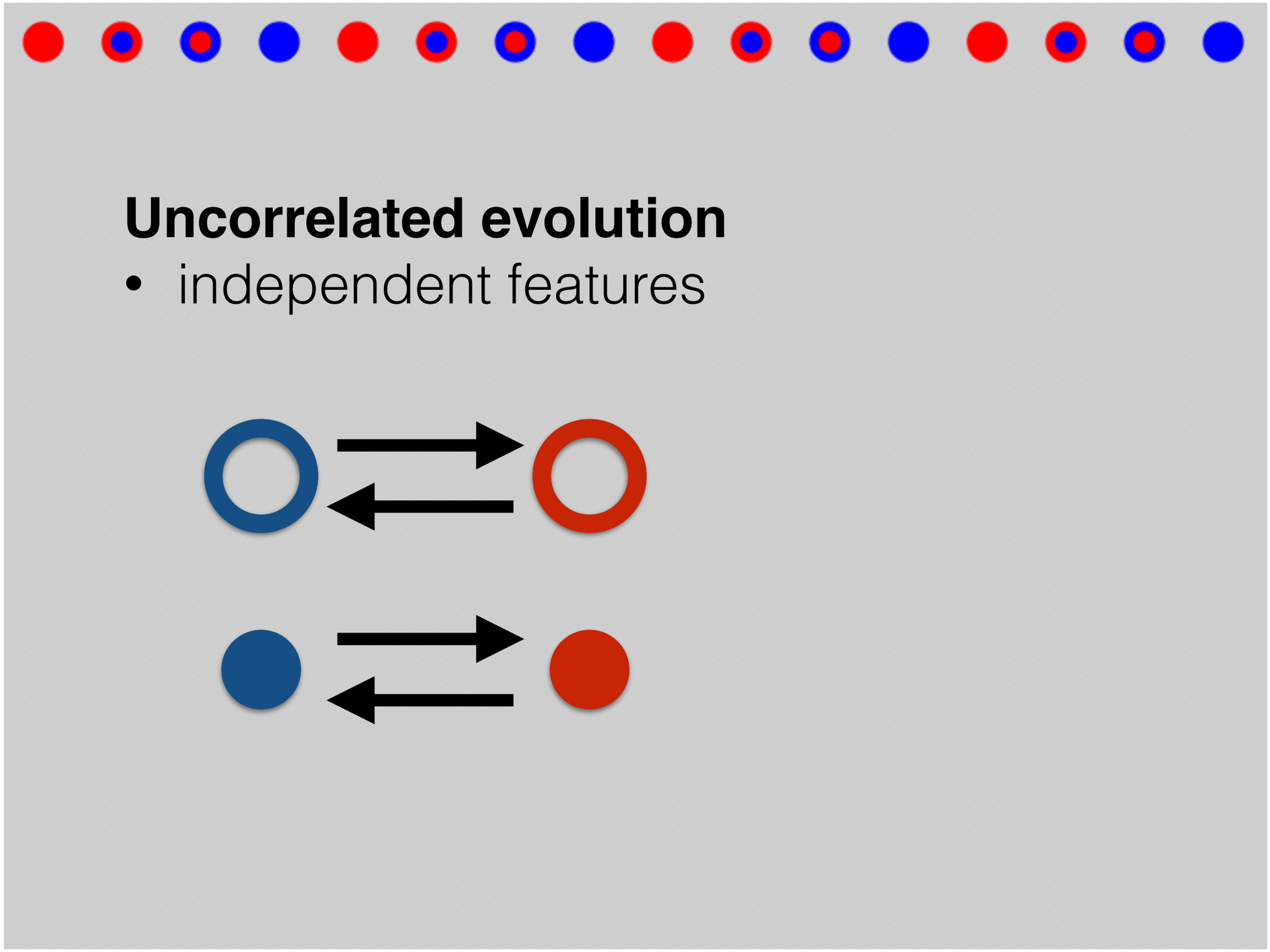


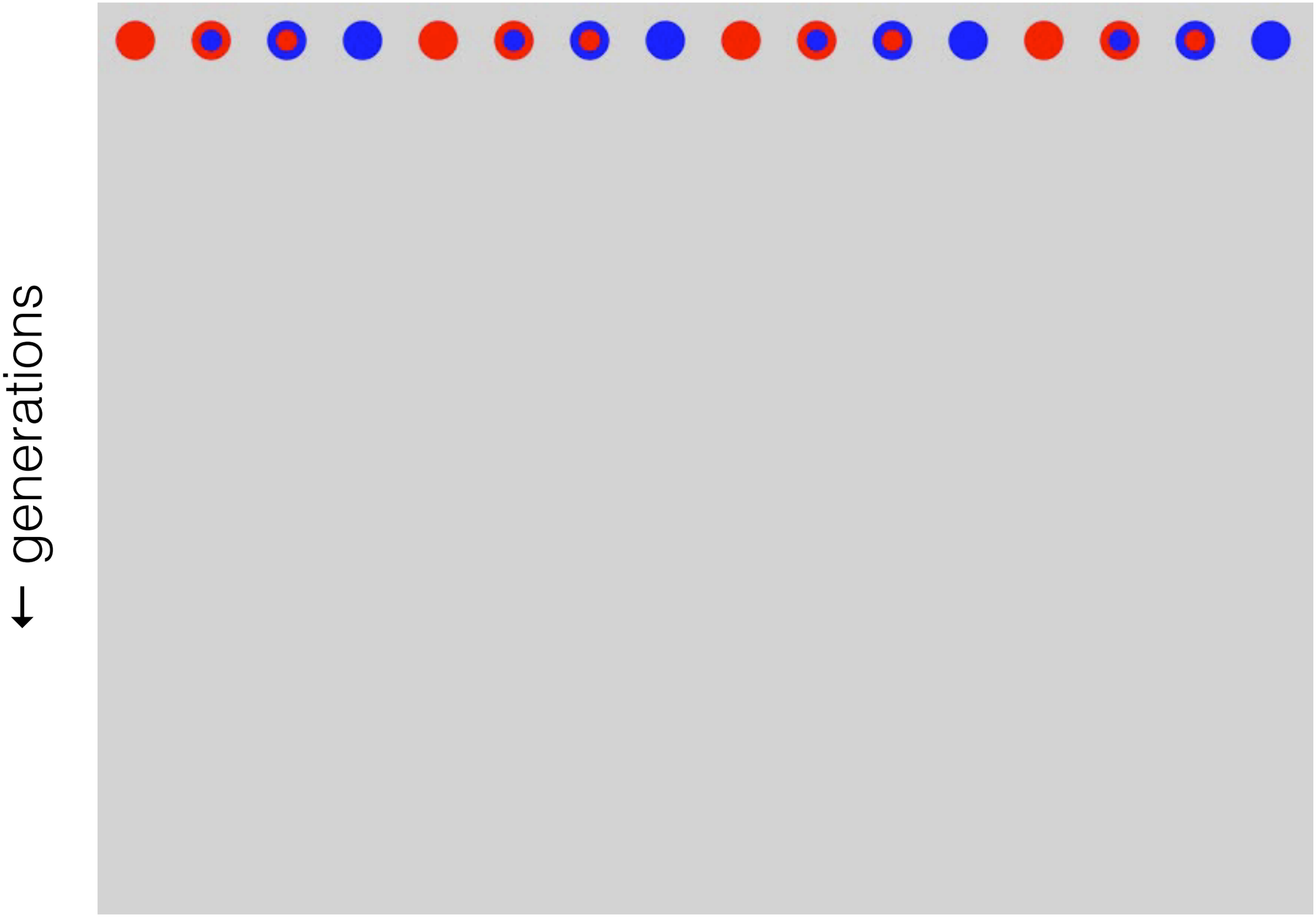
Pagel, Mark. 1994. “Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters.” Proceedings of the Royal Society of London B: Biological Sciences 255(1342): 37–45.

↓ generations




↓ generations

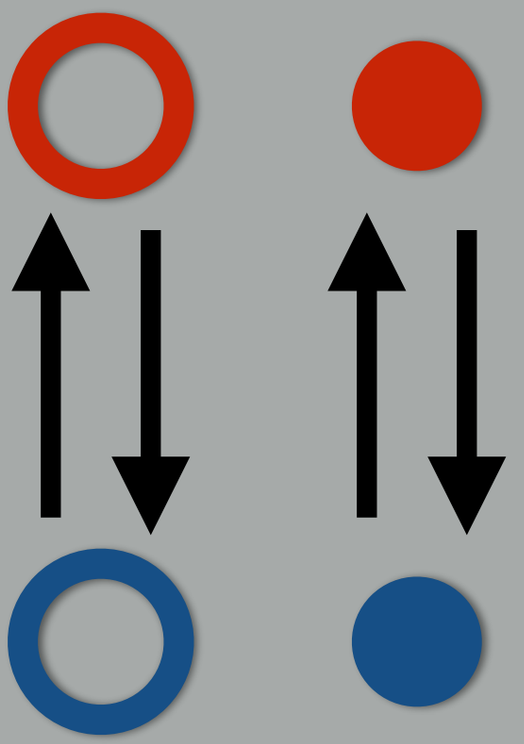

Uncorrelated evolution
• independent features
↓ generations


↓ generations

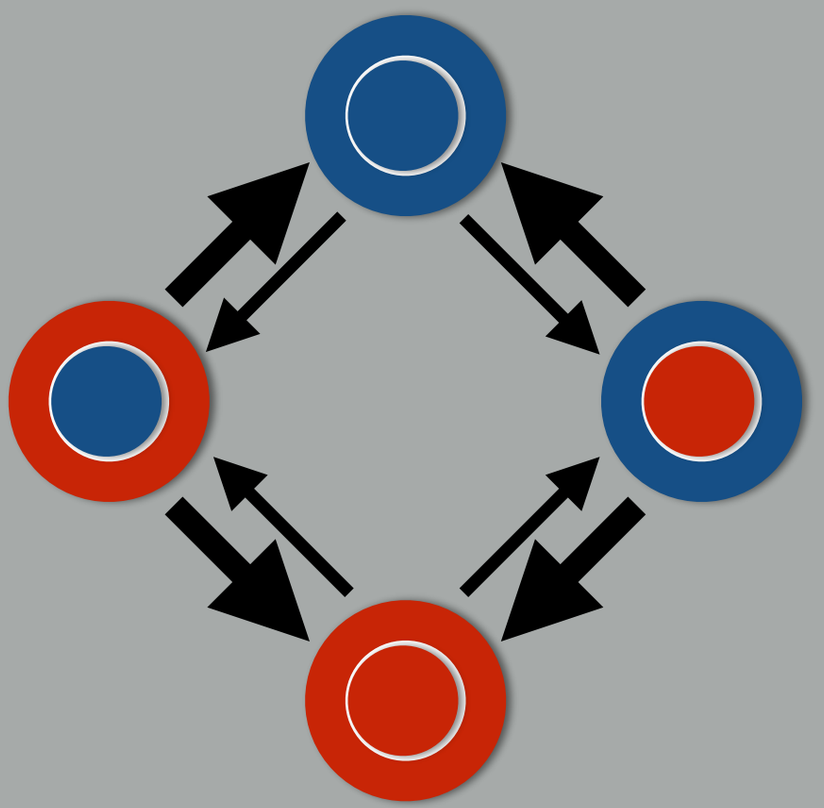
PostP, OV

PreP, VO

... and some of independent processes



Some patterns are the result of dependent processes...



Indo-European

Hittite

Waziri

Persian List

Tadzik

Baluchi

Ossetic

Bantu

Uto-Aztecan

Opata

Mekeo

Gapapaiwa

Kilivila

Saliba

Bwaidoga

Dobuan

Maututu

Bileki D

Sengseng

Mouk

Aria

Teanu

Buma

Kuanua

Kokota

Sisingga

Babatana

Banoni

Teop

Dehu

SakaoPortOlry

AmbaiYapen

Semantic typology

Semantics beyond the word

Semantic typology "the systematic cross-linguistic study of how languages express meaning by way of signs" (Evans 2010: 504)

Lexical typology "characteristic ways in which language [...] packages semantic material into words" (Lehrer 1992:249)

Lexical, Grammatical and Prosodic sign

- Big questions in semantics, systems rather than individual terms, e.g.

- Which subsystems encode which kinds of meanings?
  - Our ideas are most developed for morphological inflection (many publications on different, and new, grammatical categories)

  - Most poorly developed is prosody (What kind of meanings can be expressed by intonations? What are the cross-linguistic patterns?)

  - Some areas of the lexicon are explored (kinship, colour, ethnobiology), other areas neglected (facial types), other areas just beginning (smell, temperature)

Cross-linguistic comparison of meaning

- Relativist position stresses “incommensurability of different conceptual traditions, and the unsatisfactory nature of translation across languages”
  - “a meaning of a sign is its place in the system”
- The problem of representing meaning
  - logic-based metalanguage (e.g. studies of quantifiers)
  - diagrams (cognitive semantics)
  - non-linguistics reference (colour chips, species identifiers)

Cross-linguistic comparison of meaning

Etic vs. Emic

… seeks to set out all logically distinguishable possibilities regardless of whether individual languages group them together

… seeks to characterise what is common to all members of a category from the perspective of another language

Think "phonetic" vs. "phonemic"

Etic Grid

	♂ referent		♀ referent		
	♀ speaker	♀ speaker	♀ speaker	♀ speaker	
(elder)	1	3	5	7	Maximal
(younger)	2	4	6	8	
(elder)	brother		sister		English
(younger)					
(elder)	kakak				Indonesian
(younger)	adik				
(elder)	ani		otōto		Japanese
(younger)	ane		imōto		

Compare how the logically possible types are divided up in different languages

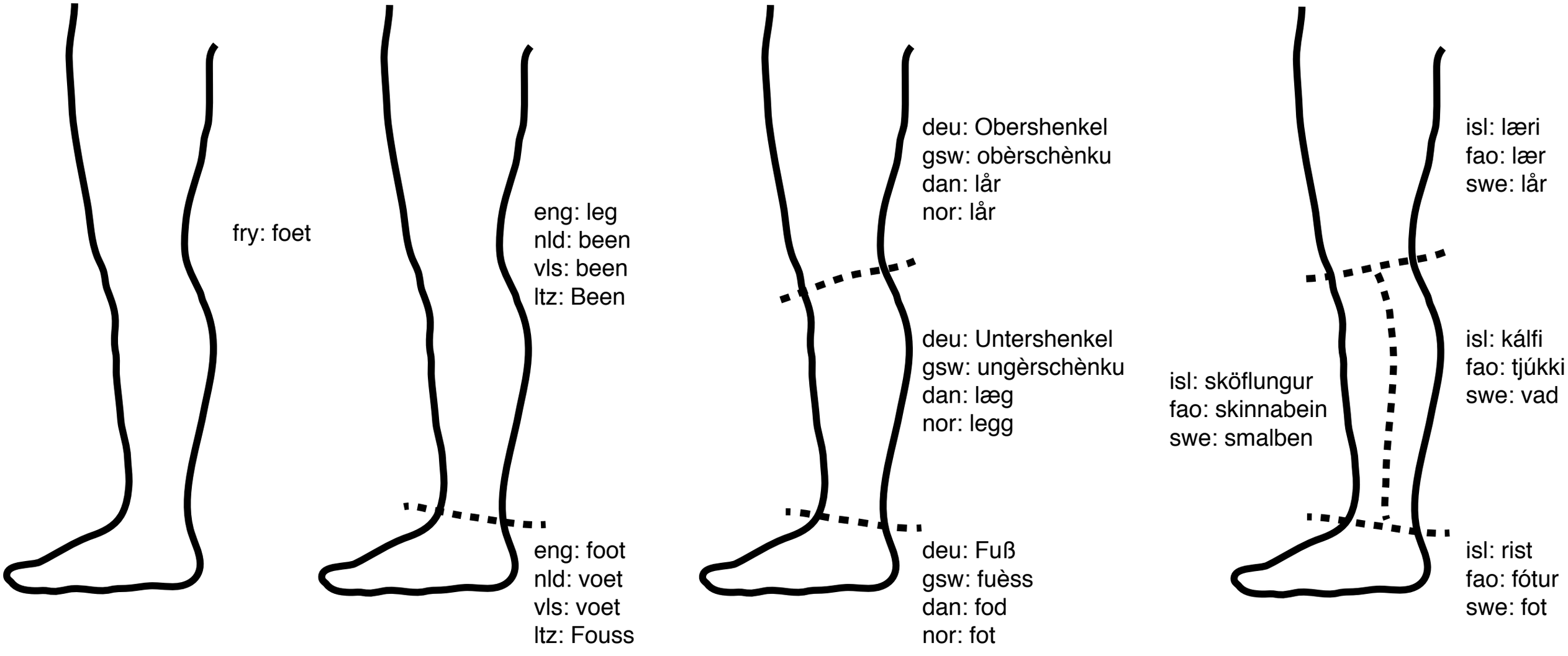
Beyond the etic grid

Kayardild sibling terms (etic)

	♂ referent		♀ referent	
	♂ speaker	♀ speaker	♂ speaker	♀ speaker
(elder)	thabuju	kularrind	kularrind	yakukathu
(younger)	duujind	kularrind	kularrind	duujind

Determine the language-internal (emic) logic of the system

Granularity



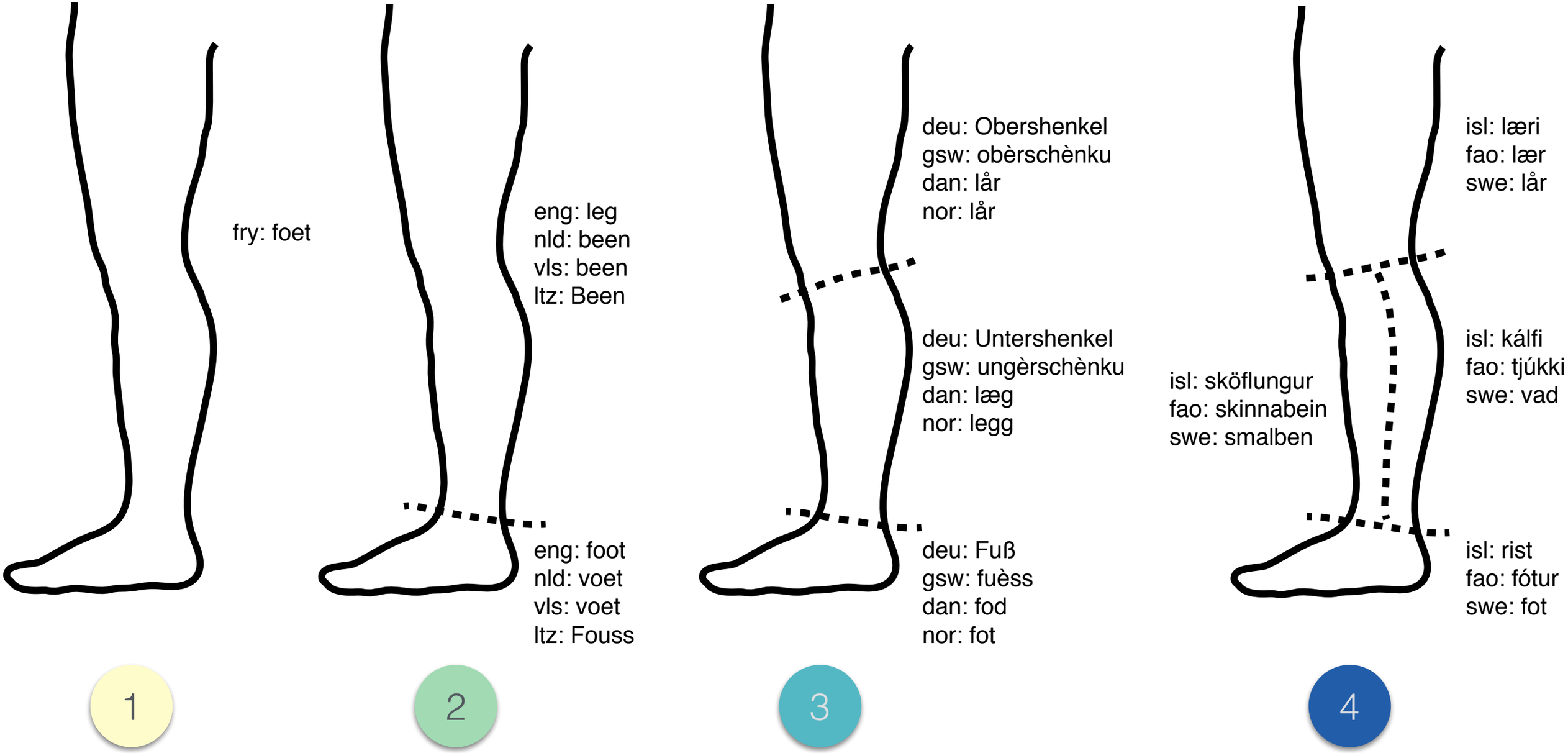
Grouping and dissection

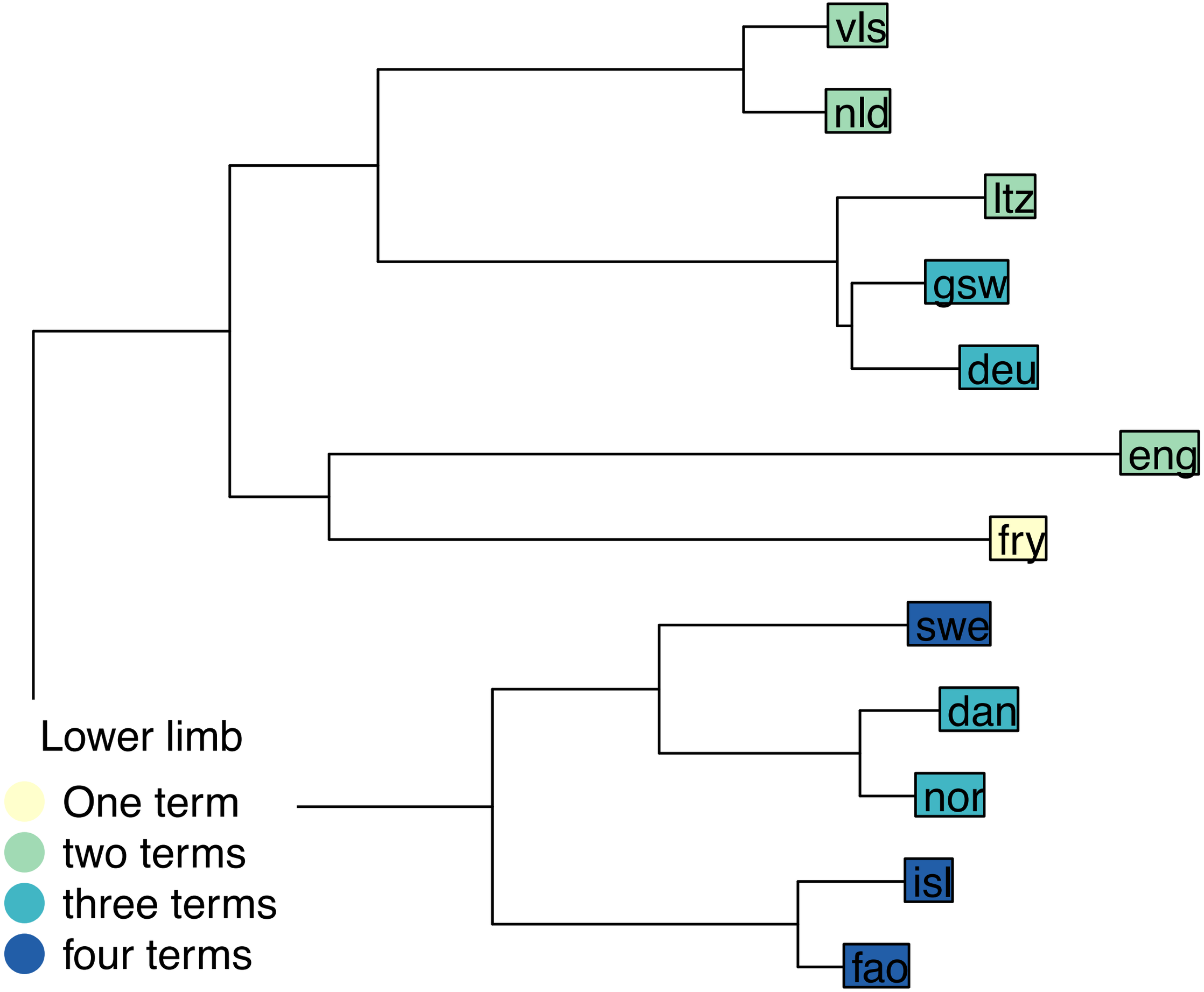
English	*hand*	*finger*	*foot*	*toe*
Russian	*ruka*	***palets***	*noga*	***palets***

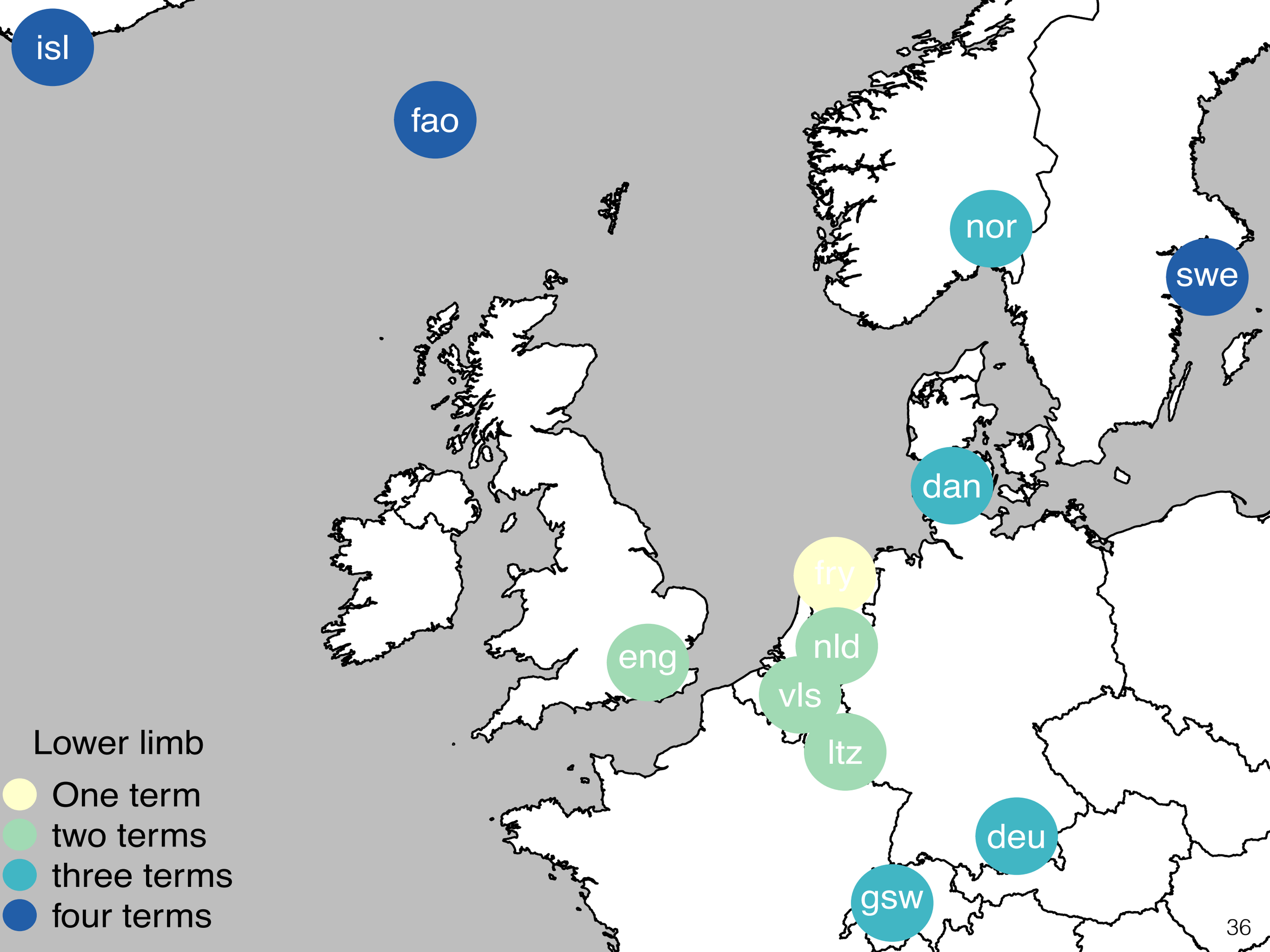
Phylogenetics

+ Semantics

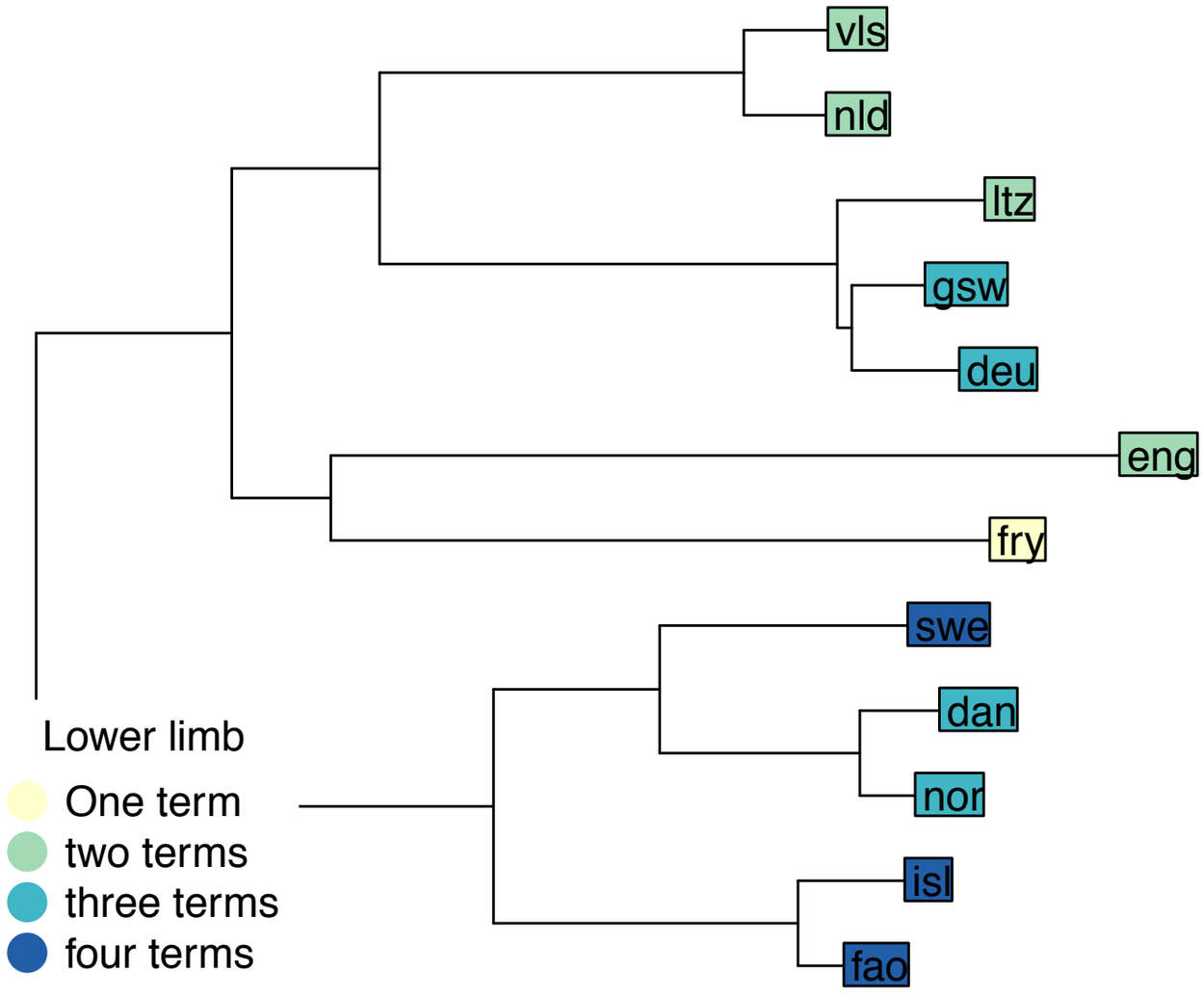
Putting it all together

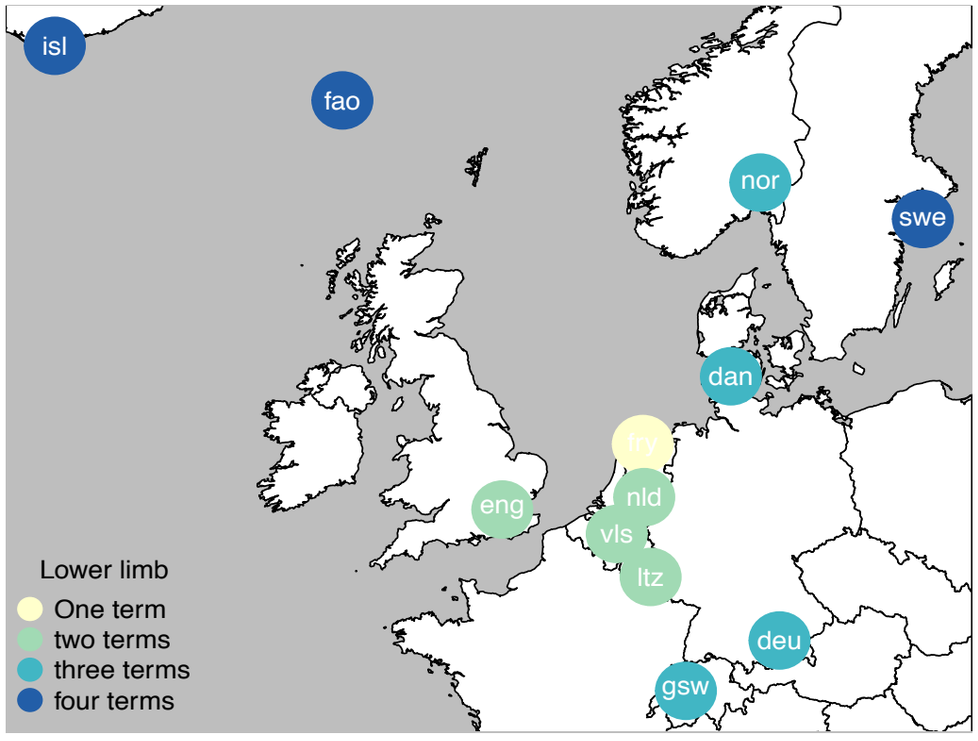

fry: foet
eng: leg
nld: been
vls: been
ltz: Been
eng: foot
nld: voet
vls: voet
ltz: Fouss
deu: Obershenkel
gsw: obèrschènku
dan: lår
nor: lår
deu: Untershenkel
gsw: ungèrschènku
dan: læg
nor: legg
deu: Fuß
gsw: fuèss
dan: fod
nor: fot
isl: læri
fao: lær
swe: lår
isl: sköflungur
fao: skinnabein
swe: smalben
isl: kálfi
fao: tjúkki
swe: vad
isl: rist
fao: fótur
swe: fot
1
2
3
4


vls
nld
ltz
gsw
deu
eng
fry
swe
dan
nor
isl
fao
Lower limb
One term
two terms
three terms
four terms


isl
fao
nor
swe
dan
fry
nld
eng
vls
ltz
deu
gsw
Lower limb
One term
two terms
three terms
four terms

What determines this diversity?
A mixture of inheritance and contact





Emic variation
Etic comparison

- How can we compare how two (many) languages segment a semantic domain?

- Qualitatively

- Quantitatively

  - so we can attempt to disentangle which factors determine the of patterns variation

yes

no

yes

English
‘leg’?

https://commons.wikimedia.org/wiki/File:HBGM.jpg

da

da!

da

English
‘leg’?

Russian
‘noga’?

https://commons.wikimedia.org/wiki/File:HBGM.jpg

yes da

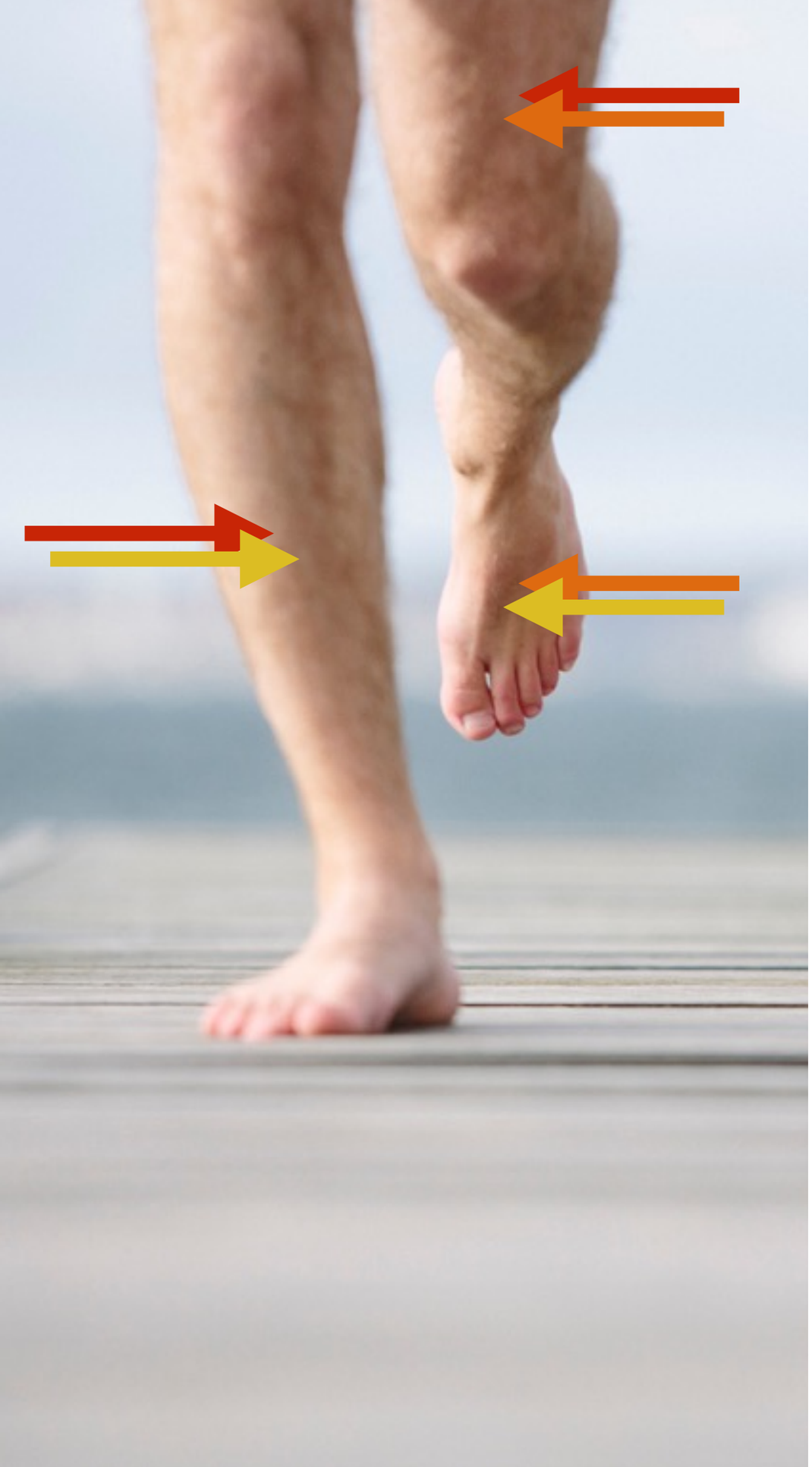
no da

yes da

English
‘leg’?

Russian
‘noga’?

https://commons.wikimedia.org/wiki/File:HBGM.jpg



Do these two parts have the same label?

English	Russian	same?
yes	**yes**	**1**
no	**yes**	**0**
no	**yes**	**0**
		33%

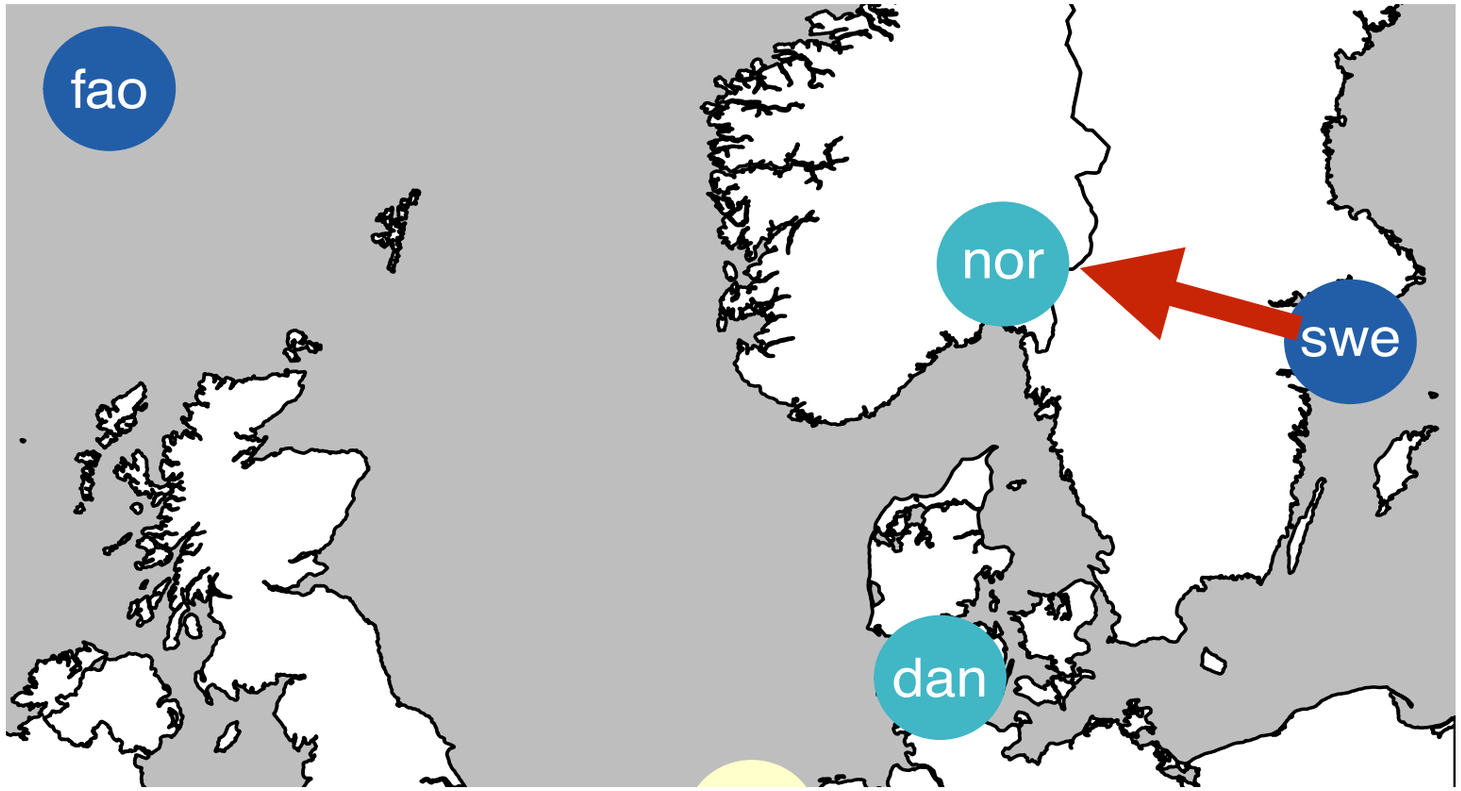
- ‘Pairwise similarity’
- Compares lexical classification without paying attention to lexical form

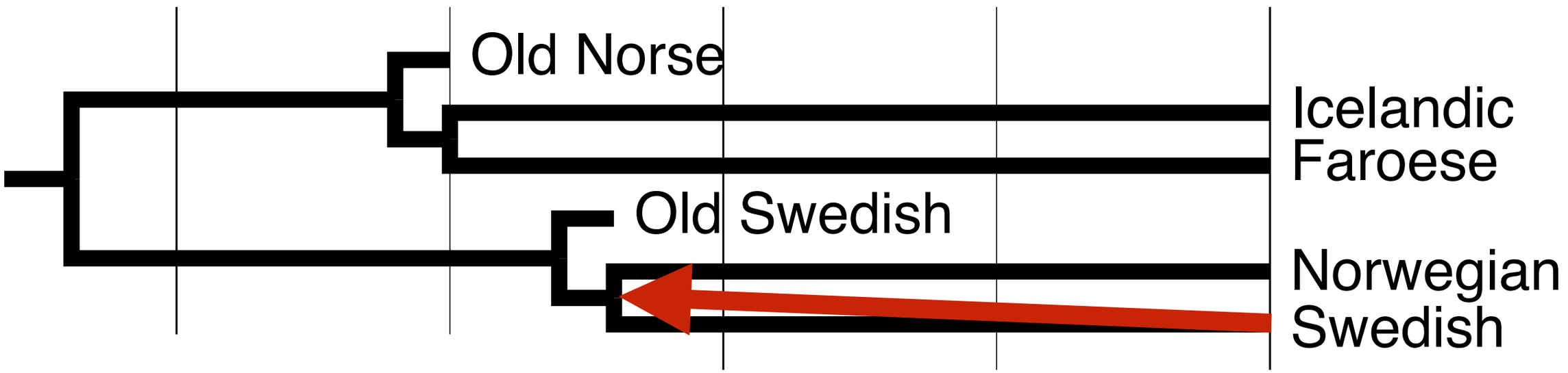
Distances

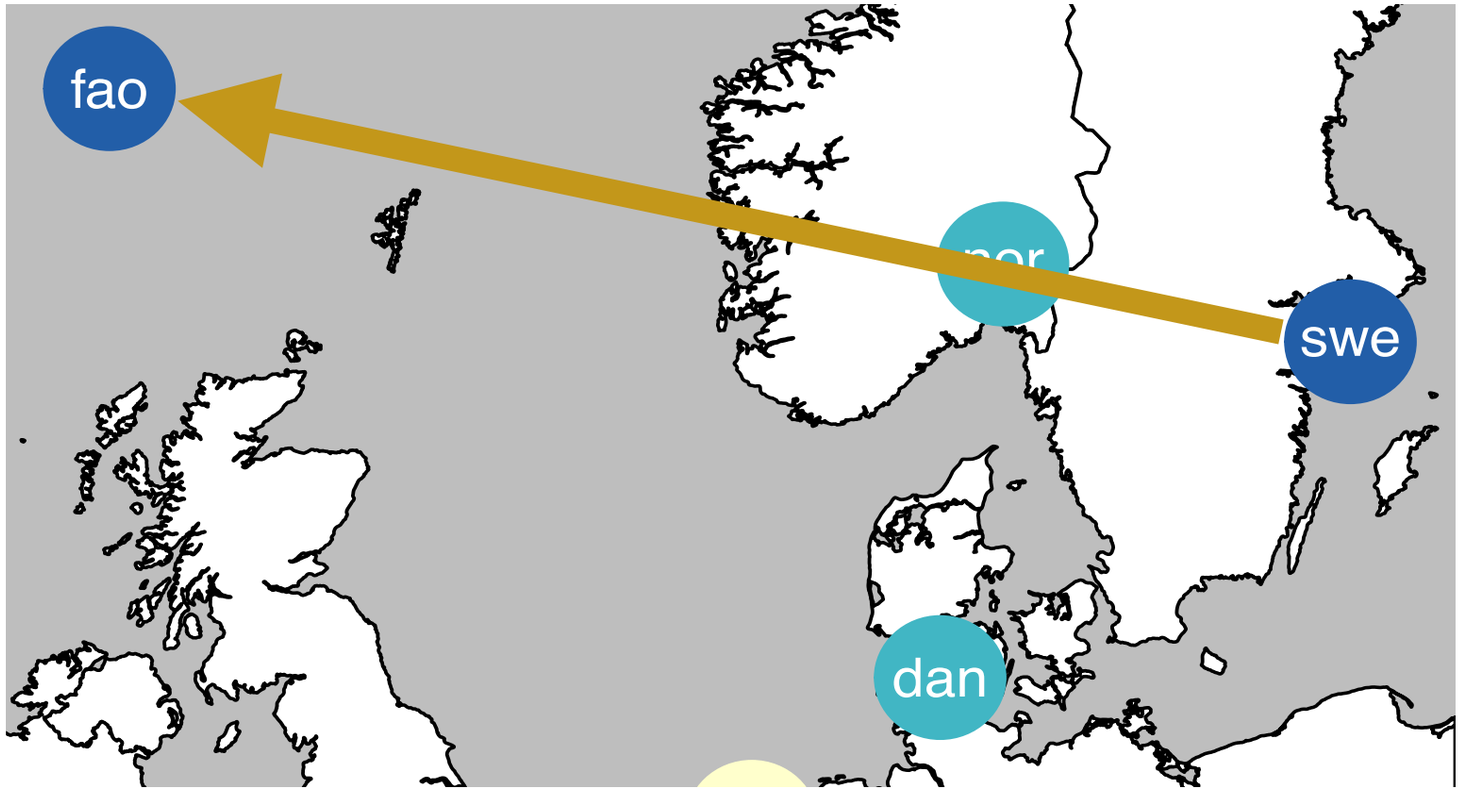
- Pairwise similarity → A semantic distance measure per domain

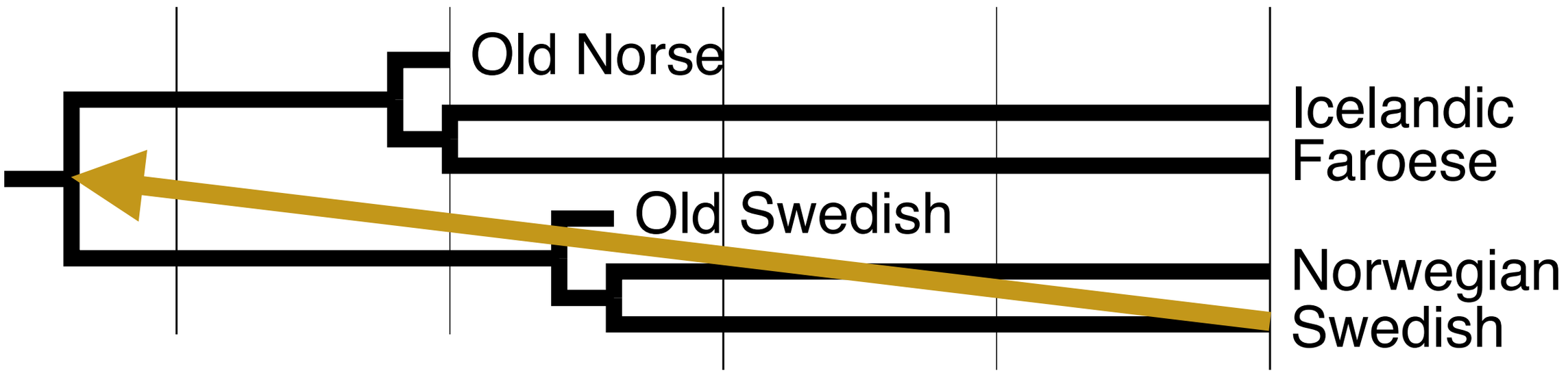
Two other kinds of distance:

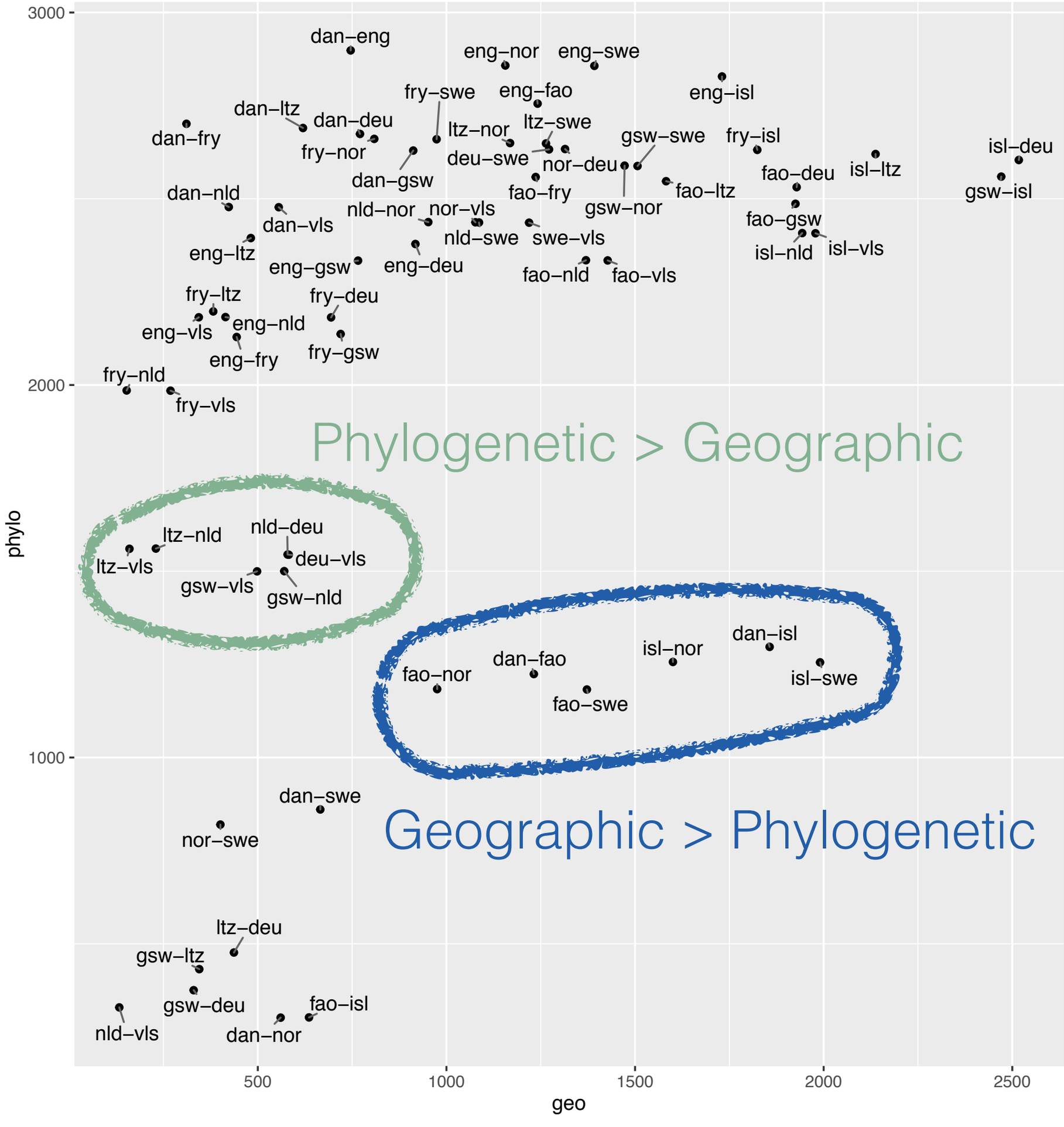
- Geographic distance
- Phylogenetic distance


fao
nor
swe
dan


Old Norse
Old Swedish
Icelandic
Faroese
Norwegian
Swedish


fao
nor
swe
dan

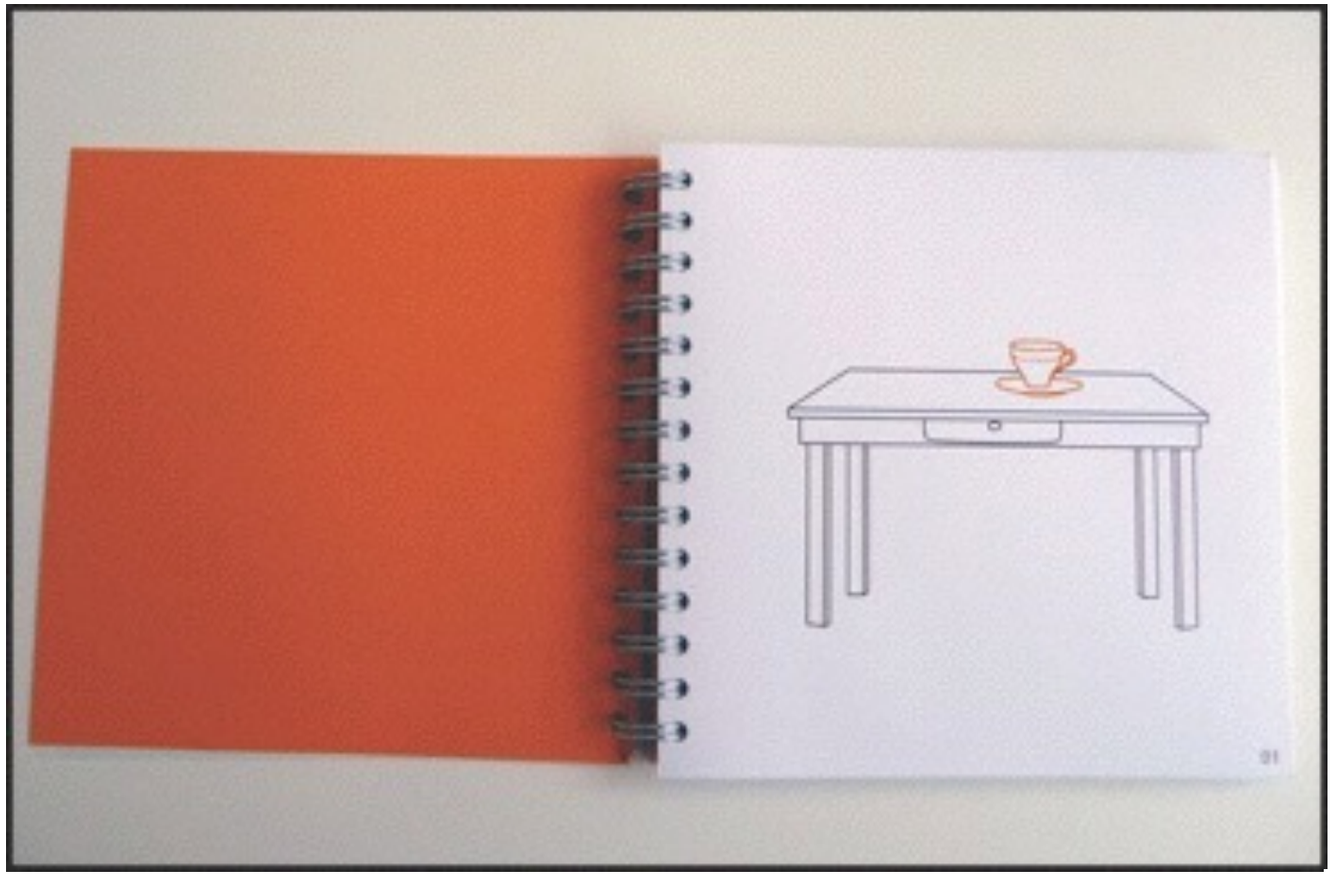

Old Norse
Old Swedish
Icelandic
Faroese
Norwegian
Swedish

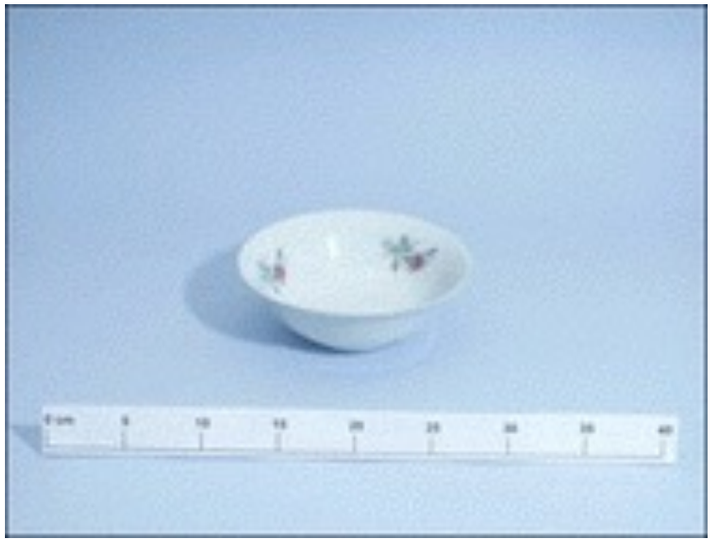

3000
2000
1000
phylo
dan–eng
eng–nor
eng–swe
eng–isl
fry–swe
eng–fao
dan–ltz
dan–deu
ltz–nor
ltz–swe
dan–fry
fry–nor
deu–swe
nor–deu
gsw–swe
fry–isl
isl–deu
isl–ltz
fao–deu
dan–gsw
fao–fry
fao–ltz
gsw–isl
dan–nld
nld–nor
nor–vls
gsw–nor
fao–gsw
dan–vls
nld–swe
swe–vls
eng–ltz
isl–nld
isl–vls
eng–gsw
eng–deu
fao–nld
fao–vls
fry–ltz
fry–deu
eng–nld
eng–vls
eng–fry
fry–gsw
fry–nld
fry–vls
Phylogenetic > Geographic
ltz–nld
nld–deu
ltz–vls
deu–vls
gsw–vls
gsw–nld
dan–isl
isl–nor
dan–fao
fao–nor
isl–swe
fao–swe
dan–swe
nor–swe
Geographic > Phylogenetic
ltz–deu
gsw–ltz
gsw–deu
fao–isl
nld–vls
dan–nor
500
1000
1500
2000
2500
geo

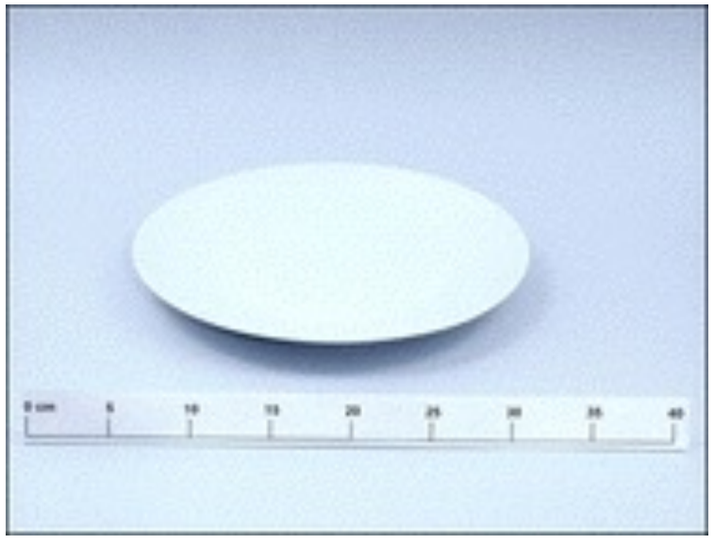
Data

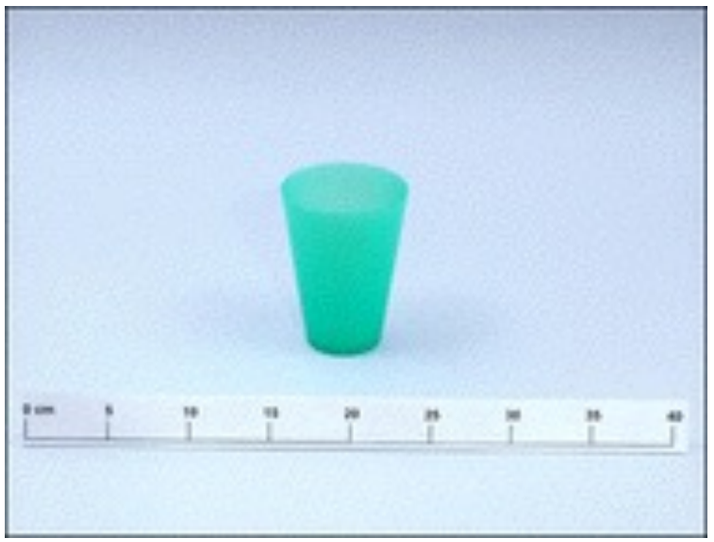
- For each of 50 Indo-European languages…
  - for each of four semantic domains…
    - for each of 20 participants…
      - elicit names of each stimulus item (~80 per domain)

http://eoss.nl

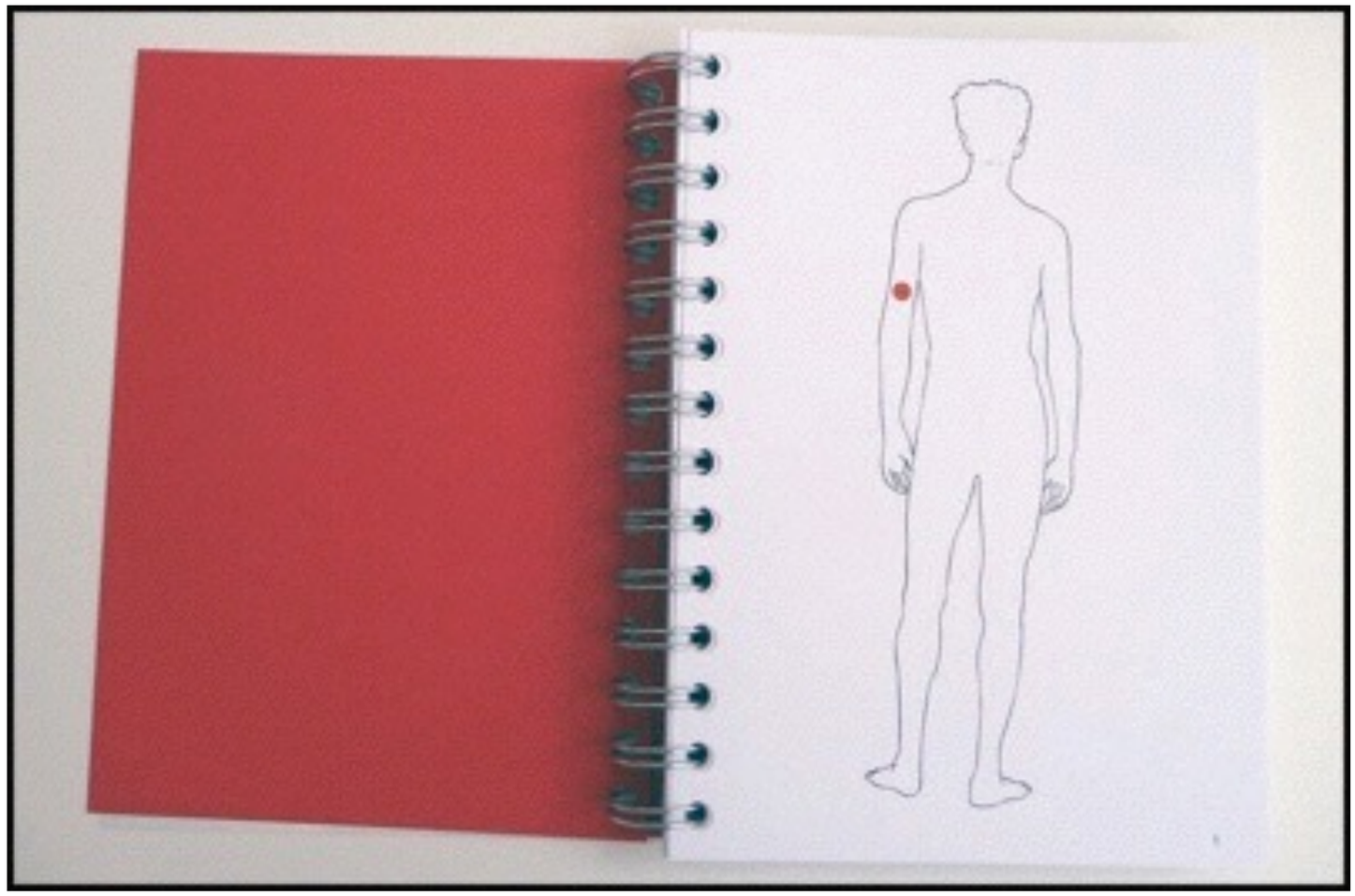


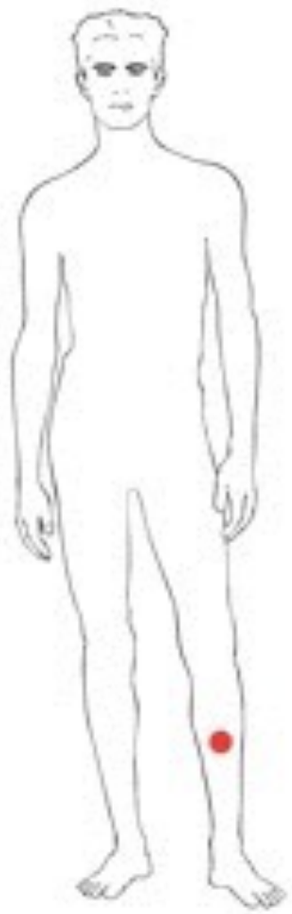






Ameel, Storms, Malt & Sloman. 2005. “How Bilinguals Solve the Naming Problem.” Journal of Memory and Language 53(1):60-80





Contact vs. Inheritance

- Test for
  - correlation between semantic distance and phylogenetic distance
  - correlation between semantic distance and geographic distance
- Correct for the dependency between phylogenetic and geographic distance

[unpublished data removed]

Conclusions

- *Semantic systems* are tractable for diversity studies
- Biology and ecology have tools that can address the questions we are interested in
- It is possible to make predictive models of semantic variation
- Some semantic domains are more susceptible to contact effects, others track genealogy