Typological Enclaves

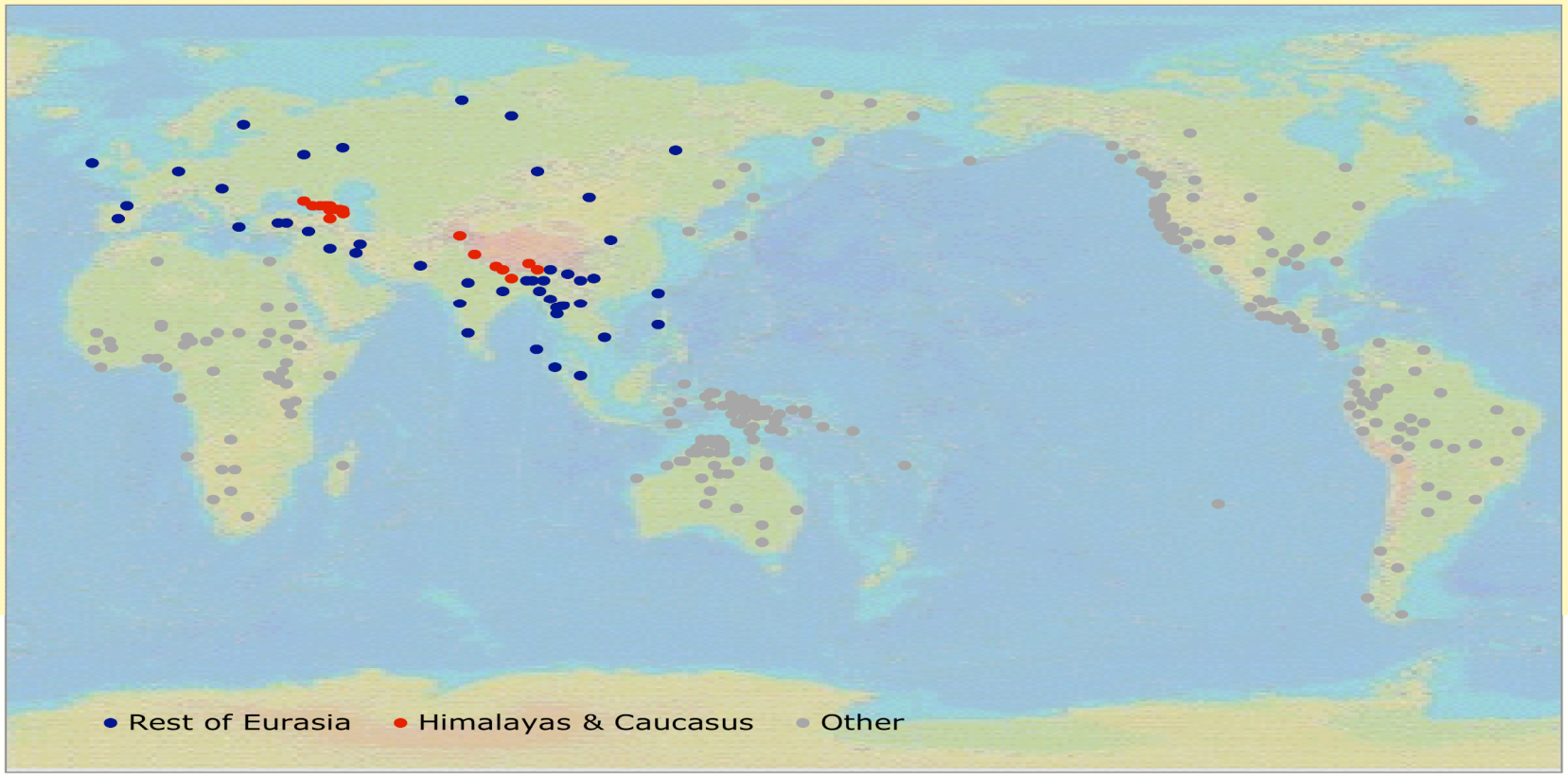
Balthasar Bickel & Johanna Nichols

U. Leipzig, UC Berkeley

http://www.uni-leipzig.de/~autotyp

The phenomenon

- **Pan-Eurasian Macroarea** typological profile defined by several variables (Jakobson 1932).
- **But:** some variables reveal deviating zones in the Himalayas and/or in the Caucasus.



The phenomenon

- **Pan-Eurasian Macroarea** typological profile defined by several variables (Jakobson 1932).
- **But:** some variables reveal deviating zones in the Himalayas and/or in the Caucasus.

- These zones are **typological enclaves:**
  - they are accretion zones at the margins of major spread zones (Silk Road, SEA spreads)
  - they can but need not be areally homogenous; they can be ancient or recent.

Method

1. Survey variables in database suspected to have rare values and to possibly reveal enclave effects
2. Test enclave hypothesis
   - extremely skewed tables: need randomization tests
   - small cell values and high miss-sampling risk: need tolerance test based on re-sampling methods (Janssen, Zúñiga, and Bickel 2003)
3. Interpret in historical scenario model
4. Test hypotheses based on this model
   - same tests as above

Variables surveyed

1a. verb-inflectional synthesis

1b. polypersonal agreement

2. bipartite stems

3. conjunct/disjunct marking

4. multiple possessive classes

5. radically double-marking (objects and possessors)

6. perhaps AND/WITH conjunction (Stassen 2000)

#1a - Synthesis: sampling

A. Survey of categories that are

- **verbal**, e.g. TAM, AGR, NEG, but exluding transcategorial clitics/particles (such as Belhare or Garo interrogatives)
- **inflectional**, i.e. the marker must somehow be sensitive to the syntactic environment (e.g. AGR) or impose or require morphological paradigm/allomorphy choices (e.g. NEG interacting with tense or agreement)
- **synthetic**, i.e. not a syntactic word on its own (e.g. Khasi modal auxiliaries)
- **not necessarily phonologically bound** (e.g. Lai, Khasi TAM particles)

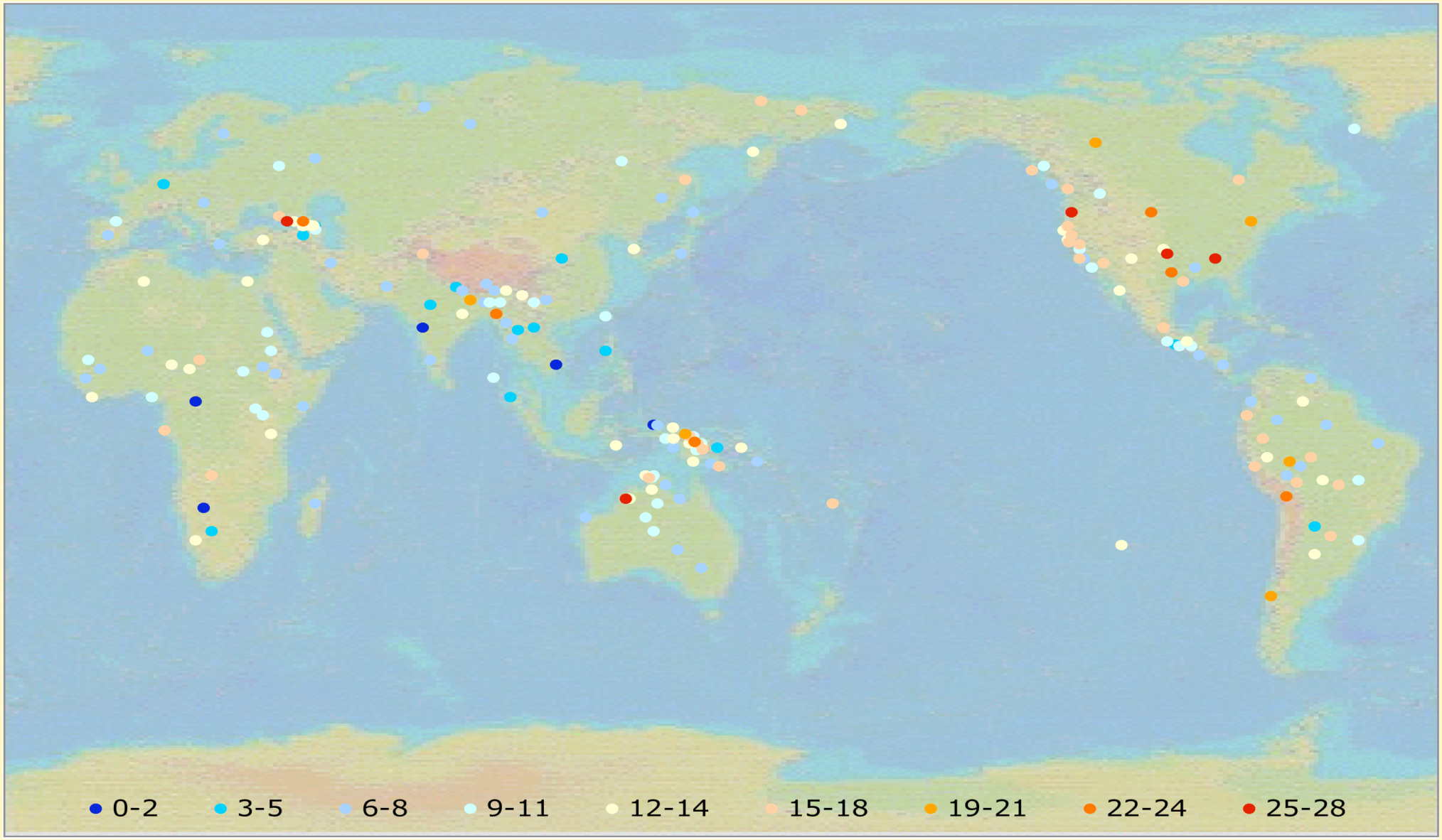
#1a - Synthesis: sampling

B. Survey of formative slots or morphological layers (including tonal layers)

Synthesis Index: sample maximal numbers of categories and formatives

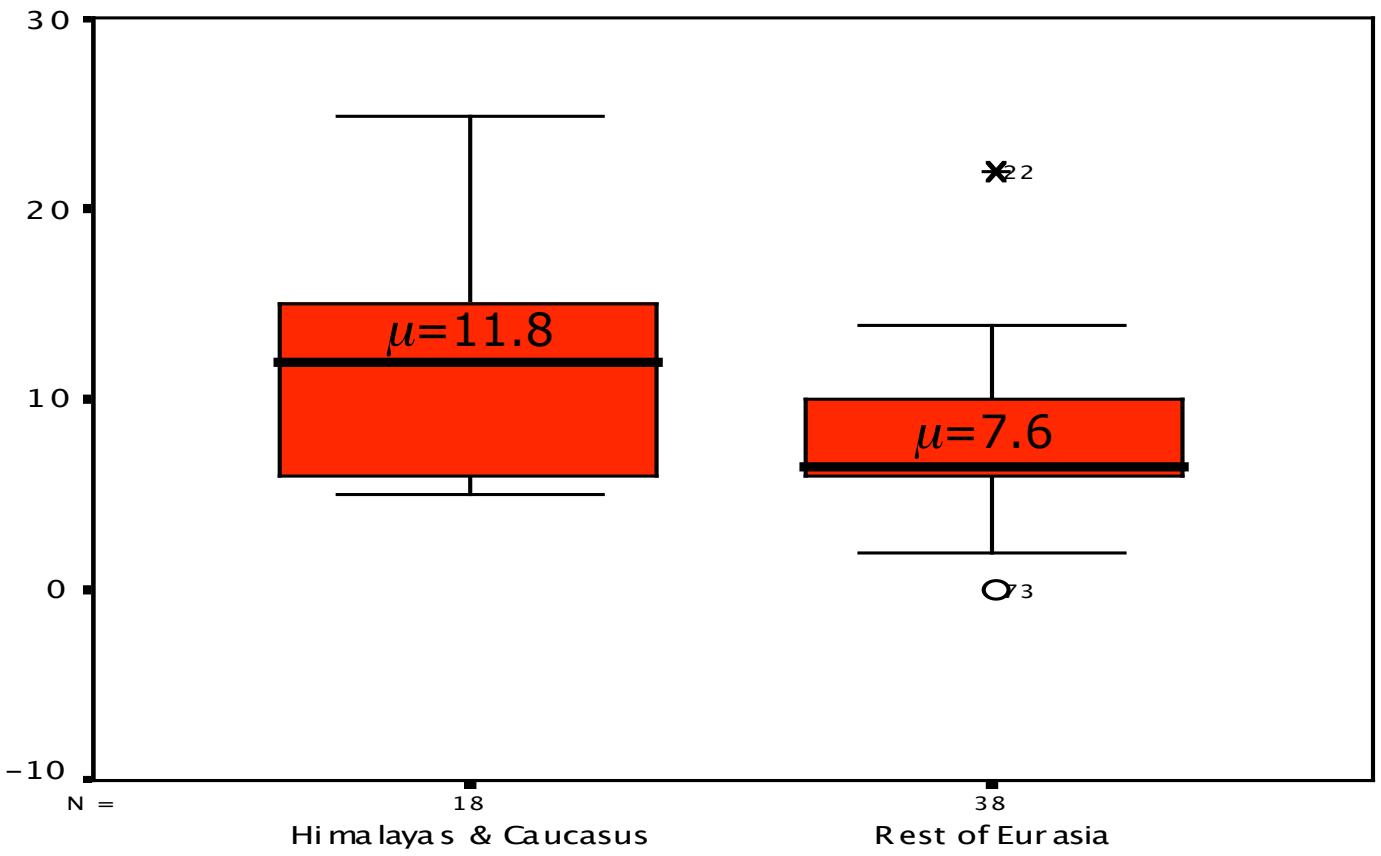
SYN = *n* (categories) + *n* (formatives)

#1a - Synthesis: map



N = 199

#1a - Synthesis: analysis



N	*Mann-Whitney's*	*Randomized ANOVA*	*.05-Tolerance*
56	U=206.5, p=.017	F=8.73, p=.004	T_{sd}=1.2

N (randomizations) = 10,000

#1b - Polyagreement: sampling

Unconditionally obligatory verb agreement with more than one argument

Example: Belhare (Sino-Tibetan, Himalayas)

ma-ŋ-lur-he
1SG.P-3PL.A-tell-PAST
'They told me'

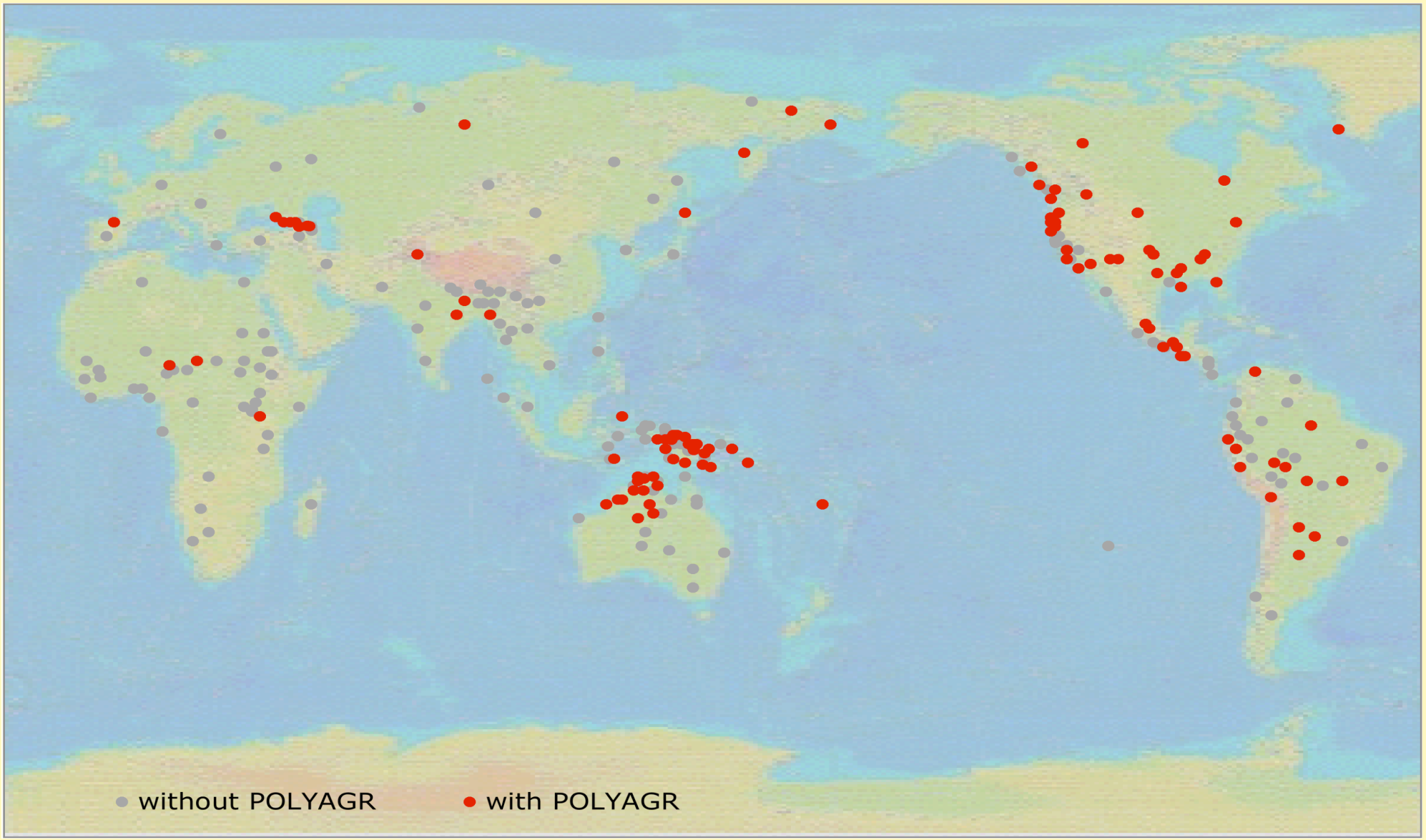
Excluding

- optional agreement as in, e.g., many Bantu lgs (Swahili example):

u-me-(ki-)leta *ki-tabu?*
2SG.A-PERF-VII.P-bring VII-book
'Have you brought a (the) book?'

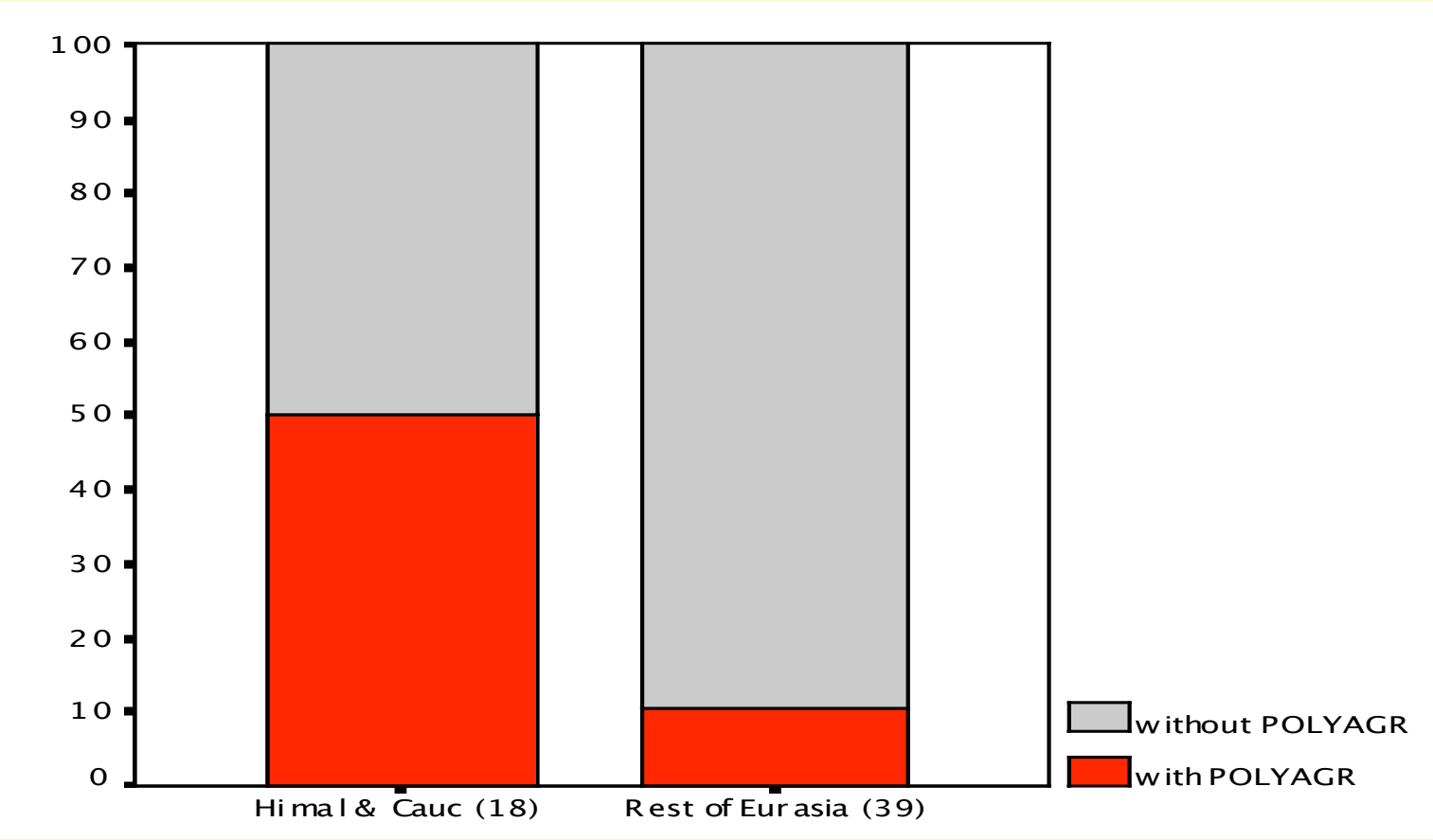
- all kinds of pronominal agreement (agreement in complementary distribution with some NP position)

#1b - Polyagreement: map



N = 277

#1b - Polyagreement: analysis



N	*Fisher's*	*Randomized Chi2*	*.05-Tolerance*
57	p=.002	χ^2=8.91, p=.002	T_{add}=4, T_{sub}=2

N (randomizations) = 10,000

#2 - Bipartite stems: sampling

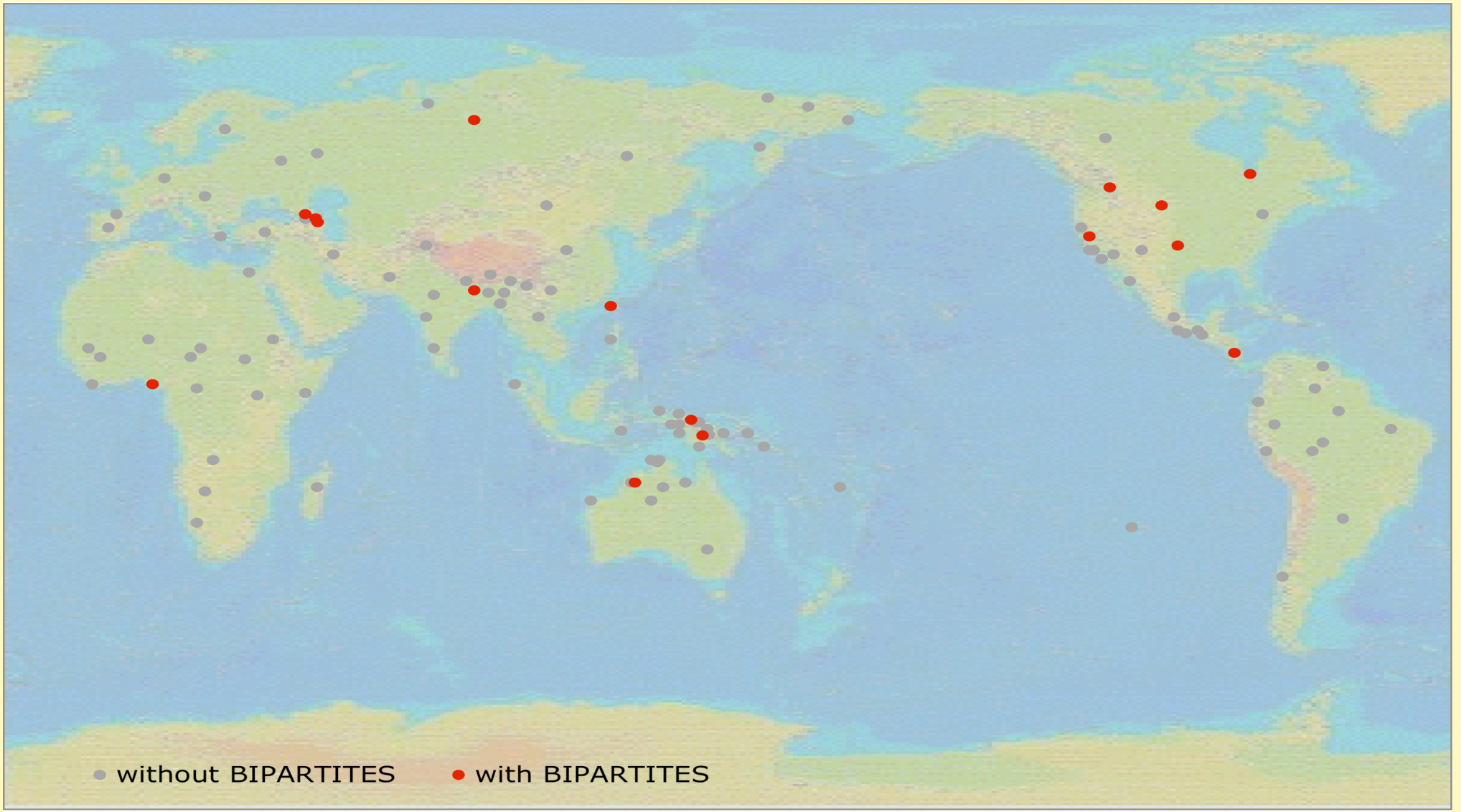
- First identified by Jacobsen 1980 in Washo (isolate, N. California).
- DeLancey 1996: the **Bipartite Stem Belt** in the Pacific Northwest.
- Himalayan example from Belhare:

  laʔu*-ma* 'dance-INF'

  la ***ŋŋ-u****-yakt-he*
  dance 3nsg.S-dance-IMPERFECTIVE-PAST
  'They were dancing.'

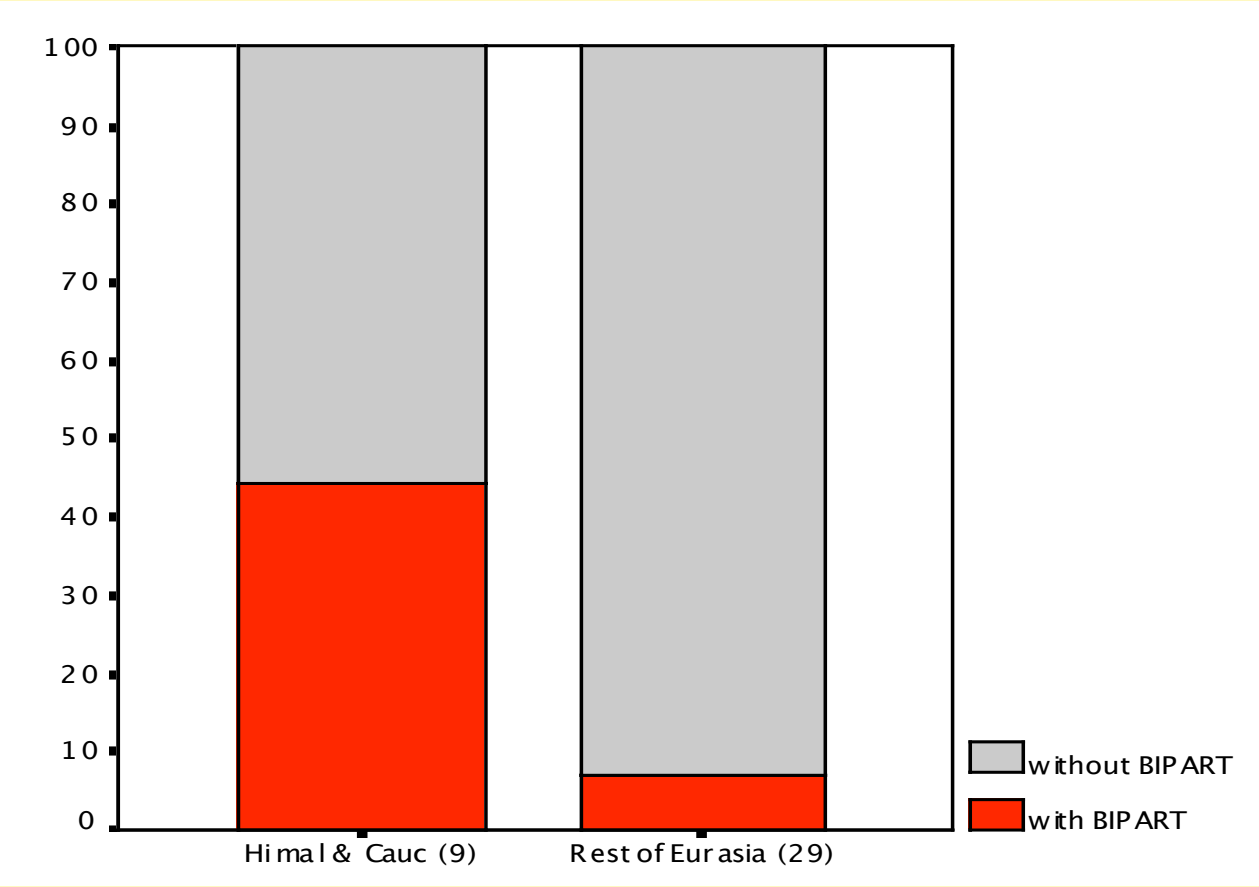
- Data collection incomplete; current sample *N* = 122

#2 - Bipartite stems: map



N = 122

#2 - Bipartite stems: analysis



N	*Fisher's*	*Randomized Chi2*	*.05-Tolerance*
38	p=.020	χ^2= 4.73, p=.024	$T_{add}=1, T_{sub}=0$

N (randomizations) = 10,000

#3: Conjunct/disjunct: sampling

CONJ = **Informant** is {agent, …}

DISJ = **Informant** is not {agent, …}

Newar (Hargreaves 1990)

- *a:pwa* *ton-**ā*** ***/la?***
- too.much drink-**CONJ**.PAST **Q**
- ‘I drank too much’ / ‘Did you drink too much?’

- *a:pwa* *ton-**a*** ***/la?***
- too.much drink-**DISJ**.PAST **Q**
- ‘You/s/he drank too much’ / ‘Did I/s/he/drink too much?’

#3: Conjunct/disjunct: sampling

Awa Pit (Barbacoan; Ecuador - Columbia border region; Curnow 2000)

kïn-ka=na, *na=na* *Santos=ta* *izh-ta-**w**.*
dawn-when=TOP 1SG[NOM]=TOP S.=ACC see-PT-**CONJ**.SUBJ
'At dawn I saw Santos.'

shi *ayuk=ta=**ma*** *libro* *ta-ta-**w**?*
what inside=LOC=**Q** book put-PT=**CONJ**.SUBJ
'Under what did you put the book?'

pïna *alu* *ki-matï-zi.*
very rain do-PFV-PT-**DISJUNCT**
'It rained heavily.'

#3: Conjunct/disjunct: results

Attested in the Himalayas, and only once elsewhere (in S. America). Nowhere else in a database of 350 languages!

#4 - Multiple poss. classes: sampling

Possessive declension classes or lexically determined constructional possessive class distinctions.

Typical examples:

- Diegueño (Yuman, California)

  ʔ-ətal^{y} 'my-mother' vs. *ʔ-ən^{y}-ewaː* 'my-house'

- Warndarang (Maran, Australia)

  ng-baba 'my/our-father' vs. *wuradburru ngini* 'country my'

#4 - Multiple poss. classes: sampling

Multiple possessive classes:

Anêm (New Britain Stock; Papua New Guinea)

	‘water’	‘child’	‘leg’	‘mat’
1sg	*kom-i*	*gi-ng-e*	*ti-g-a*	*mîk-d-at*
2sg	*kom-î*	*gi-ng-ê*	*ti-g-îr*	*mîk-d-ir*
3sgM	*kom-u*	*gi-ng-o*	*ti-g-î*	*mîk-d-it*
3sgF	*kom-îm*	*gi-ng-êm*	*ti-g-î*	*mîk-d-it*

(4 out of 20 classes)

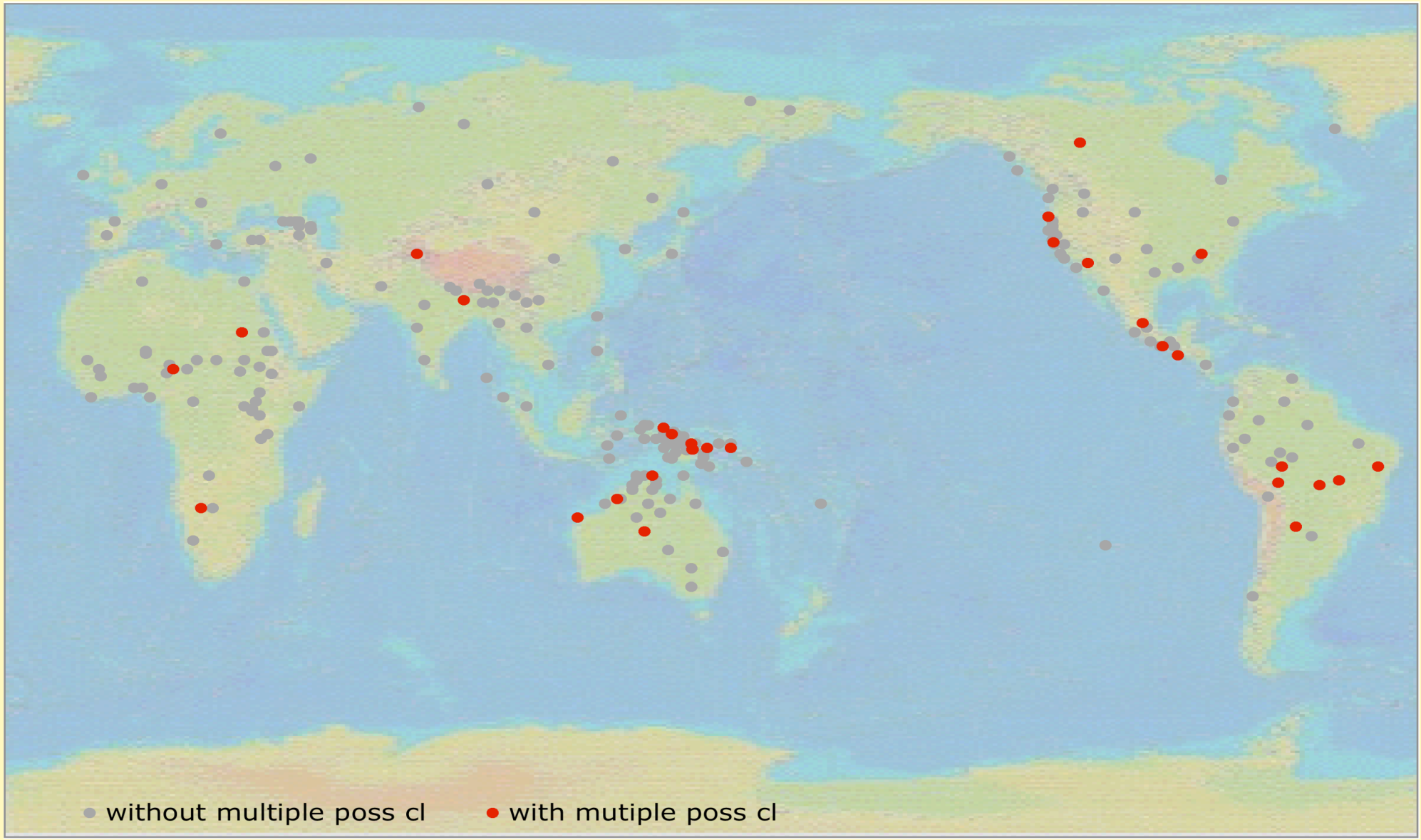
#4 - Multiple poss. classes: sampling

Himalayan example:

Limbu (Sino-Tibetan, Nepal)

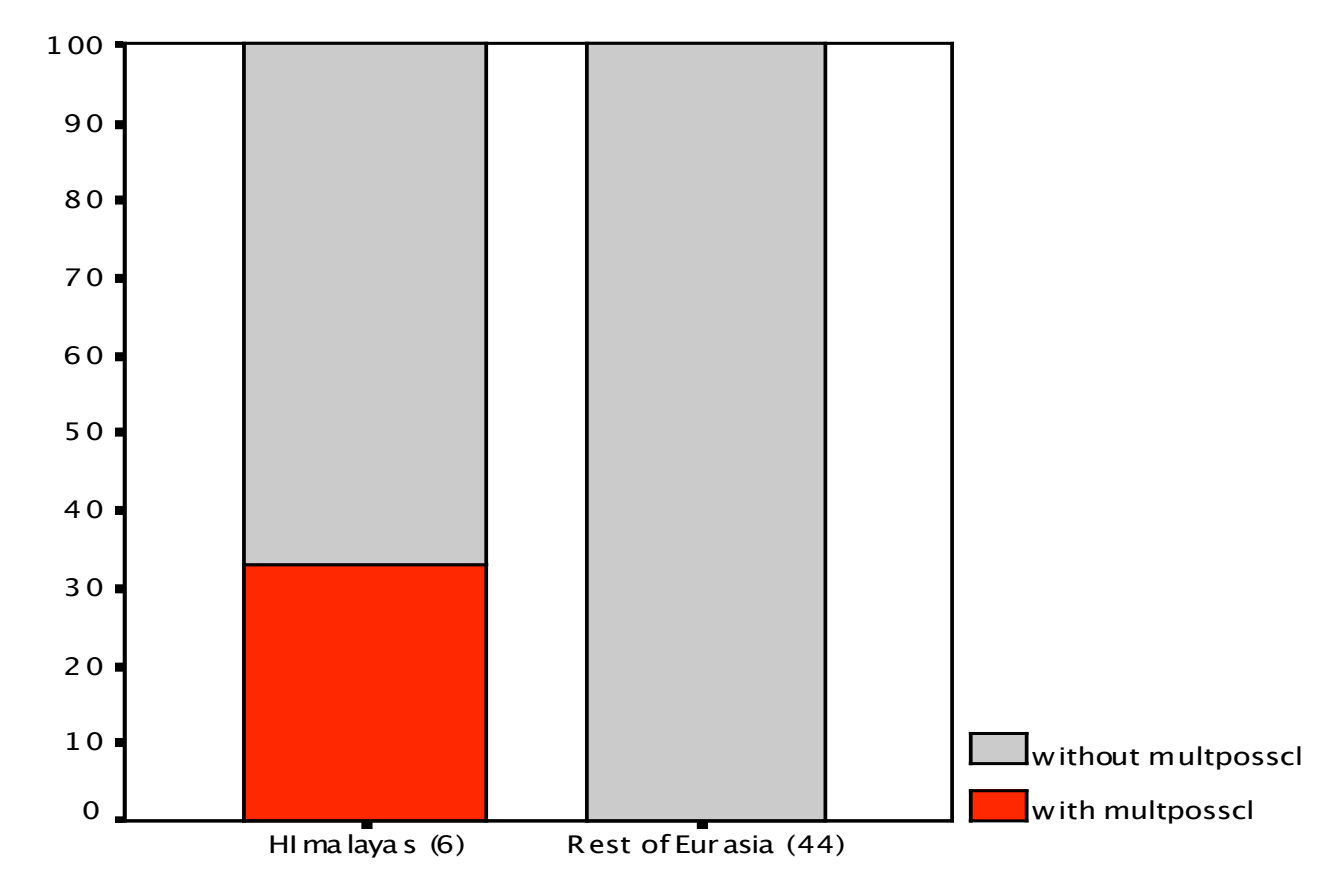
	Class I	Class II	Class III (default)
Effect	Nasalization	Stem reduction	
1sg form	*am-bhɔŋaʔ* 'my uncle'	*a-nsaʔ* (< *nusaʔ*) 'my sibling'	*a-yuma* 'my grandmother'
sample members	friend, father, mother, aunt etc.	head, older sister, moustache, sibling, etc.	

#4 - Multiple possessive classes: map



N = 231

#4 - Multiple possessive classes: analysis



N	*Fisher's*	*Randomized Chi2*	*.05-Tolerance*
50	p=.012	χ^2= 7.83, p=.012	$T_{add}=1, T_{sub}=0$

N (randomizations) = 10,000

#5 - Radically 2-marking: sampling

- Double-marking (optional or unconditional) of both Possessor and Object (any kind) relations.

  Himalayan example:

  a. *sa-**ti*** *lui-t-**u**?*
  who-**ABS.SG** tell-NPT-**3SG.P**

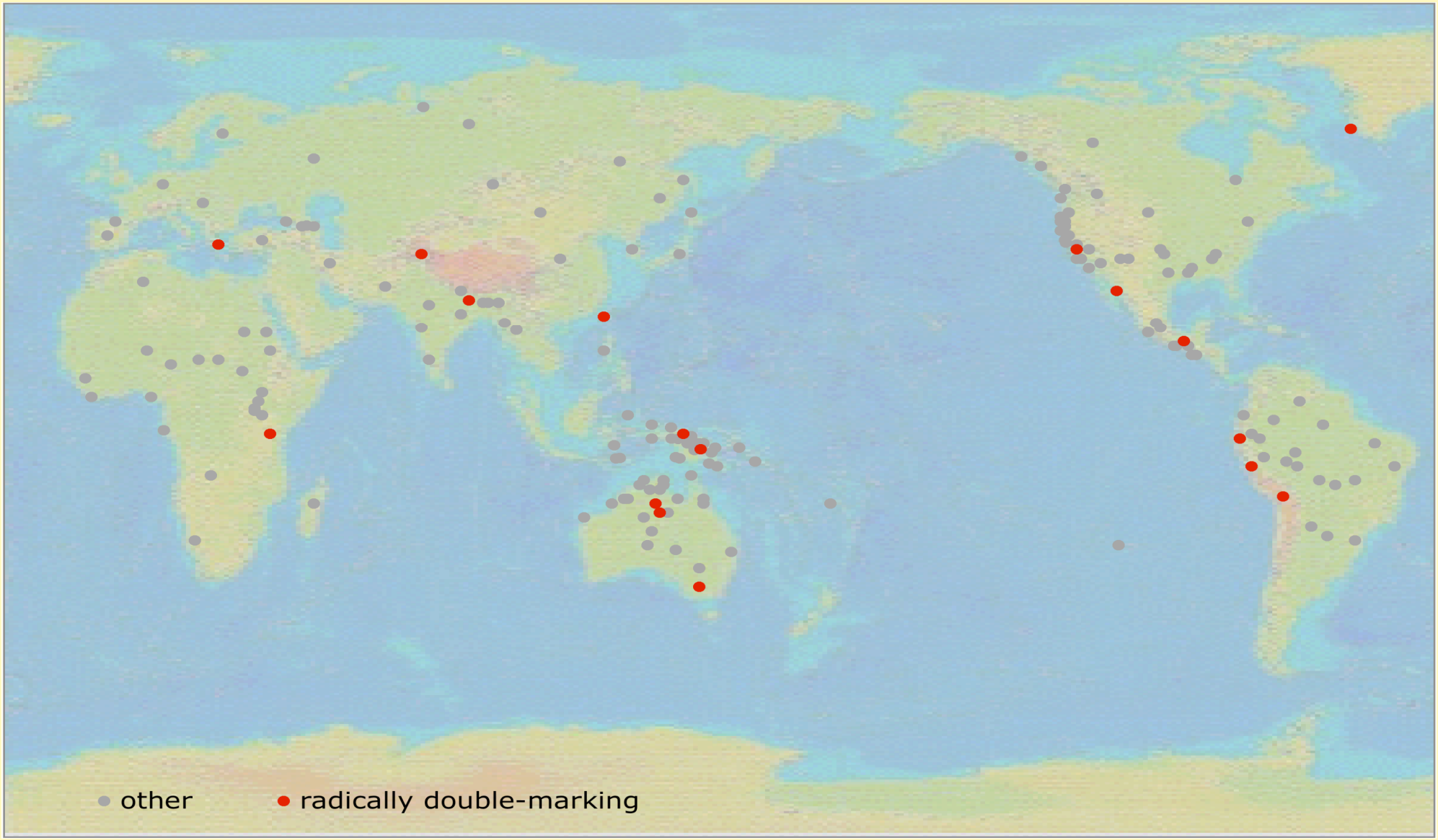
  'Who did s/he tell?'

  b. *sa-**ha*** ***u**-khim?*
  who-**GEN** **3SG.POSS**-house

  'Whose house?'

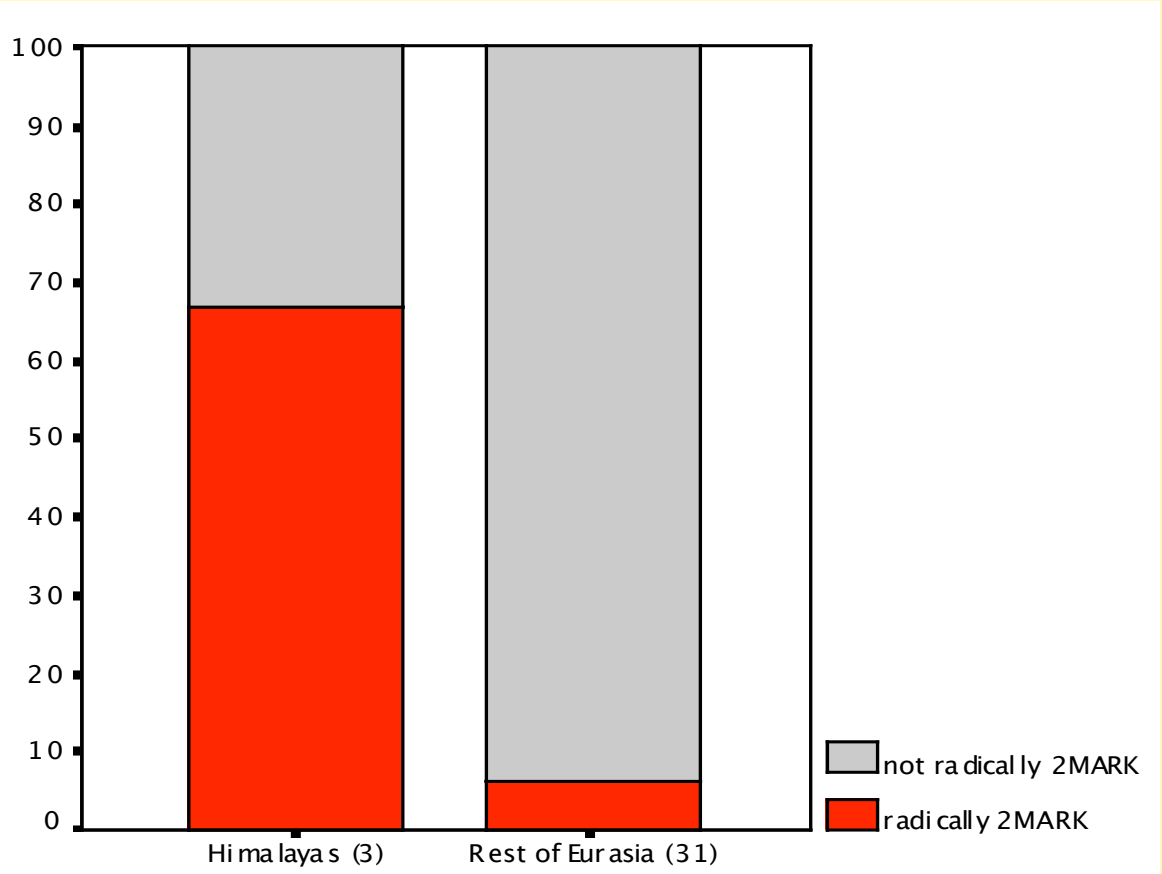
- Subject (S, A) relation excluded from sampling because of universal trend of marking this anyway.

#5 - Radically 2-marking: map



N=186

#5 - Radically 2-marking: analysis



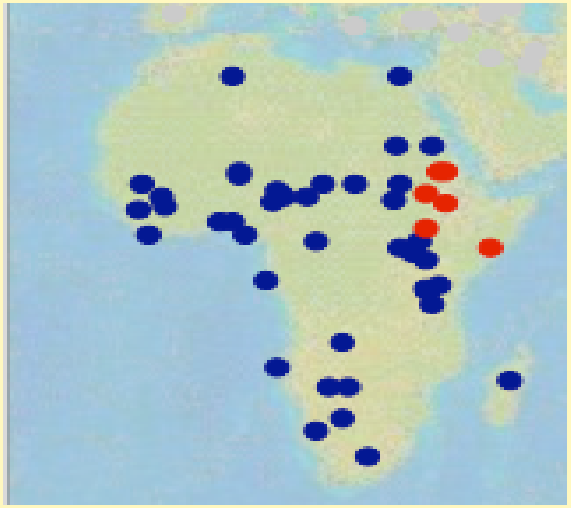
N	*Fisher's*	*Randomized Chi2*	*.05-Tolerance*
34	p=.031	χ^2= 4.63, p=.031	$T_{add}=0$, $T_{sub}=0$

N (randomizations) = 10,000

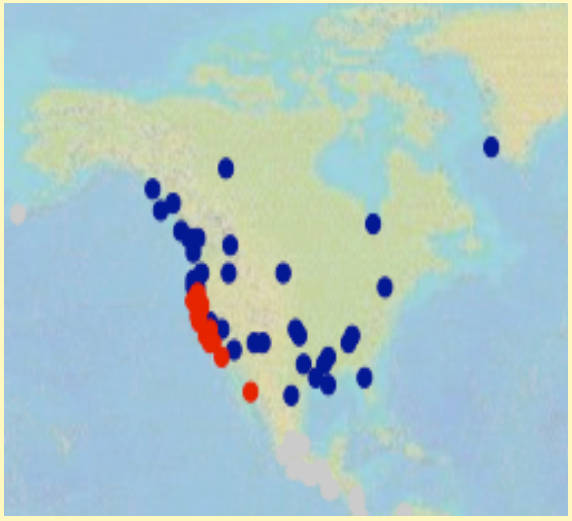
Summary of findings

- highly significant and relatively reliable enclave effects:
  - synthesis (T_{sd}=1.2): Himalayas and Caucasus
  - polyagreement (T_{add}=4, T_{sub}=2): Himalayas and Caucasus
  - conjunct/disjunct systems: Himalayas only
- significant, but relatively less reliable ($T_{add/sub}$<2) enclave effects:
  - bipartite stems: Himalayas and Caucasus (sample incomplete!)
  - radically double-marking: Himalayas only
  - multiple possessive classes: Himalayas only

Enclaves as special accretion zones



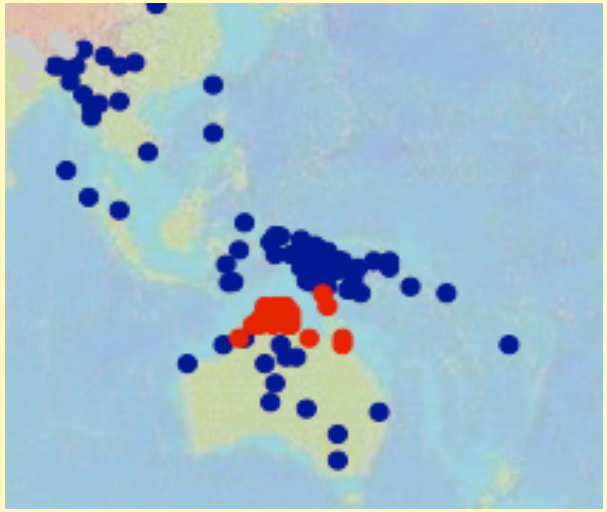
	accr	rest	Nonparametric	Randomization
SYNTH	7.3	10.6	ns	ns
POLYAGR	0%	9%	ns	ns
C/D	0%	0%	n/a	n/a
BIPART	0%	7%	ns	ns
2MARK	0%	5%	ns	ns
MPOSS	0%	9%	ns	ns



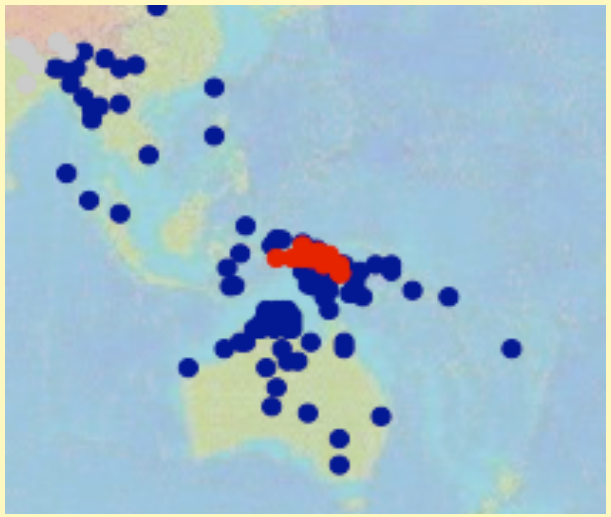
	accr	rest	Nonparametric	Randomization
SYNTH	13	16.7	ns	ns
POLYAGR	45%	81%	p(FE)=.047	χ^2=3.59, p=.046
C/D	0%	0%	n/a	ns
BIPART	0%	50%	ns	ns
2MARK	18%	4%	ns	ns
MPOSS	8%	19%	ns	ns

N(rnd)=10,000

Enclaves as special accretion zones



	accr	rest	Nonpar.	Randomization
SYNTH	13	10.3	ns	ns
POLYAGR	50%	35%	ns	ns
C/D	0%	0%	n/a	n/a
BIPART	0%	10%	ns	ns
2MARK	0%	11%	ns	ns
MPOSS	14%	13%	ns	ns



	accr	rest	Nonpar.	Randomization	.05-Toler.
SYNTH	15.5	9.6	U=76**	F=12.295**	T_{sd}=2.2
POLYAGR	41%	37%	ns	ns	
C/D	0%	0%	n/a	n/a	
BIPART	0%	11%	ns	ns	
2MARK	17%	8%	ns	ns	
MPOSS	13%	13%	ns	ns	

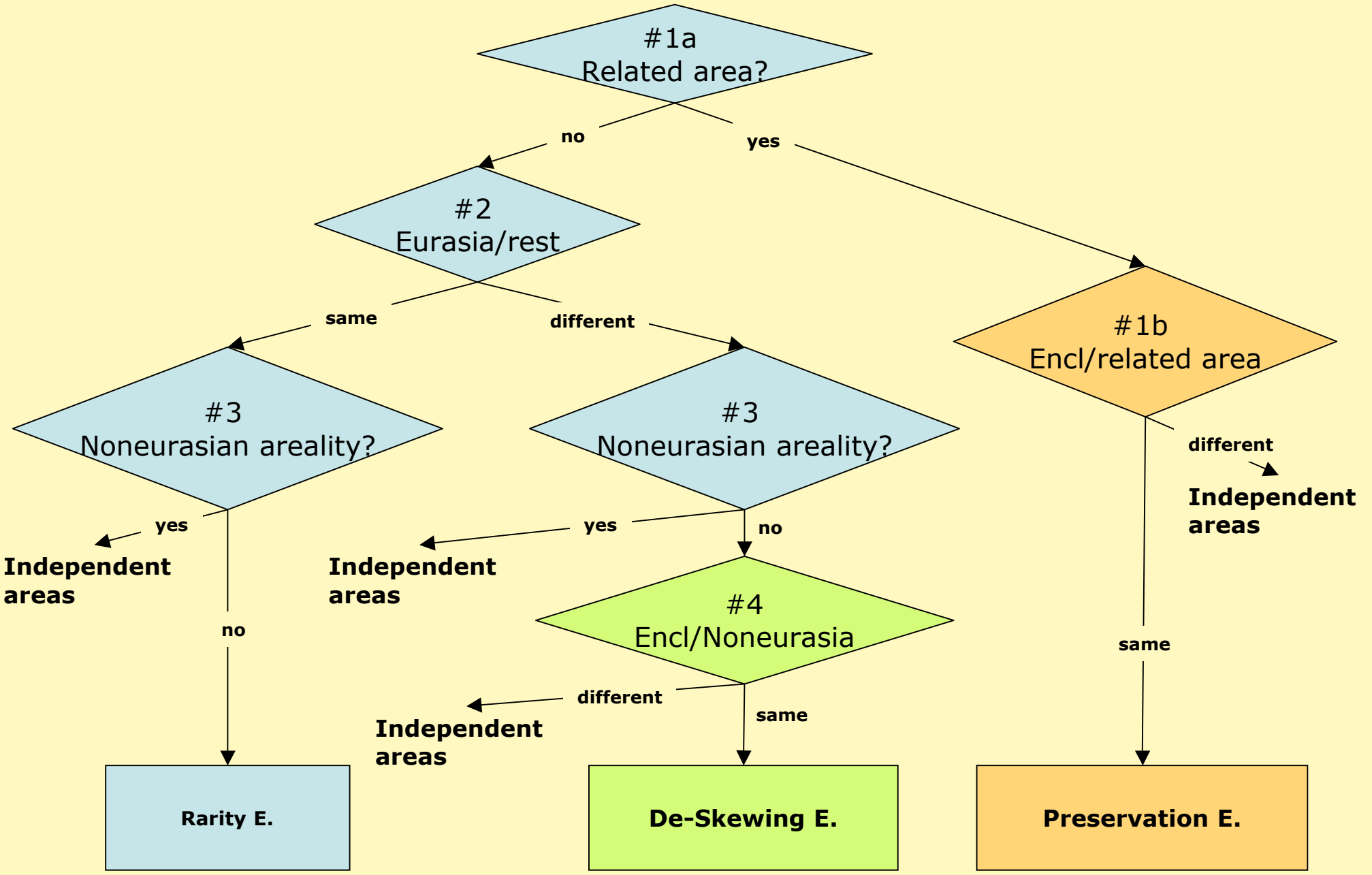
** p< .001, N(rnd)=10,000/1,000

- **Enclave effects are not general accretion zone effects and demand locally specific explanations.**

Types of enclave distributions

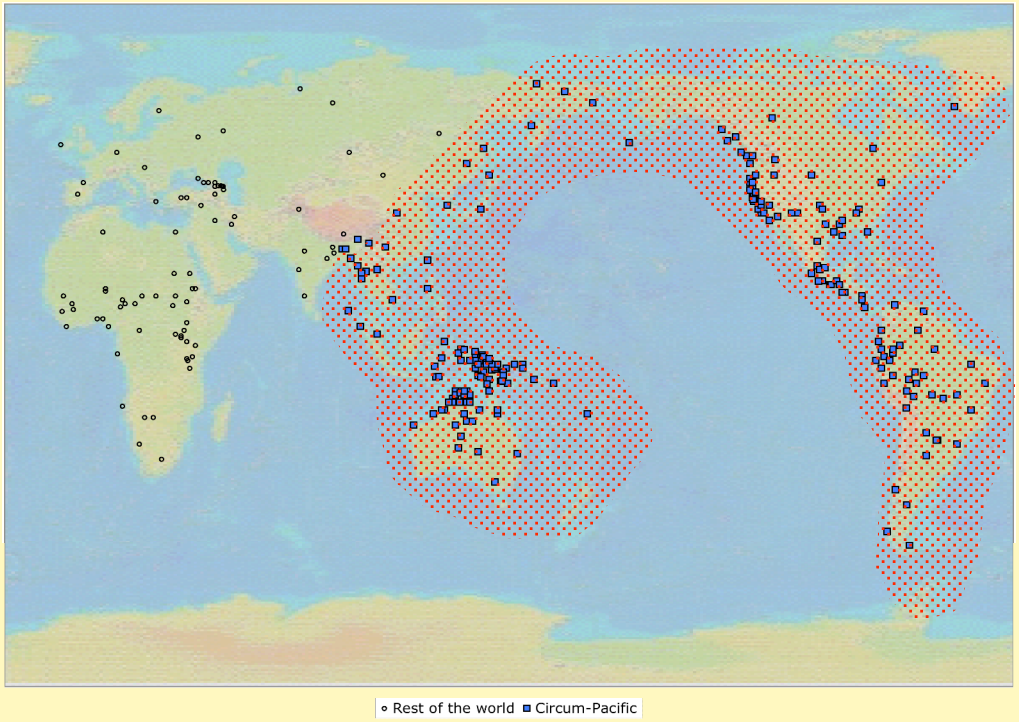
Scenario	*Distribution*	*Explanation*
De-Skewing Enclave	The surrounding macroarea deviates from the universal norm; the enclave follows the universal norm.	Relative isolation allowed the enclave to revert to the universal default, unhampered by areal skewing pressure
Rarity Enclave	The enclave contains a singularity or rarity found nowhere or almost nowhere else.	Relative isolation allowed undisturbed development and short-term preservation of (possibly unstable) rarities
Preservation Enclave	The enclave reflects a distribution that is different from the surrrounding macro-area but historically connected to it.	Relative isolation allowed preservation of a profile that characterized the surrounding macroarea before the population(s) of this macroarea migrated to a macroarea with which the enclave now shares the profile.

Testing enclave scenarios



Testing enclave scenarios

Test 1 - possibly related area: Circumpacific (CP)



	1a CP areality	1b Encl/CP
SYNTH	U=3070.5** F=10.5**	U=1104.5^{ns} F<1^{ns}
POLYAGR	χ^2=18.03**	χ^2<.01^{ns}
C/D	same	n/a
BIPARTIT	χ^2<.01^{ns}	n/a
2MARK	χ^2<.01^{ns}	n/a
MPOSSCL	χ^2=2.9^{ns}	n/a

** p<.001, * p<.05, ^{ns} p > .2, N (rnd) = 10,000

- SYNTH, POLYAGR have the same distribution in the enclaves and the CP
- Preservation enclave (reliable finding, T_{add}>4; T_{sd}=1.2)

Testing enclave scenarios

De-Skewing vs. Rarity scenarios (Tests 2-4)

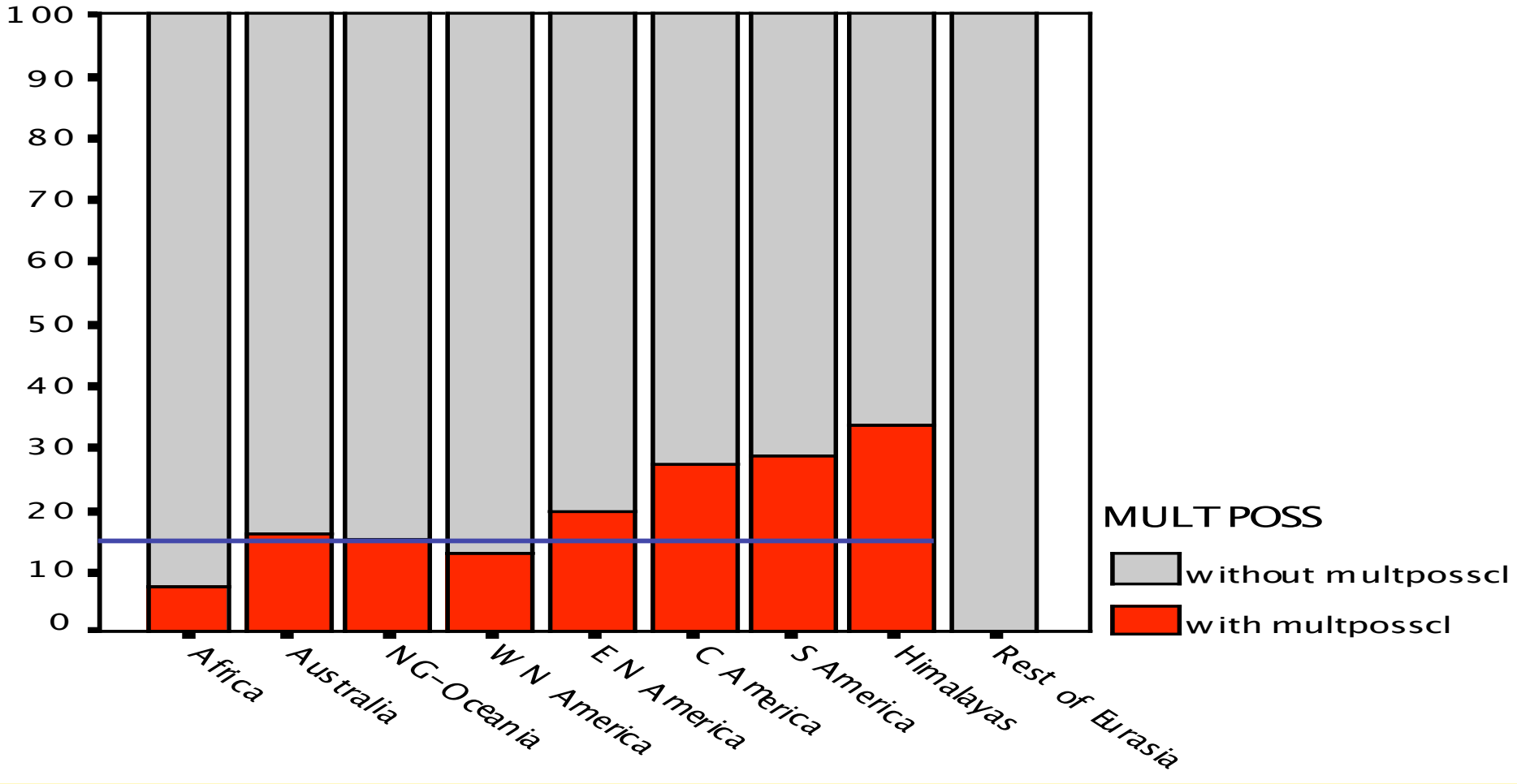
	2 Eurasia/ rest	3 Noneurasian areality	4 Enclave/ Noneurasia	∴ best-fitting scenario
C/D	same	none	n/a	**Rarity Enclave**
BIPARTIT	χ^2=.16$^{\text{n.s.}}$	**χ^2=13.4***	n/a	**Independent areality**
2MARK	χ^2<.01$^{\text{n.s.}}$	χ^2=1.9$^{\text{ns}}$	n/a	**Rarity Enclave**
MPOSSCL	**χ^2=6.42***	χ^2=5.8$^{\text{ns}}$	**χ^2<.01$^{\text{ns}}$**	**De-Skewing Enclave**

** p<.001, * p<.05, $^{\text{ns}}$ p > .2, N (rnd) = 10,000

- C/D, BIPARTIT, 2MARK enclaves do not seem to relate to any area
- Likely **Rarity enclaves** (but enclavehood T<2)

- MPOSSCL has the same distribution in the enclaves and outside Eurasia
- Likely **De-Skewing enclave** (but enclavehood T<2)

The De-Skewing Enclave Scenario

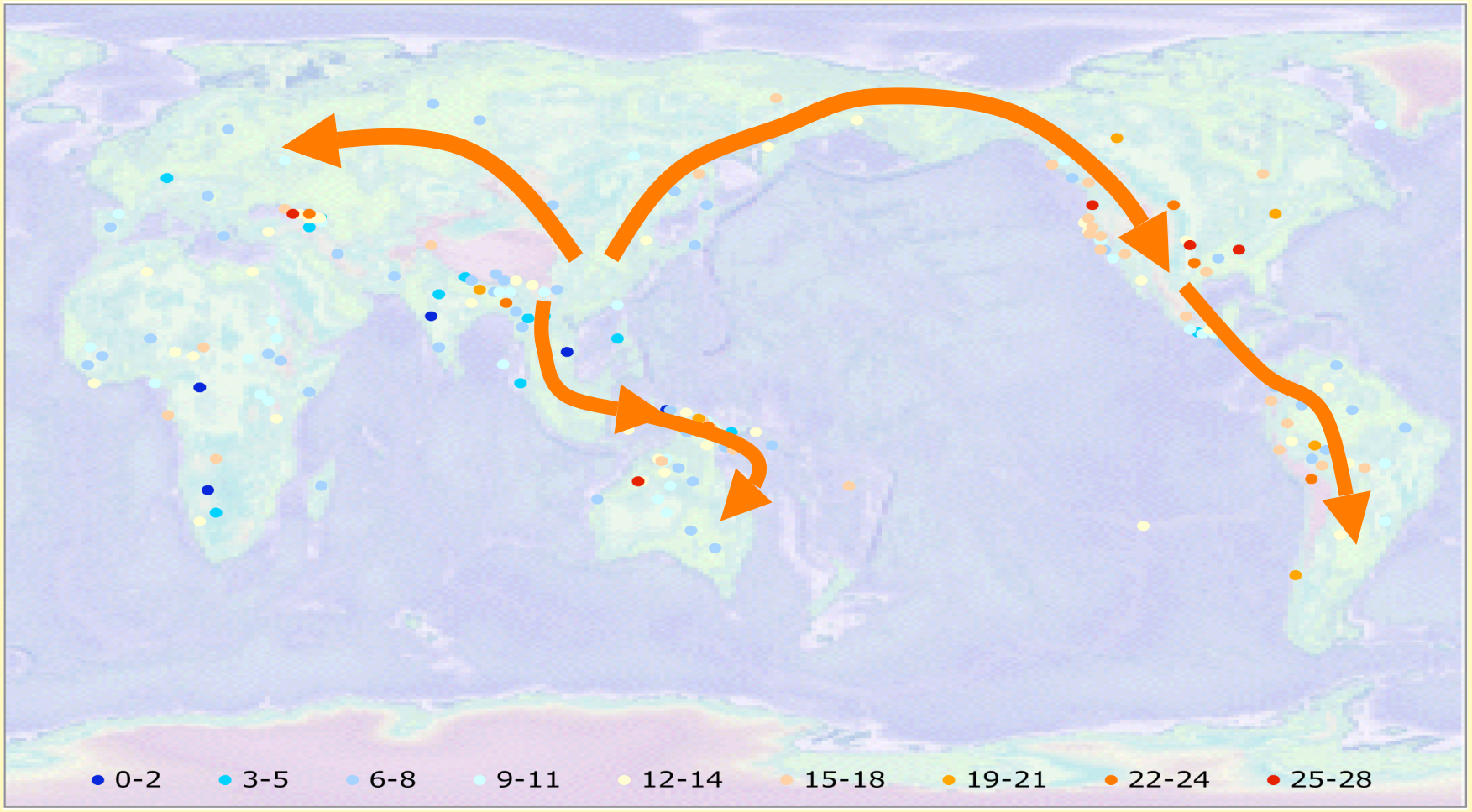


Himalayas vs. non-Eurasia: χ^2<.01^{ns}, μ = 16%

Himalayas vs. rest of Eurasia: χ^2=7.83*

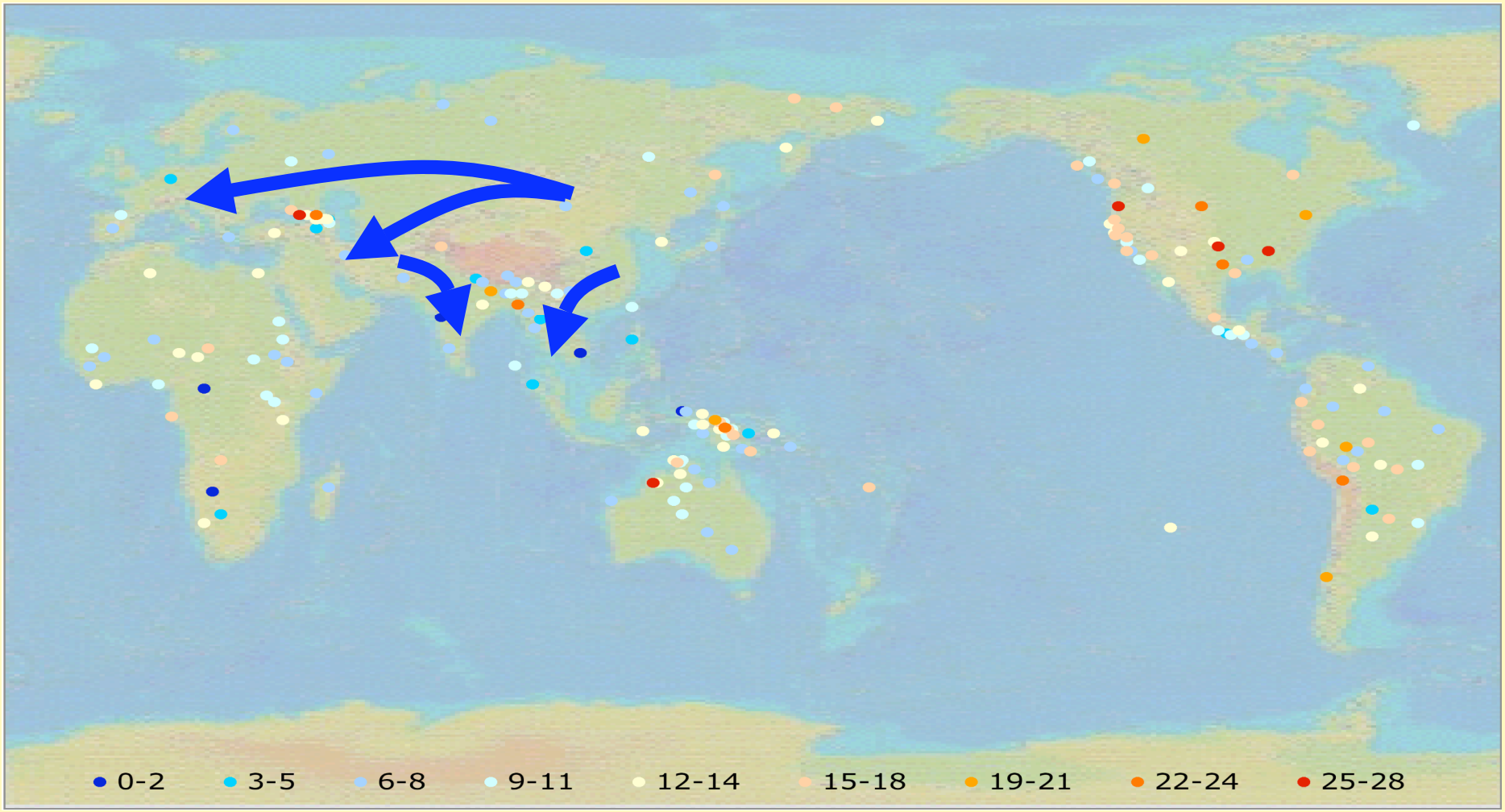
The Preservation Enclave Scenario

Enclaves reflect what used to be the Eurasian standard before the Eurasian Steppe and SEA spreads and before the colonializations of the Americas.



The Preservation Enclave Scenario

Then, the Eurasian Steppe (Greater Silk Road) and SEA spreads revert these profiles, *except in the enclaves*.



Conclusions

- The Eurasian macroarea is confirmed for two new variables (SYNTH/POLYAGR, MULT_POSS_CL).
- Himalayan and Caucasian languages are special because they deviate in several respects from the surrounding Eurasian typological profile.
- One of these deviations (SYNTH/POLYAGR) is best explained by the fact that the Himalayas and Caucasus are accretion zones that are originally connected (either by descent or contact) to the same circumpacific populations that colonialized the Americas.
- One deviation (MULT_POSS_CL) can perhaps be explained as a De-Skewing effect.
- Other deviations are probably due to independent arealities (BIPART) and/or rarities (2MARK, C/D).

Credits

- **The AUTOTYP research team** (as of September 2003)
  - Johanna Nichols (Co-Director, Berkeley)
  - Balthasar Bickel (Co-Director, Leipzig)
  - Tracy Alan Hall (Associate researcher, Leipzig)
  - Fernando Zúñiga (Post-Doc, Leipzig)
  - Kristine Hildebrandt (Post-Doc, Leipzig)
  - *RAs in Berkeley:* Gabriela Caballero, Suzanne Wilhite
  - *RAs in Leipzig:* Michael Riessler, Alena Witzlack-Makarevich, Sven Siegmund, Sindy Poppitz, Franziska Crell, Kathi Stutz
  - *Past team members:* Sandra Biewald, Aimee Lahaussois-Bartosik, Dave Peterson, Rebecca Voll, Keith Sanders

- **Funding:** Swiss NSF Grant Nos. 08210-053455 and 610-0627 (Bickel), German DFG Grant No. BI 799/2-1 (Bickel & Hall), US NSF Grant No. 96-16448 (Nichols)