The dynamics of areas and universals

Balthasar Bickel
University of Zürich

Central question and plan

What kinds of linguistic data give most insight into contact effects and large-scale area formation?

- Explore this in a relatively well-established large area: *Eurasia*
- Take issue with traditional ideas of
  - "controlling for" genealogical relatedness in language families
  - putting research on areas in opposition to research on universals
- Propose a new approach for both

Eurasia as a linguistic area

- Jakobson 1931: эвразийский языковий сойуз
- Nichols 1992, 1998: the Eurasian spread zone
- **Predictive Areality Theory** (Bickel & Nichols 2006):

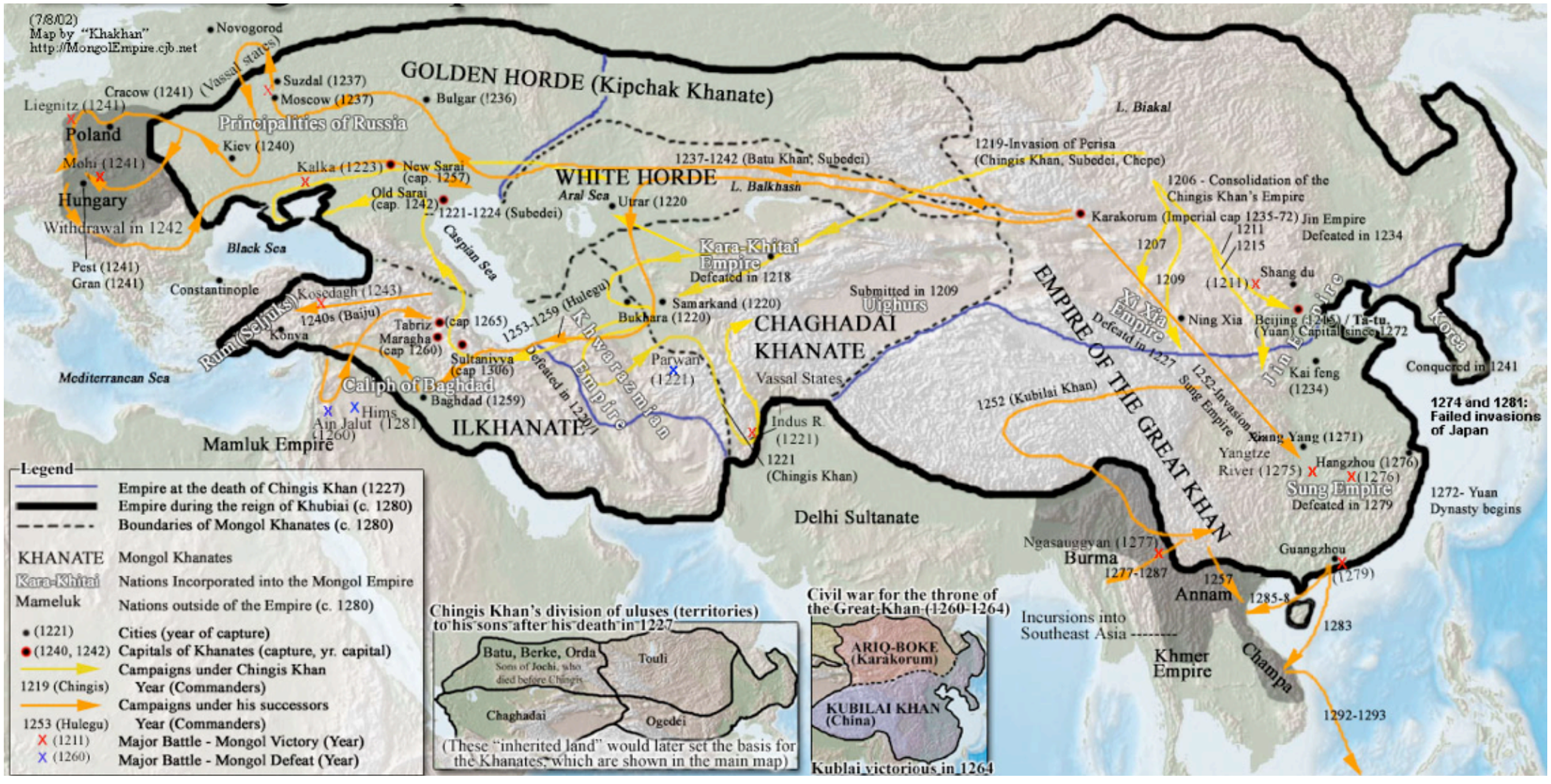
  in order to avoid circularity, linguistic areas cannot be identified by typological data but need to be grounded in **non-typological facts**

  - archeology, history
  - language family spreads and concomitant language shift and contact events
  - population genetics

Eurasia as a linguistic area: non-typological evidence

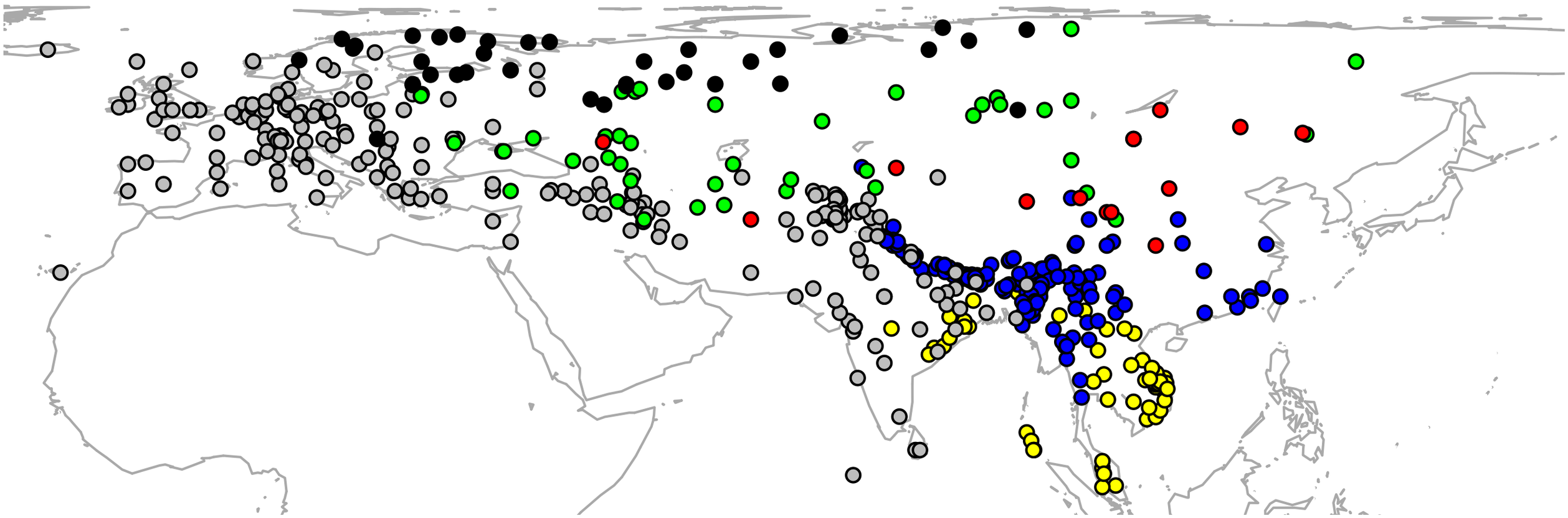
Historical record (Nichols 1998):

- mounted nomadism for about 4ky
- Iranian and later, Turko-Mongolic spreads



Eurasia as a linguistic area: non-typological evidence

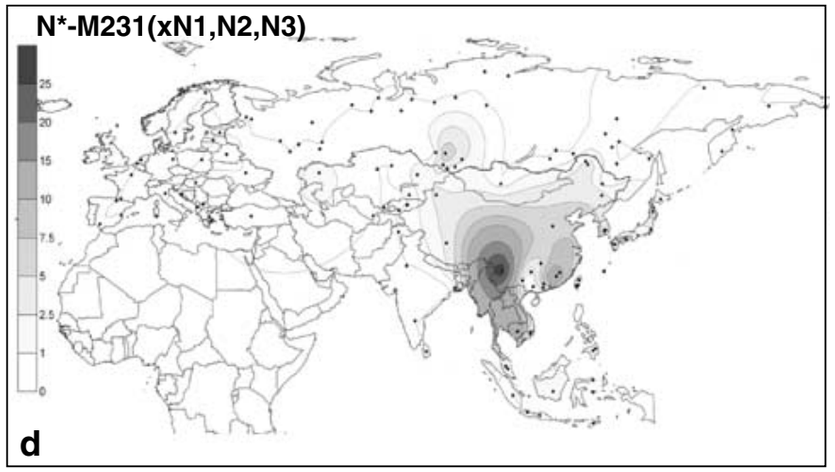
Spreads of major families (Nichols 1998):

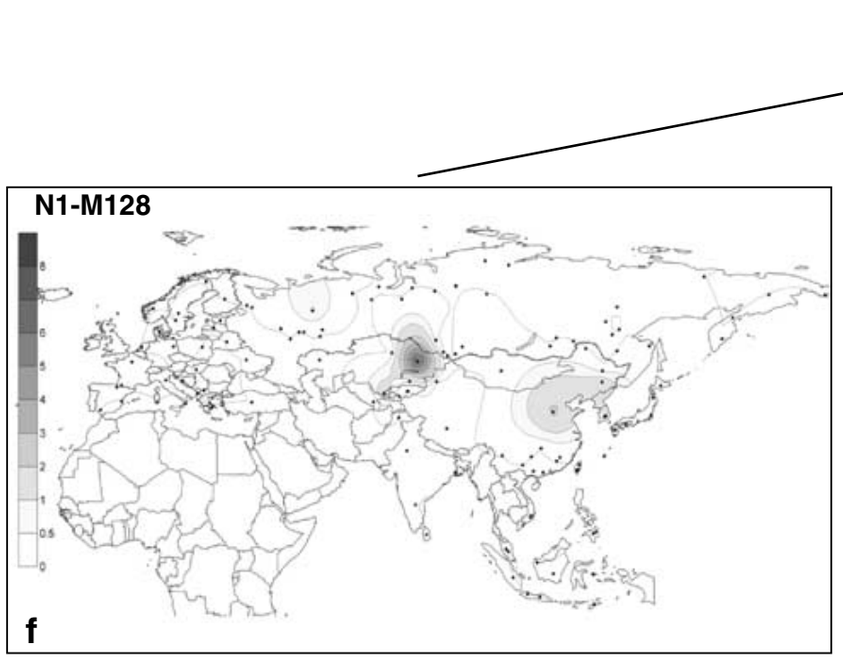


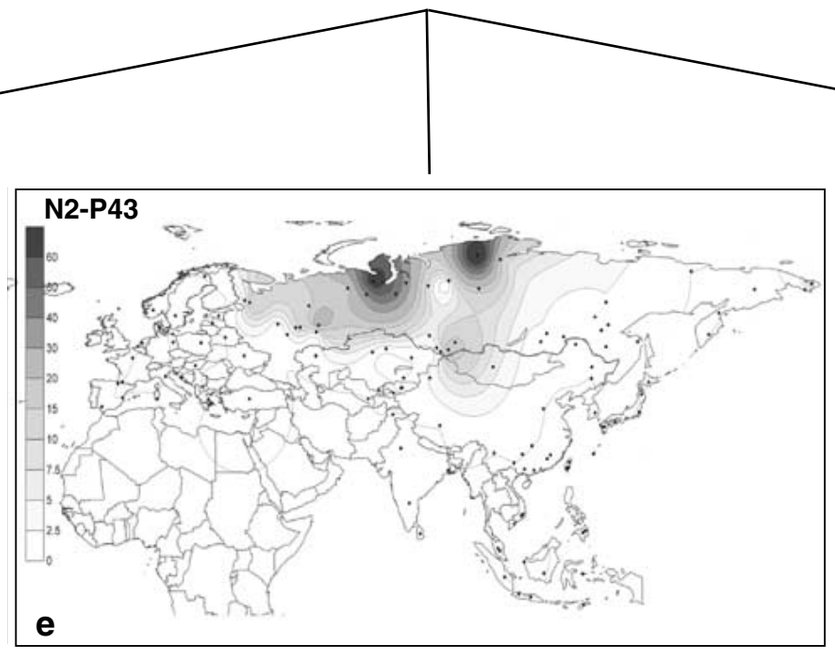
Eurasia as a linguistic area: non-typological evidence

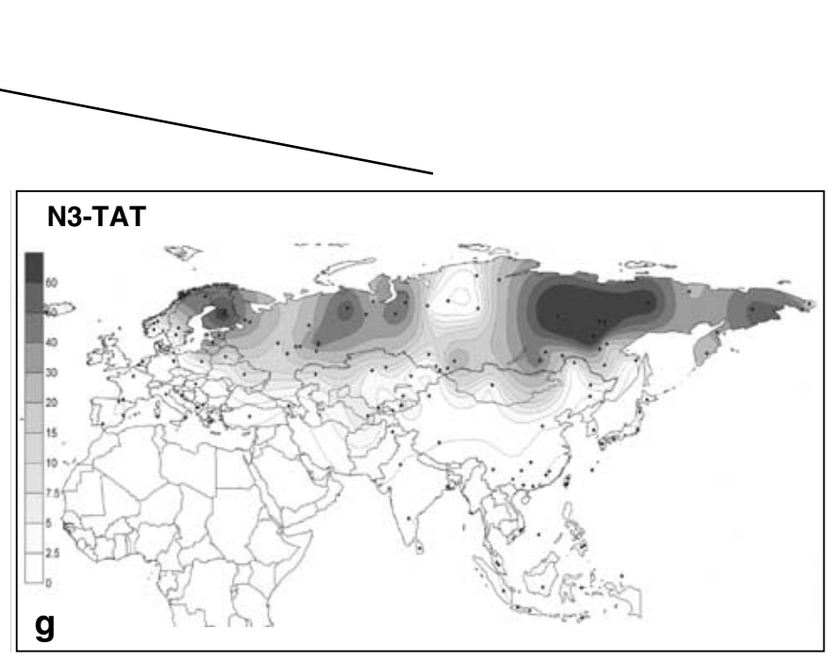
Population genetics (Rootsi et al. 2007): Y-chromosome haplogroup *N*

14 ±4 ky







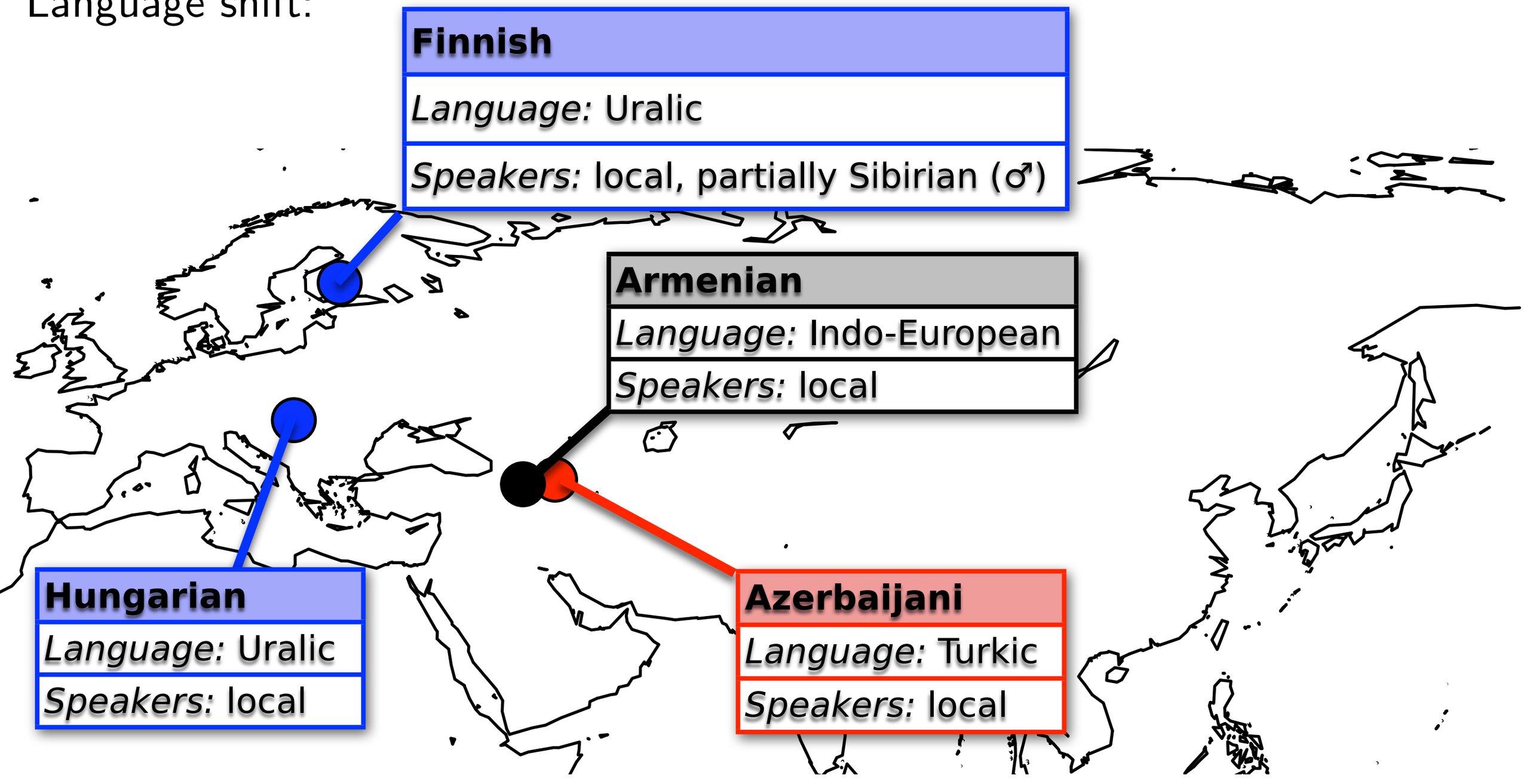


No comparable pattern in mtDNA

No evidence for N subclades in Native Americans

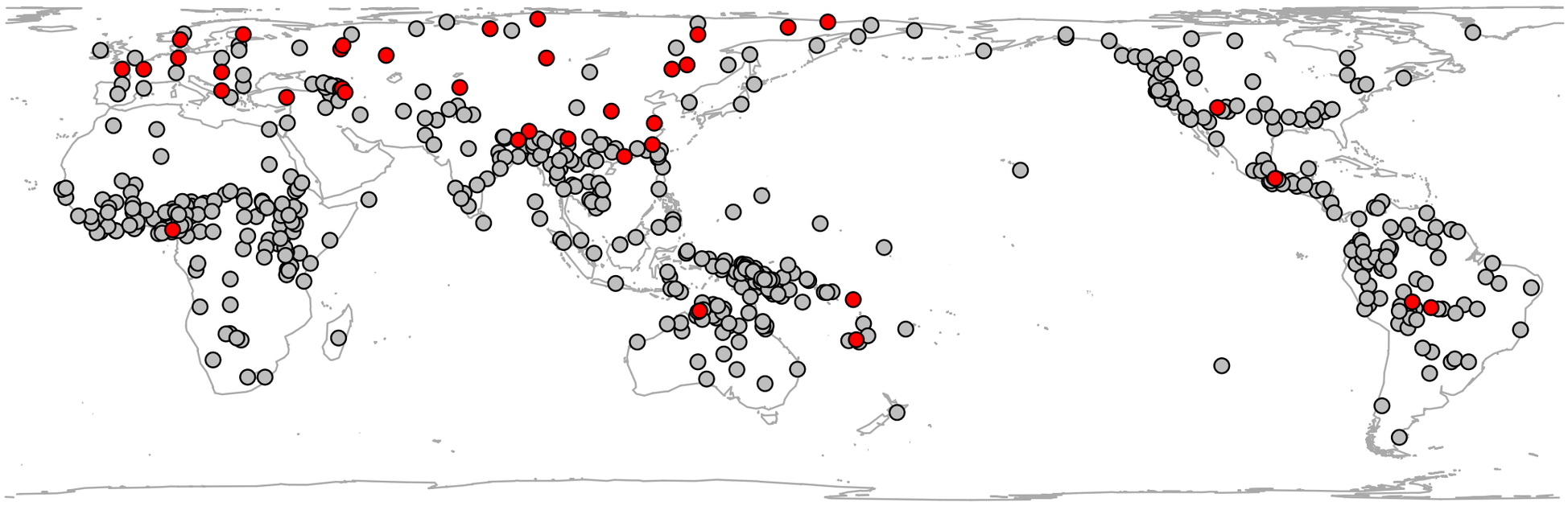
Eurasia as a linguistic area: non-typological evidence

Language shift:

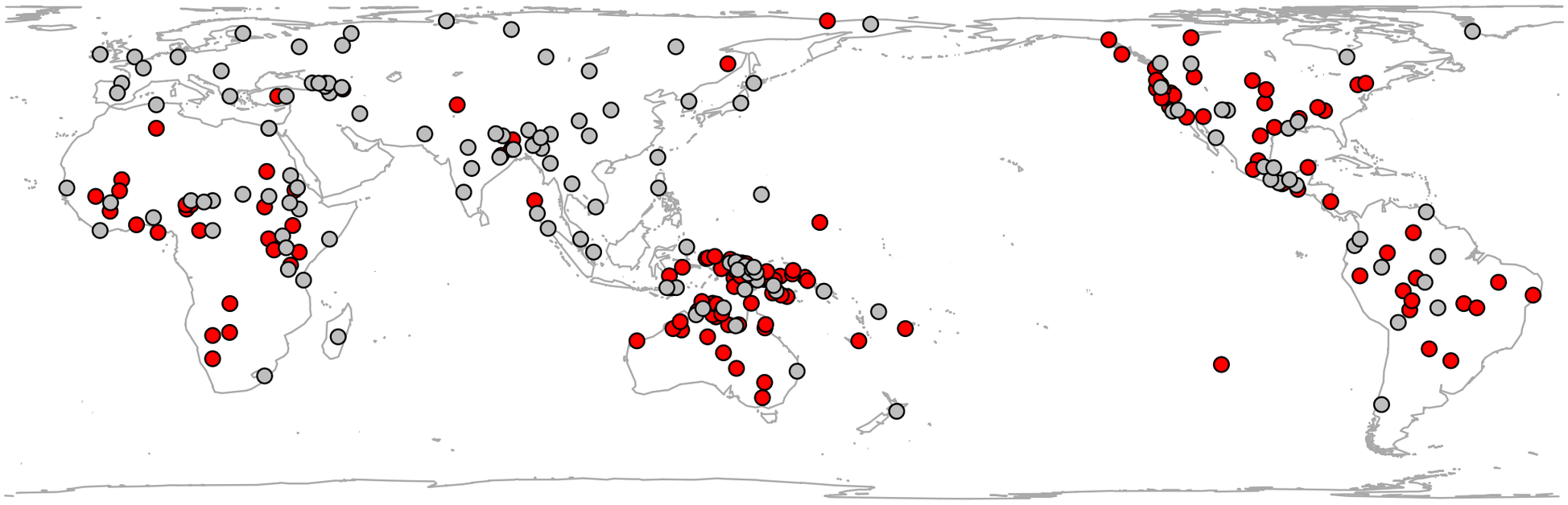


Kittles et al. 1998 *Am. J. Hum Gen.* 62; Semino et al. 2000 *Eur. J. Hum. Gen.* 8, Nasidze et al. 2003 *Hum Gen.* 112

Cartographic impressions...

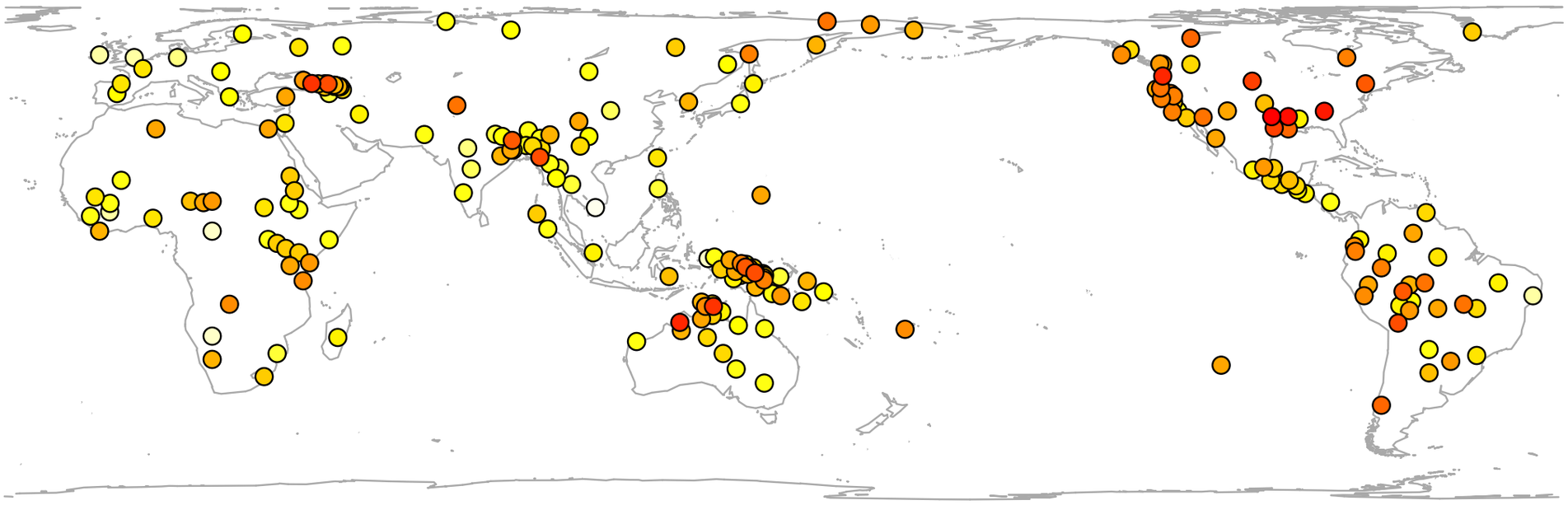


/y/

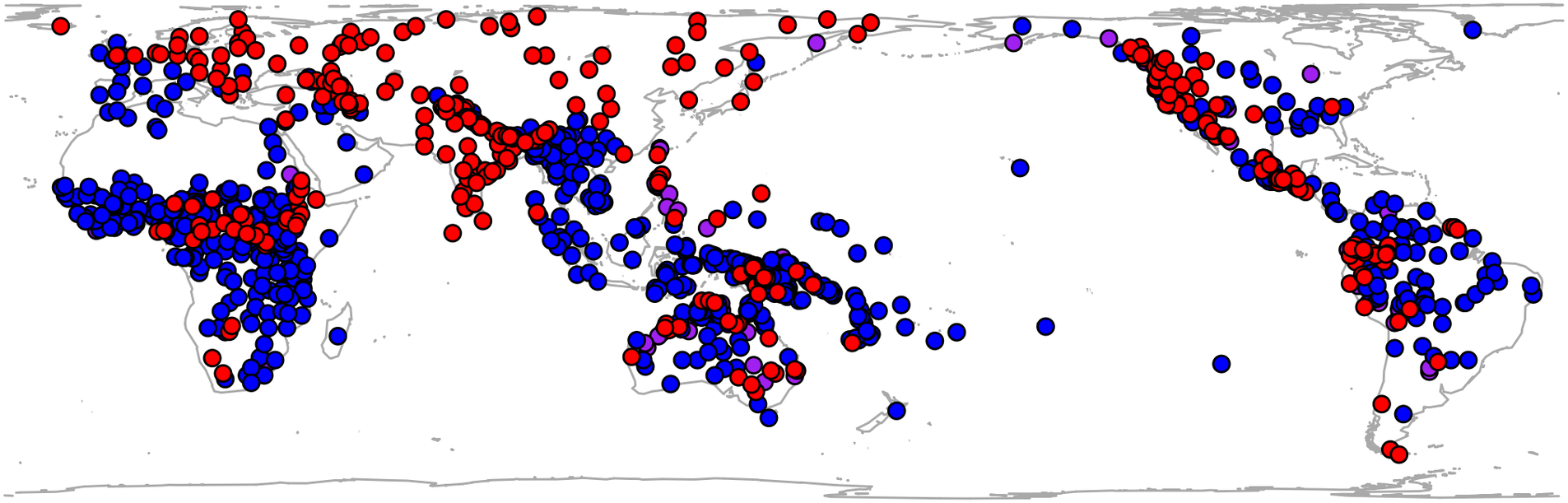


Absence of possessive classes

Cartographic impressions...

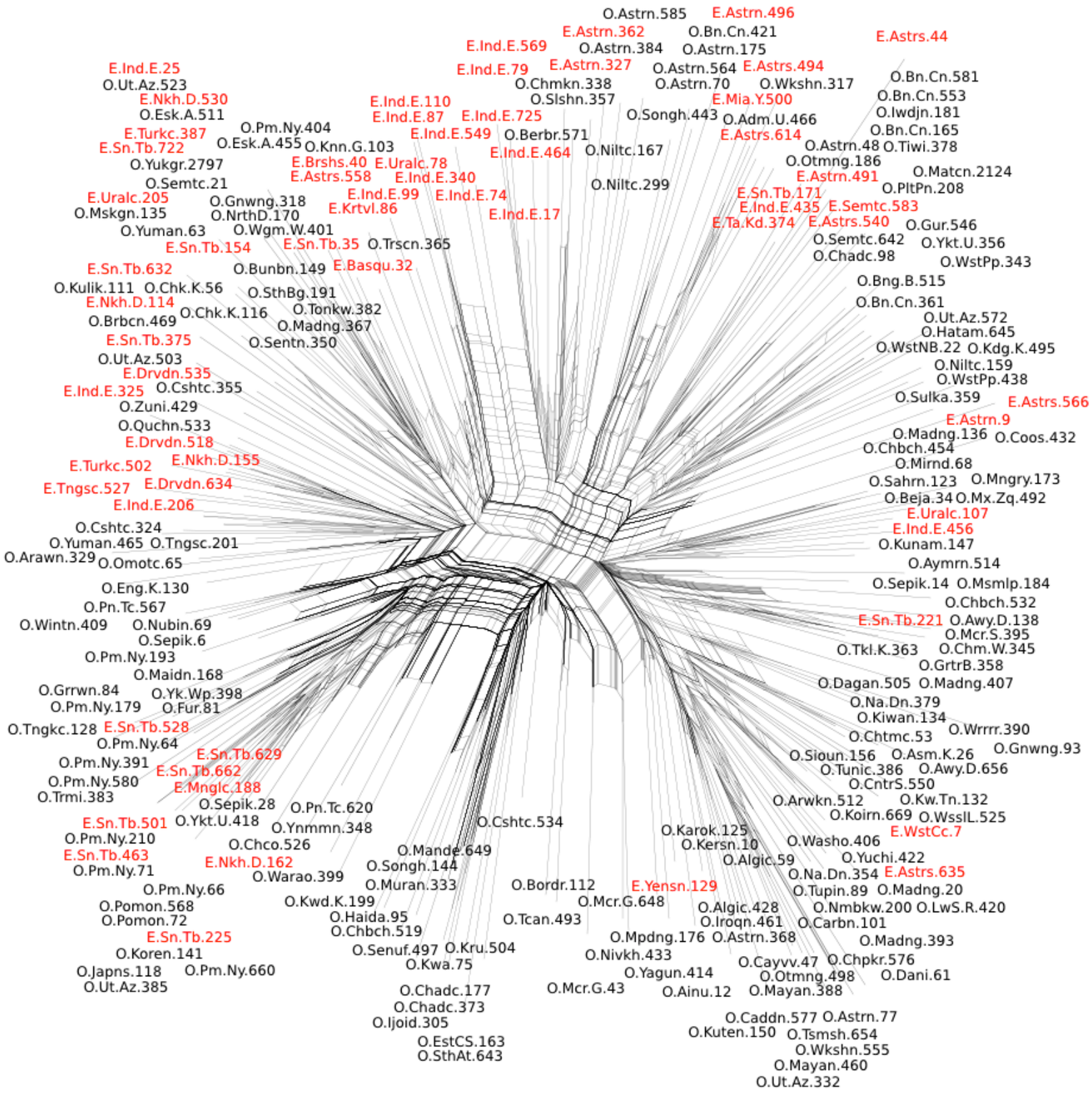


Low verbal inflectional synthesis



AdjN order

Dissimilarity Analysis of 246 languages coded for 507 typological variables



Data from AUTOTYP and WALS, reducing the number of variables and languages so as to minimize and balance the proportion of gaps in the matrix,

optimal at 32.7% NAs

(Euclidean distances for continuous, Gower distances for categorical variables)

(NeighborNet analysis, Bryant & Moulton 2004 *Mol. Biol. Evol.*)

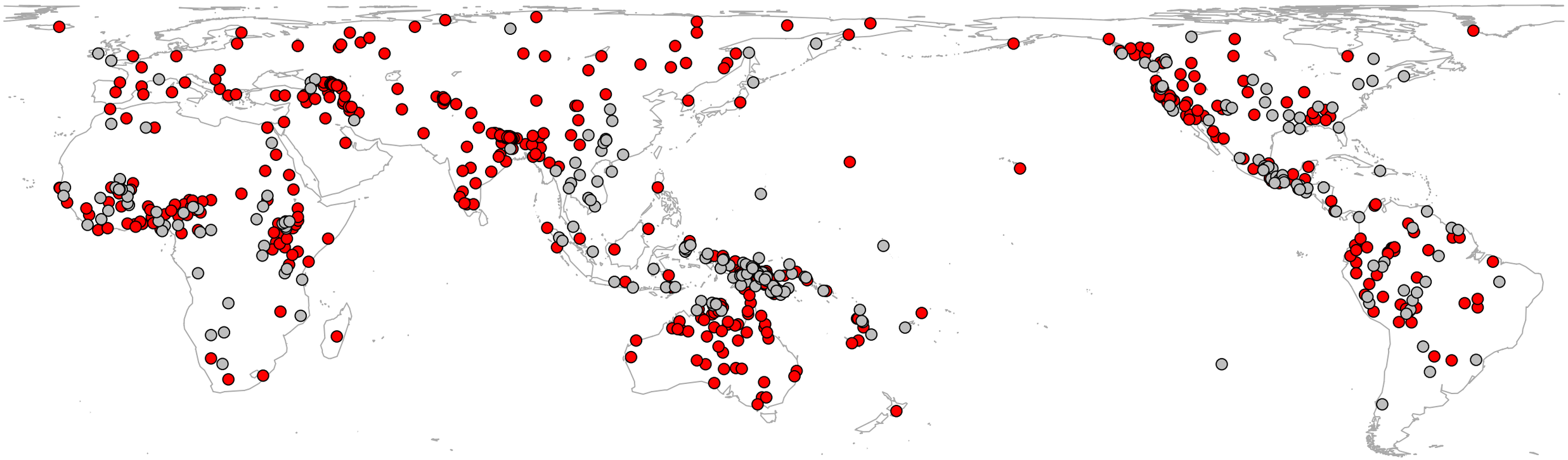
Analysis per variable

- Raw data for 355 variables, between 250 and 2550 datapoints (languages or subsystems of languages):

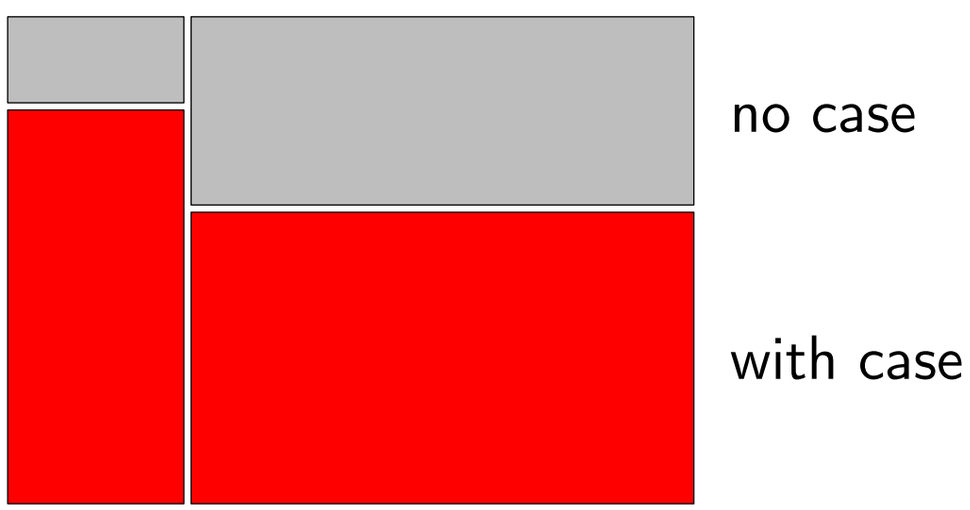
  35% variables show a difference at $p<.05$ (after Holm adjustment)

- But traditional wisdom asks for **genealogically balanced sampling ('g-sampling')**: count only once features that are shared by related languages because
  - the presence of these features may not result from contact, but from inheritance from the proto-language, independent of contact
  - evidence for areas must involve data from non-related languages
- Then, only **13%** show a difference at p<.05% (after Holm adjustment)
- An example ...

Case distinguishing A≠P at least in some NPs and in some valency classes

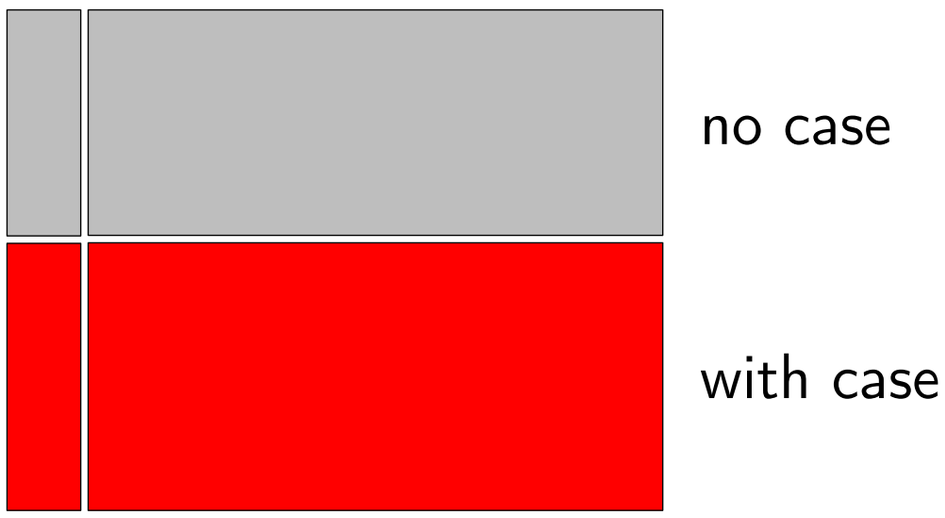


raw data:



$p<.001$

g-sampled data:



$p=.99$

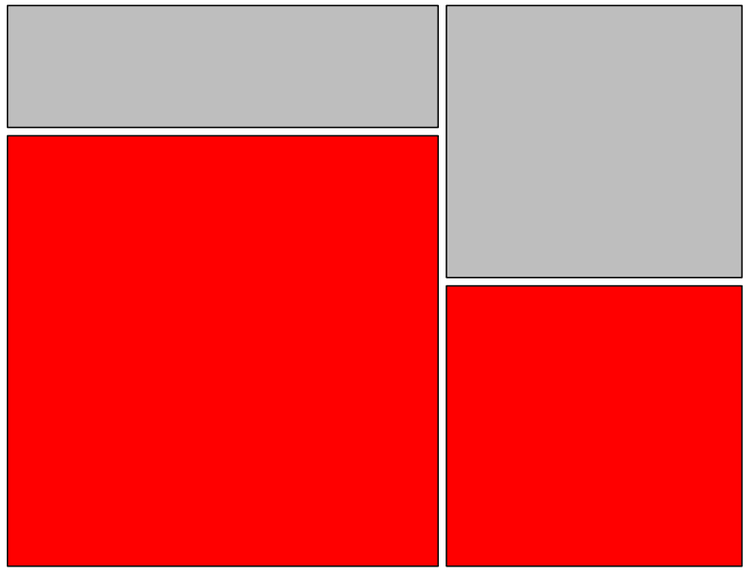
An alternative account of the presence of A≠P cases

- The presence of A≠P cases is perhaps correlated with V-final order (Greenberg 1963, Siewierska 1996, Dryer 2002, Hawkins 2004 etc.)

  [NP V] : [$\emptyset_A$ NP_P V] or [NP_A $\emptyset_P$ V]

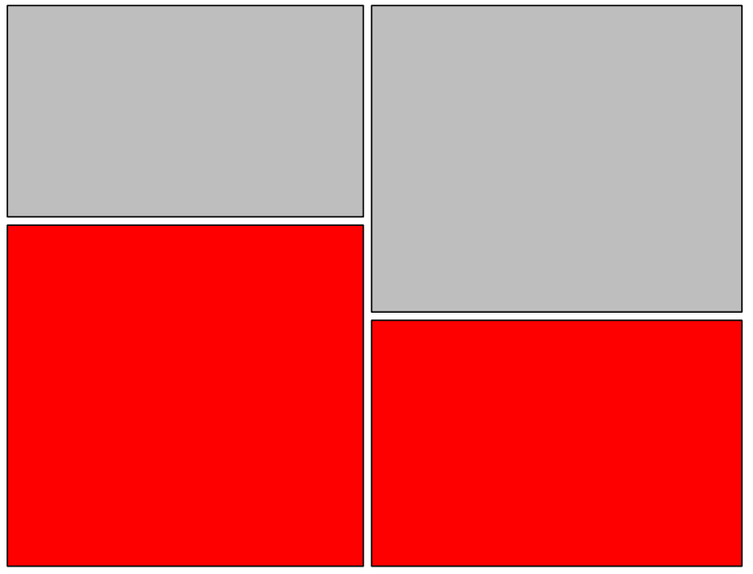
  [NP-*x* V]: [NP-x_A $\emptyset_P$ V]

raw data:



p<.001

g-sampled data:



p=.003

(Fisher Exact Tests)

The problem

- So perhaps many of the variables that seem to show an area effect are better accounted for by processing principles!
- 20th century typology was right! The current trend of asking "what's where why" (Bickel 2007) is a misguided fashion!
- But: **if you want to know about universals, control for areas!**
- and: **if you want to know about areas, control for universals...**

How to find out?

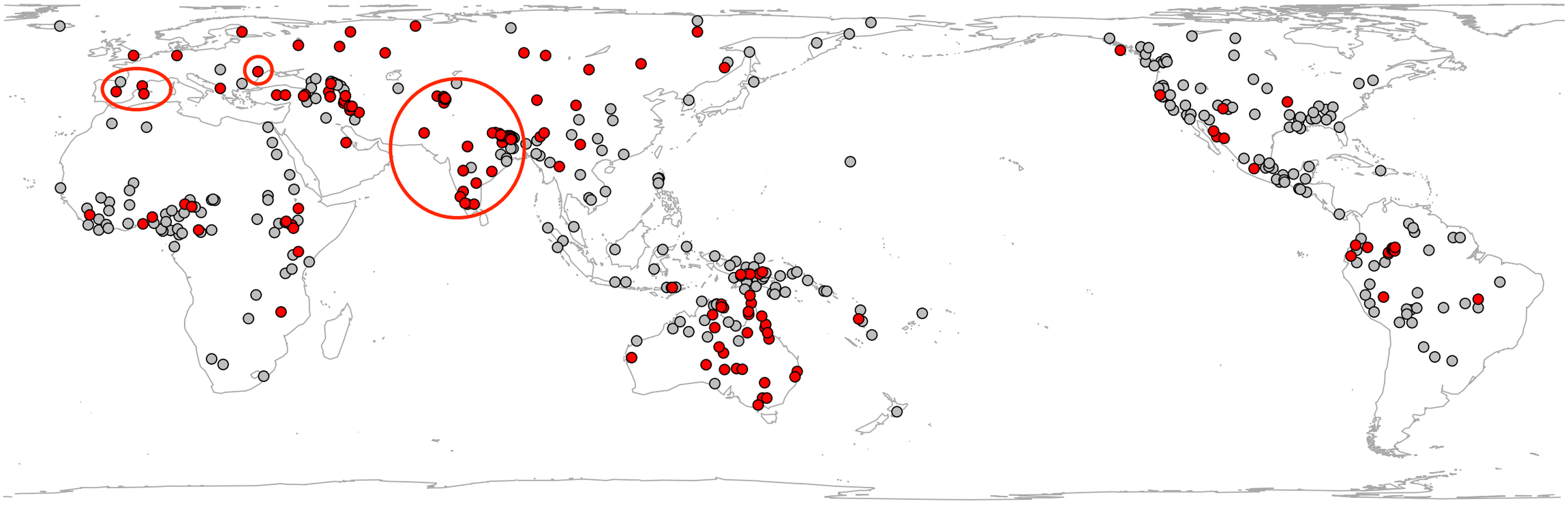
Disentangling factors

- Seemingly obvious solution:
  - build areas and other conditions (e.g. word order) as factors in a statistical model: **CASE ~ AREA × ORDER** etc.
  - and **control for genealogical relatedness** by
    - using g-sampled data (traditional), or
    - building families into the model as one more factor (Bickel et al 2009, Jaeger et al 2011)
- Three concerns...

Three concerns about controlling for genealogical relatedness

1. Shared typological features often do not reflect shared inheritance, e.g.

   - ergativity in Indo-Aryan (e.g. Hindi -ne, Nepali -le)
   - DOM in Romance (e.g. Spanish a, Romanian pe) or Indo-Iranian (e.g. Hindi -ko, Nepali -$lāi$, Persian $râ$)

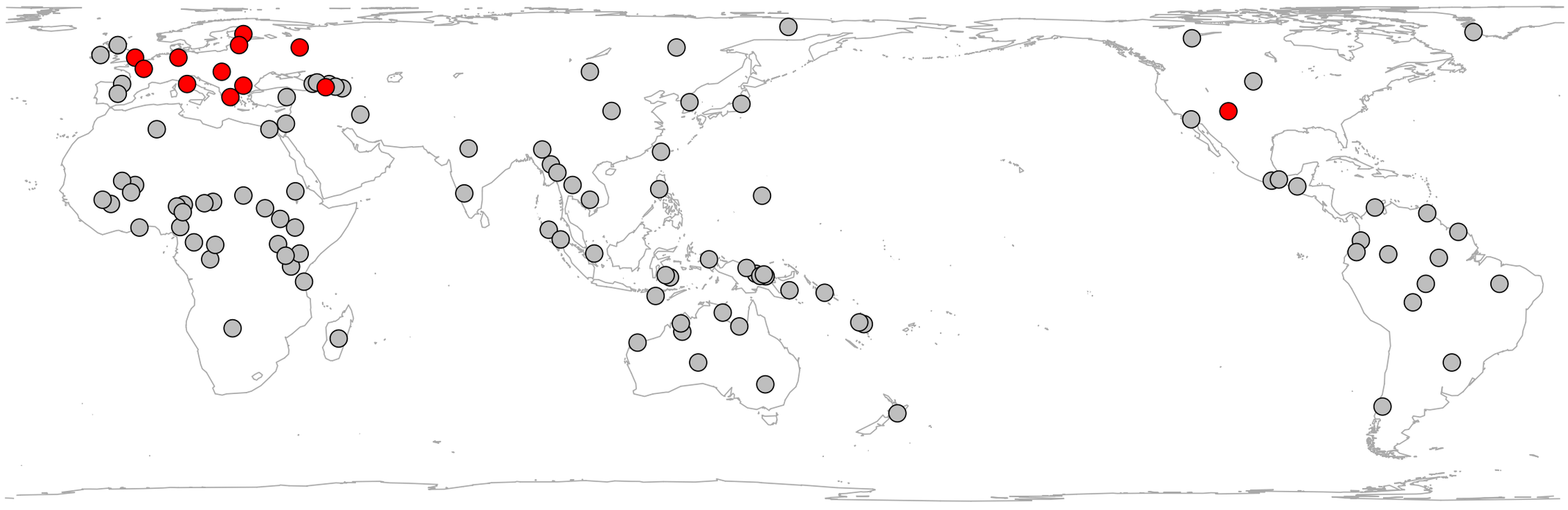


- **Discarding shared features in related languages discards possible signals of areal diffusion**

Three concerns about the traditional typological wisdom

2. Shared inheritance can reflect areal pressure, e.g.

   - it seems more likely to preserve relative pronouns if speakers are in contact with related languages that also have relative pronouns (cf. standard varieties in Europe; data from Kuteva & Comrie 2005)

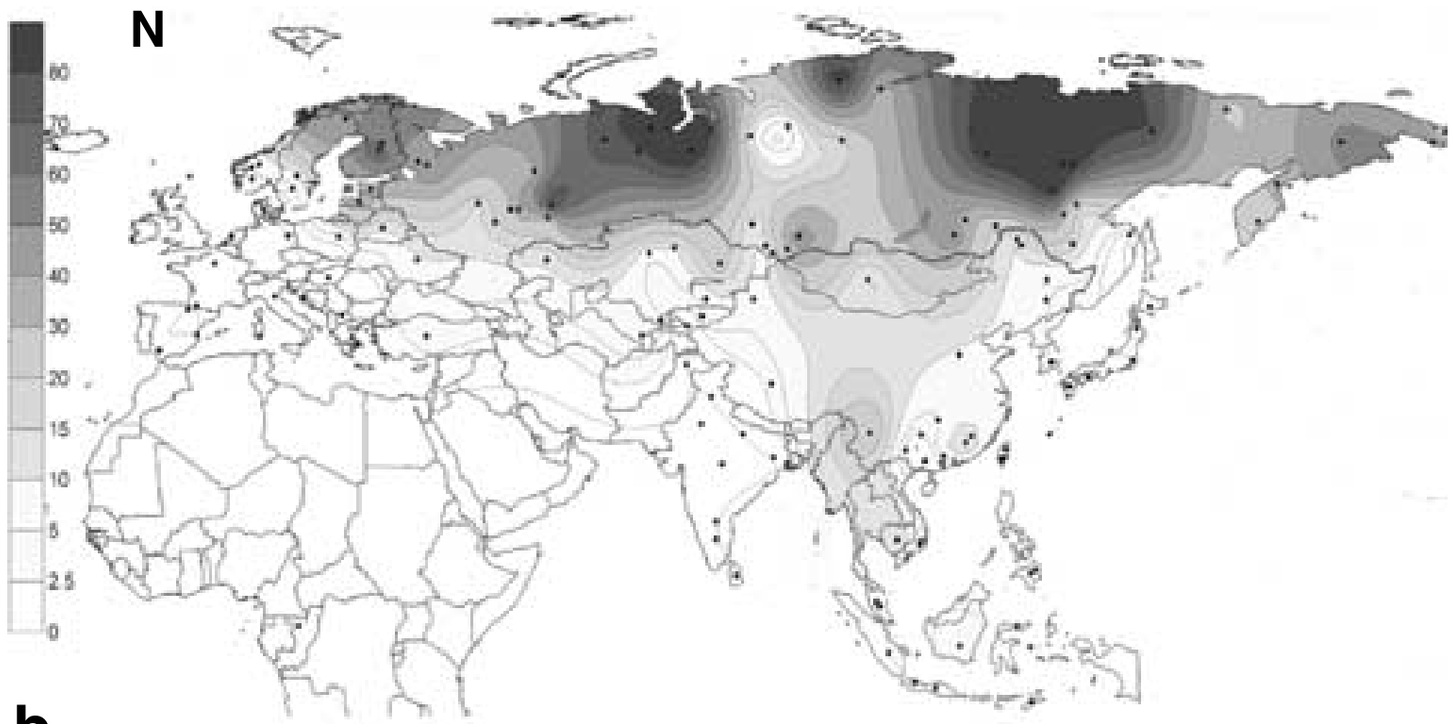


- **Again, discarding shared features in related languages may discard possible signals of areal diffusion**

Three concerns about the traditional typological wisdom

3. Contact is often not a once-off, synchronic event, but operates during long intervals,

   - e.g. thousands of years in Eurasia



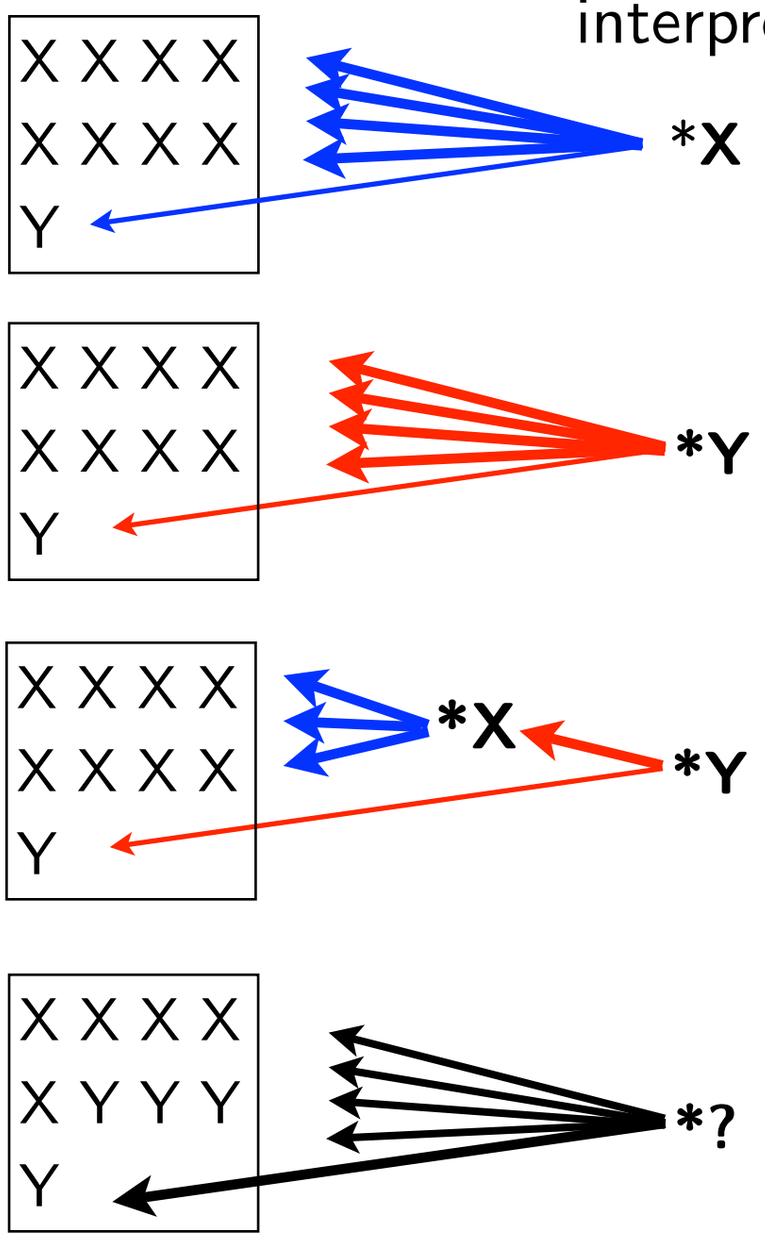
- **Need a diachronic view, but picking only features from *non-related* languages does not allow this in principle.**

Needed: an alternative approach that

- moves beyond synchronic snapshots and estimates diachronic developments
- picks up area signals from shared inheritance as much as from innovation
- allows assessing the effects of contact at the same time as any effects of universals

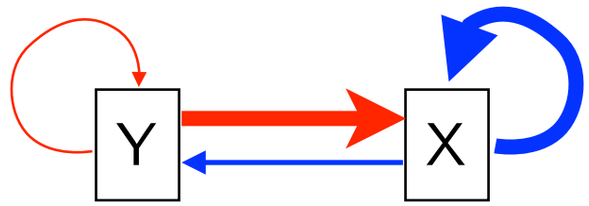
A proposal: Family Bias Method

Synchronic observations on *demonstrably related* languages:



Possible diachronic interpretations:

Conclusion: different probabilities of innovation *and* retention



$Pr(Y{>}X) > Pr(X{>}Y)$

(“**Family Bias**”)

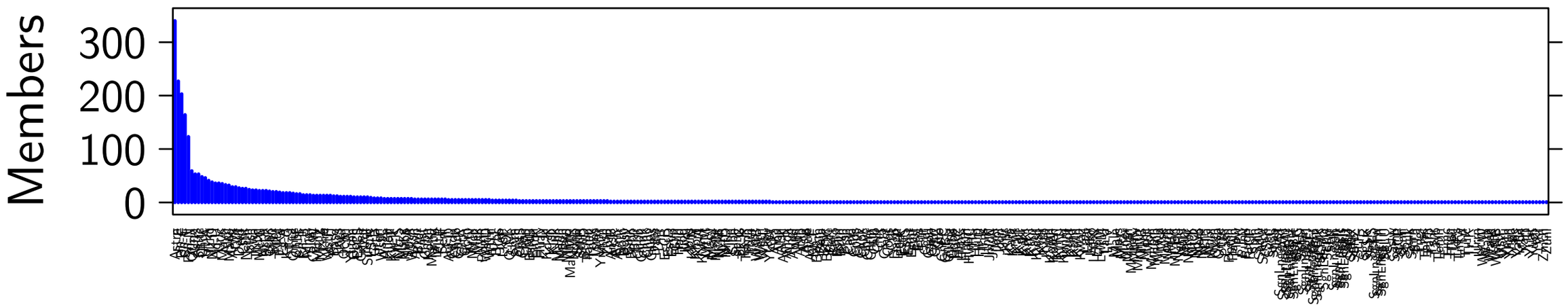
$Pr(Y{>}X) \approx Pr(X{>}Y)$

(“no bias”, “diverse”)

Family Bias Method

- Estimate biases in large families ($N \geq 5$), using binomial tests
- Extrapolate to small families based on bias probabilities of large families and the data in small families, including single-member families (isolates, or families represented only by one member in a given database)

  because, after all, this where the data are



- Software available at `http://www.uzh.ch/spw/software`

Case, Eurasia and word order

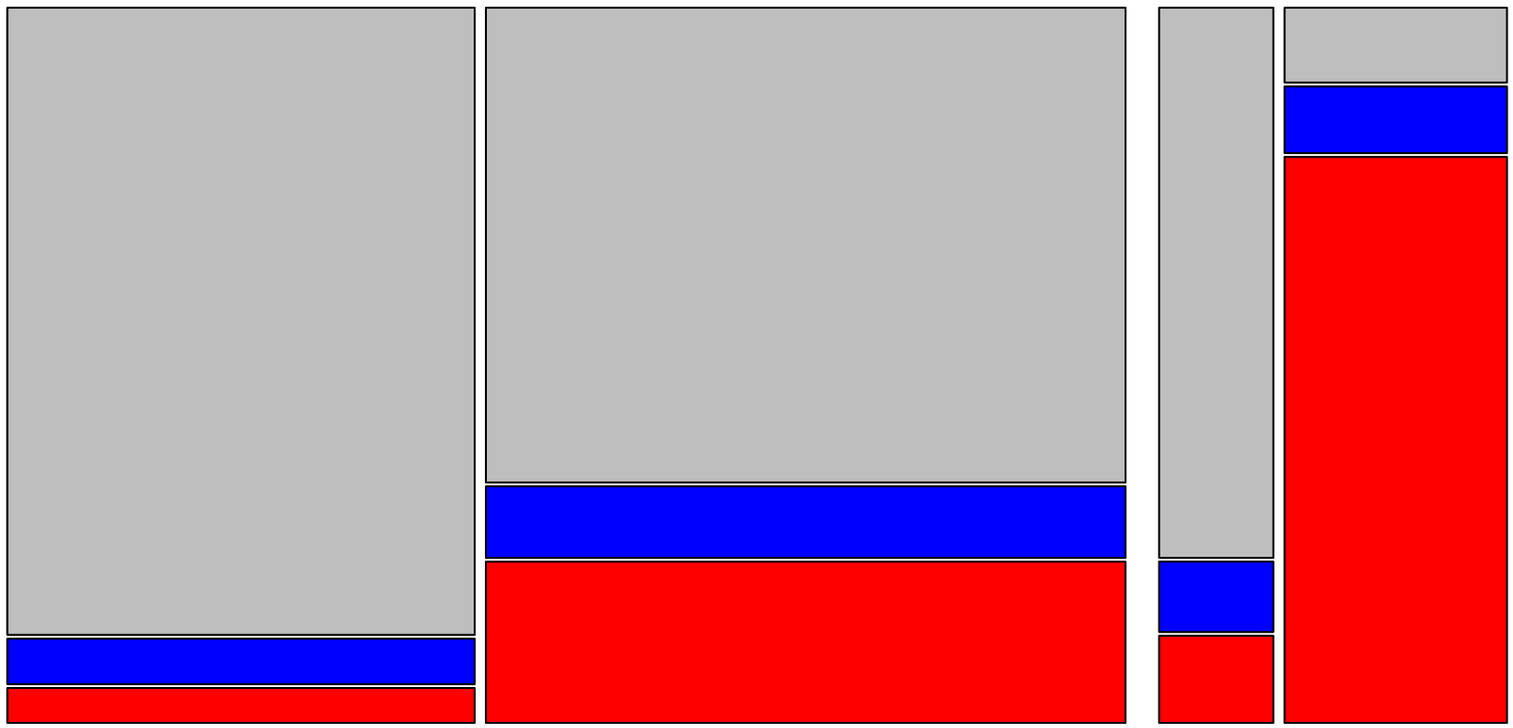
- Data from AUTOTYP (Witzlack-Makarevich et al. 2011+) on case marking and from WALS (Dryer 2005) on word order
  - 489 languages
  - 29 families with at least 5 members
  - 120 small families, including single-member families

- Bias estimates based on these top-level families (stocks), or,

  if these are split between word order types (e.g. Indo-European) or areas (e.g. Austronesian), based on subgroups (e.g. Indo-Iranian vs. Balto-Slavic wrt word order, Oceanic vs. Formosan groups wrt area)

- Extrapolation estimates tentative (need more families with more members)

Case, Eurasia and word order

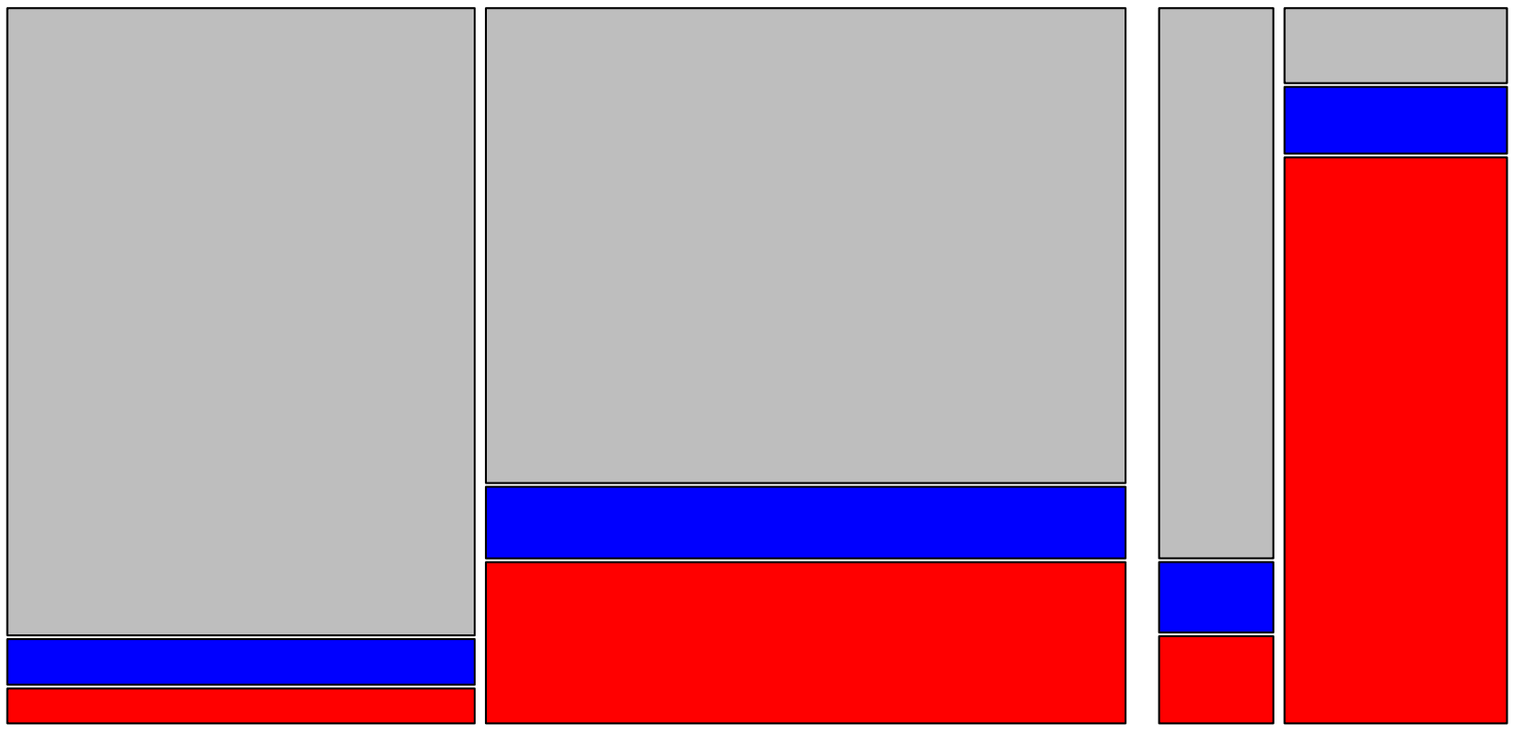


Bias for case vs. against case is determined both

- by the contact history of Eurasia: case tends to be better preserved or (re-)created if in contact with case (AREA × BIAS TYPE, p=.034)
- by processing principles: case is favored in v-final families more than others (ORDER × BIAS TYPE, p=.027)

These effects are independent of each other (three-way interaction is *n.s.*)

Case, Eurasia and word order



Diversification vs. stability is determined both

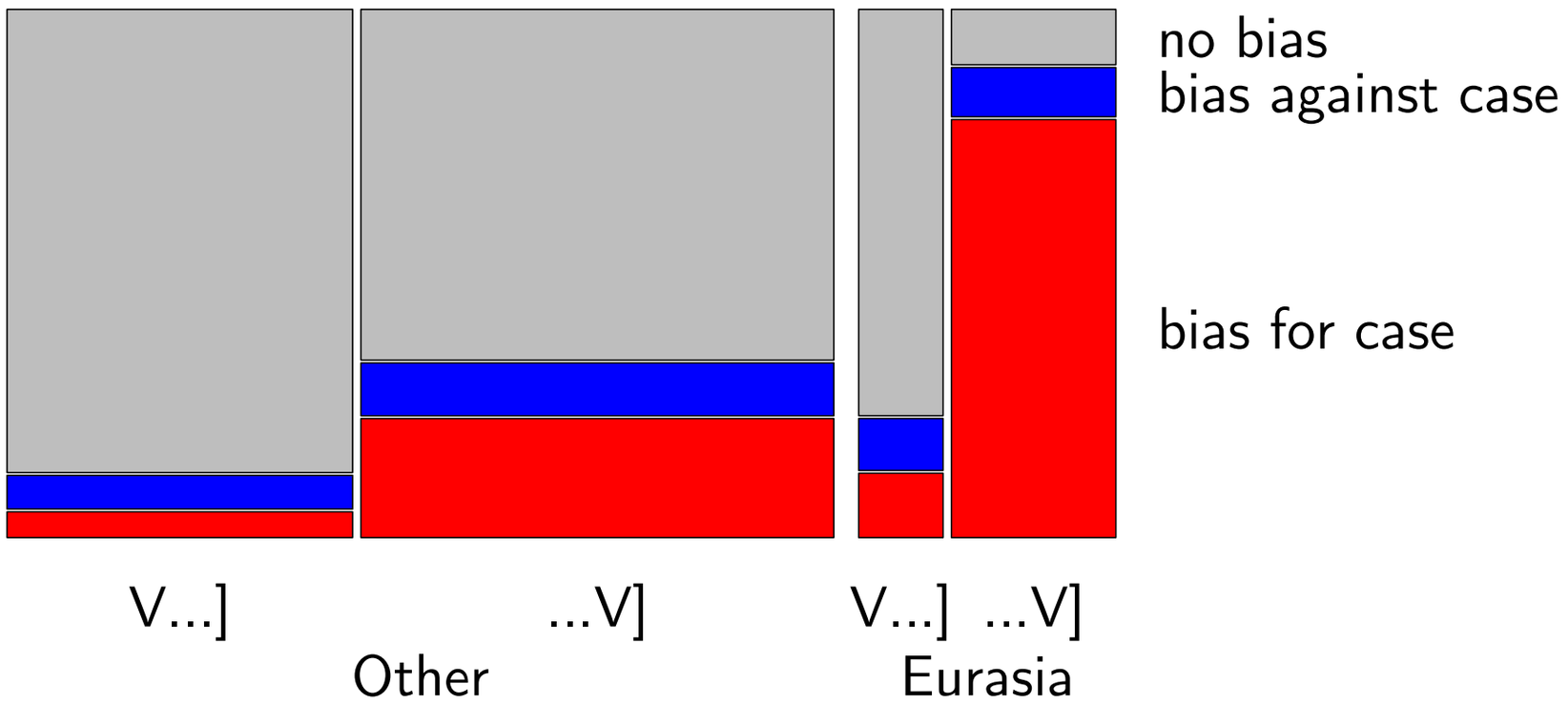
- by the contact history of Eurasia, but only in v-final groups (three-way interaction, p=.011): v-final groups diversify less in Eurasia than elsewhere (AREA × DIVERSITY, p<.001), no such effect in non-final groups
- by processing principles: v-final languages diversify less than non-v-final languages (factorial analysis across areas, both p<.001)

Interim Summary

- The method allows direct estimates on **biases *and* stability**, ***and*** **relative to other factors**
- These factors — in particular, contact histories and processing principles — need to be studied together because:
  - one can't establish one without controling the other
  - area signals may not consist in simple frequency differences but in different extents to which other factors show effects, e.g.:
    - all v-final families favor case, but the ones in Eurasia significantly more so!
    - Eurasian languages favor case, but the ones with v-final order significantly more so!

Interim Summary

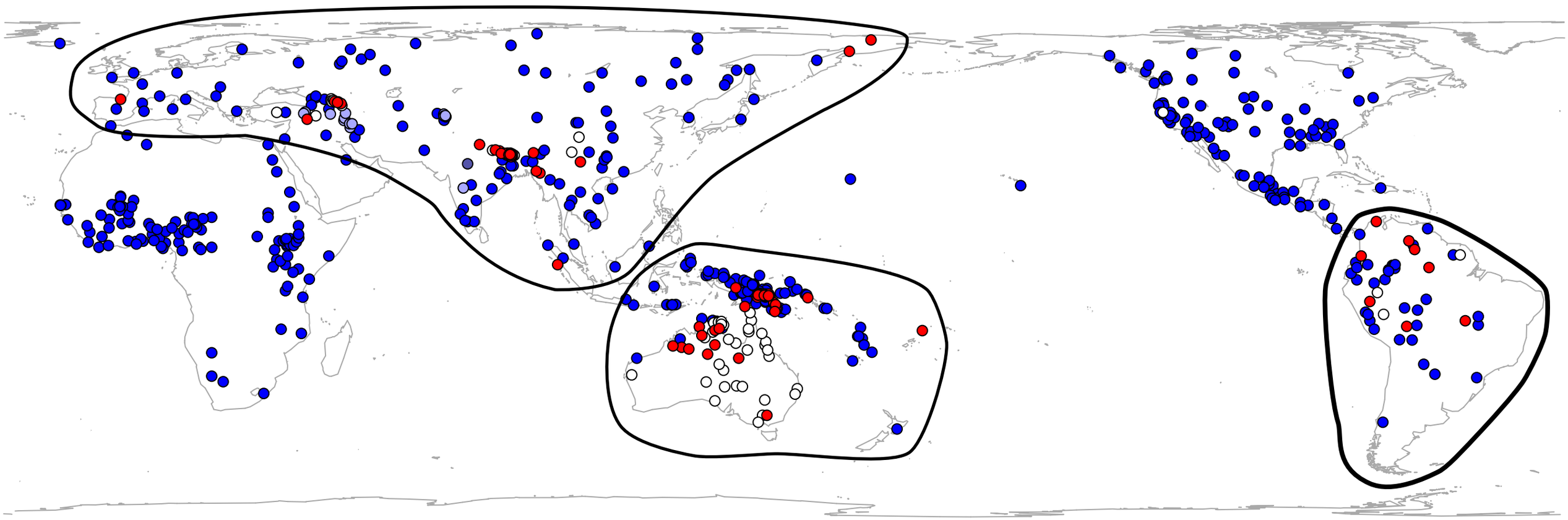
- and also the opposite of area formation — **diversification** within regions — can depend on other factors:
  - v-final languages diversify significantly less wrt to case in Eurasia than elsewhere
  - non-v-final languages tend to diversify equally across areas → no signals for area formation here



One more example

Ergativity in case-marking,

means per languages, across all NP types, clause types, and valency classes:



S≠A S=A

Areal signal?

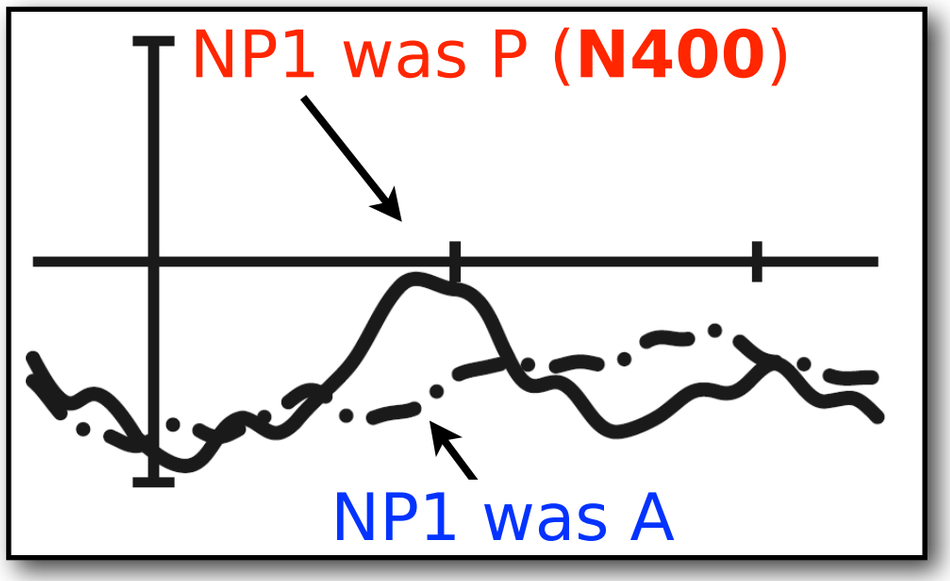
A processing principle: the anti-ergative effect

- Perhaps not: joint work with Ina-Bornessel-Schlesewsky and Alena Witzlack-Makarevich suggests a universal **anti-ergative bias** grounded in processing:

dass *Peter* *Lehrerinnen*
that Peter: ~~S~~/A/P? teachers: A/P?

- ***mögen*** [NP1 was P!]
  like
- ***mag*** [NP1 was A!]
  likes

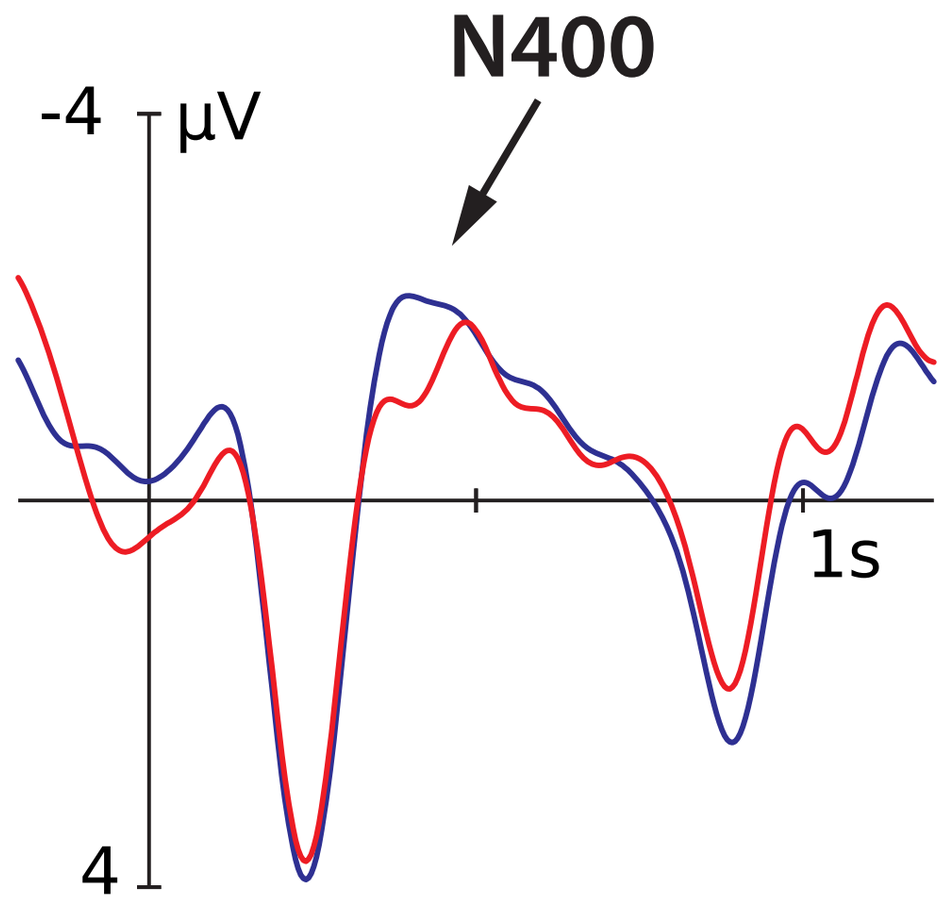
- The brain tends to first assume that NP1 is **S or A, but not P**
- If NP1 later (e.g. at the verb) turns out to be P, this costs something:
  - ERP effect ("**anti-ergative**")



A processing principle: the anti-ergative effect

- Confirmed in many languages, and even in languages with ergative case, such as Hindi

kitāb	***bec-ī***	*(Rām-ne)*
book(FEM)[NOM]	sell-PP.FEM	Ram-ERG
kitāb-ko	***bec-ā***	*(Rām)*
book(FEM)-ACC	sell-PP.MASC	R[NOM]



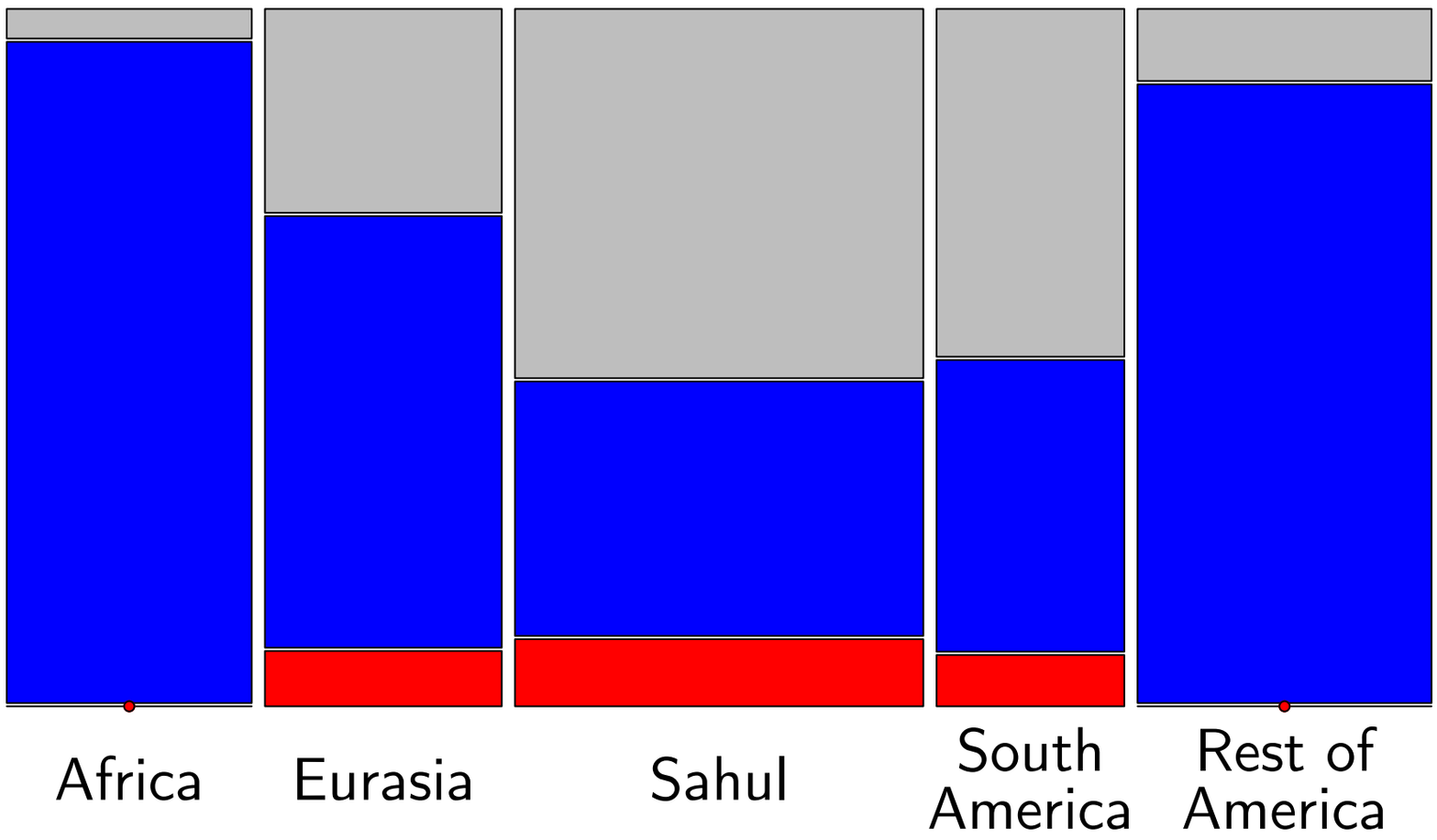
  - Although Hindi NOM structurally includes (and often prefers) a P-reading, the processor first interprets it as S/A!

- Motivated by simplicity of S and primacy of agents (A)

A processing principle: the anti-ergative effect

- Effects weak enough so that ergative cases can be processed and transmitted over generations
- But possibly strong enough to bias diachronic development away form S≠A
- Tested on 601 languages, 695 subsystems (e.g. past vs. nonpast), 158 families, of which 46 families with at least 5 members
- using again the Family Bias Method

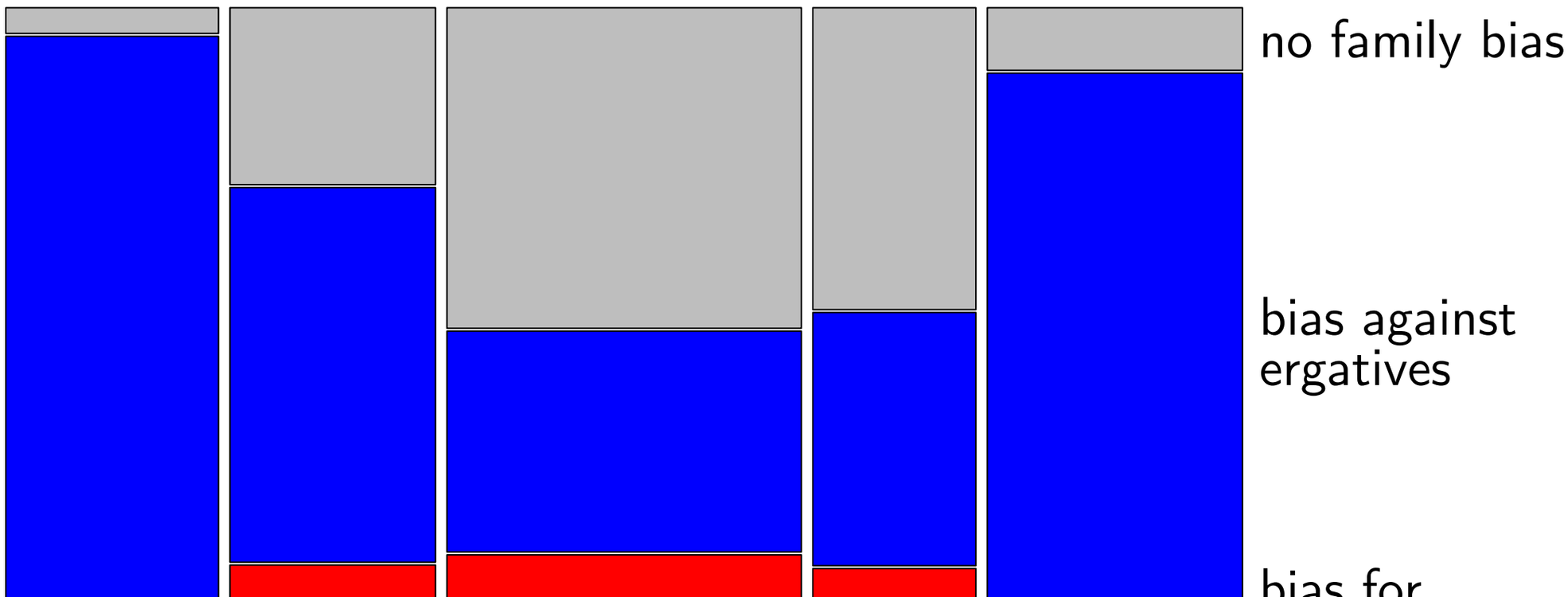
Results



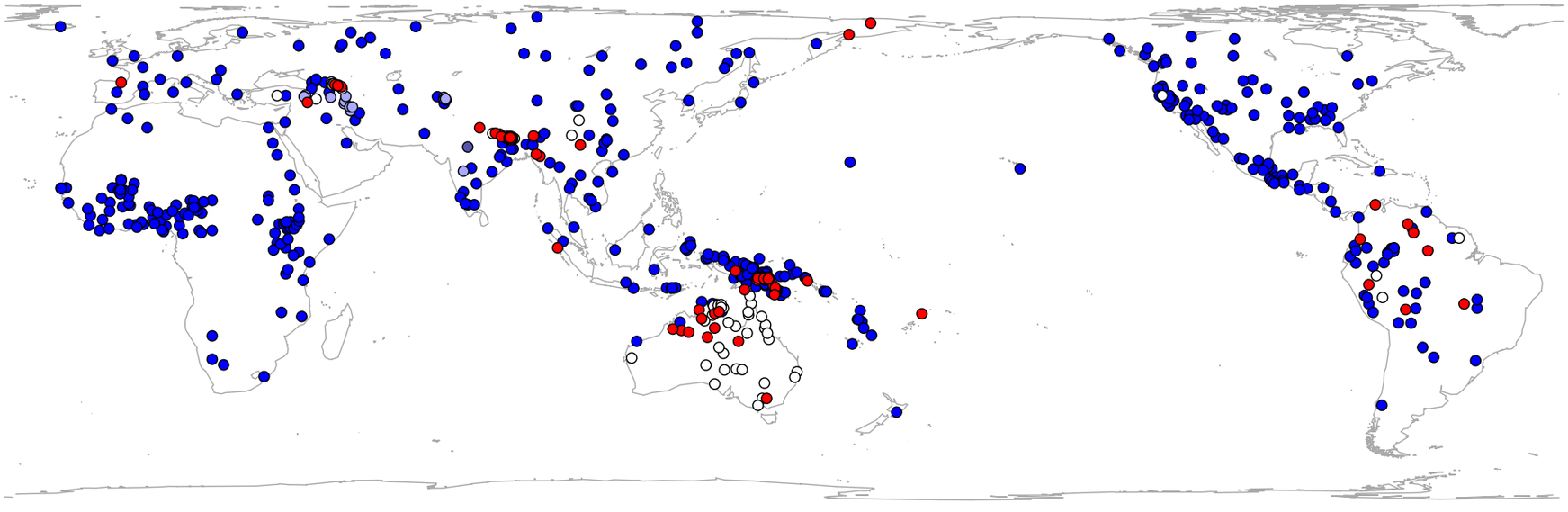
Bias for ergatives vs. against ergatives is determined both

- by contact histories (AREA × BIAS TYPE**,** p=.002)
- by processing principles: proportion of ergative biases smaller than proportion of anti-ergative biases across all areas (all ps<.05)

Results



Diversification strongly depends on area ($p < .001$)



Conclusions

1. Research on linguistic areas can't do without research on universals
   - as a control
   - and because areals signals may be hidden behind effects from processing

Conclusions

2. Like biases, the regional distribution of diversification can be subject to processing principles (e.g. less in Eurasia wrt case in v-final languages, no area effects with other orders)

   → **stability metrics need to be relativized!**

Conclusions

3. Research on linguistic areas can't afford to factor out, let alone *throw out*, data from related languages because
   - only data from related languages, from families, allow estimating diachronic biases
   - and areas are diachronic, not synchronic phenomena.

Conclusions

4. Need samples that are as exhaustive as possible

   - The more datapoints we have, the more reliable are the Family Bias estimates,
   - but the method itself is independent of sampling techniques
   - To the extent that the principles of language change did not *fundamentally* change in the past in an area (or worldwide), the results hold for the entire history in the area (or world-wide)