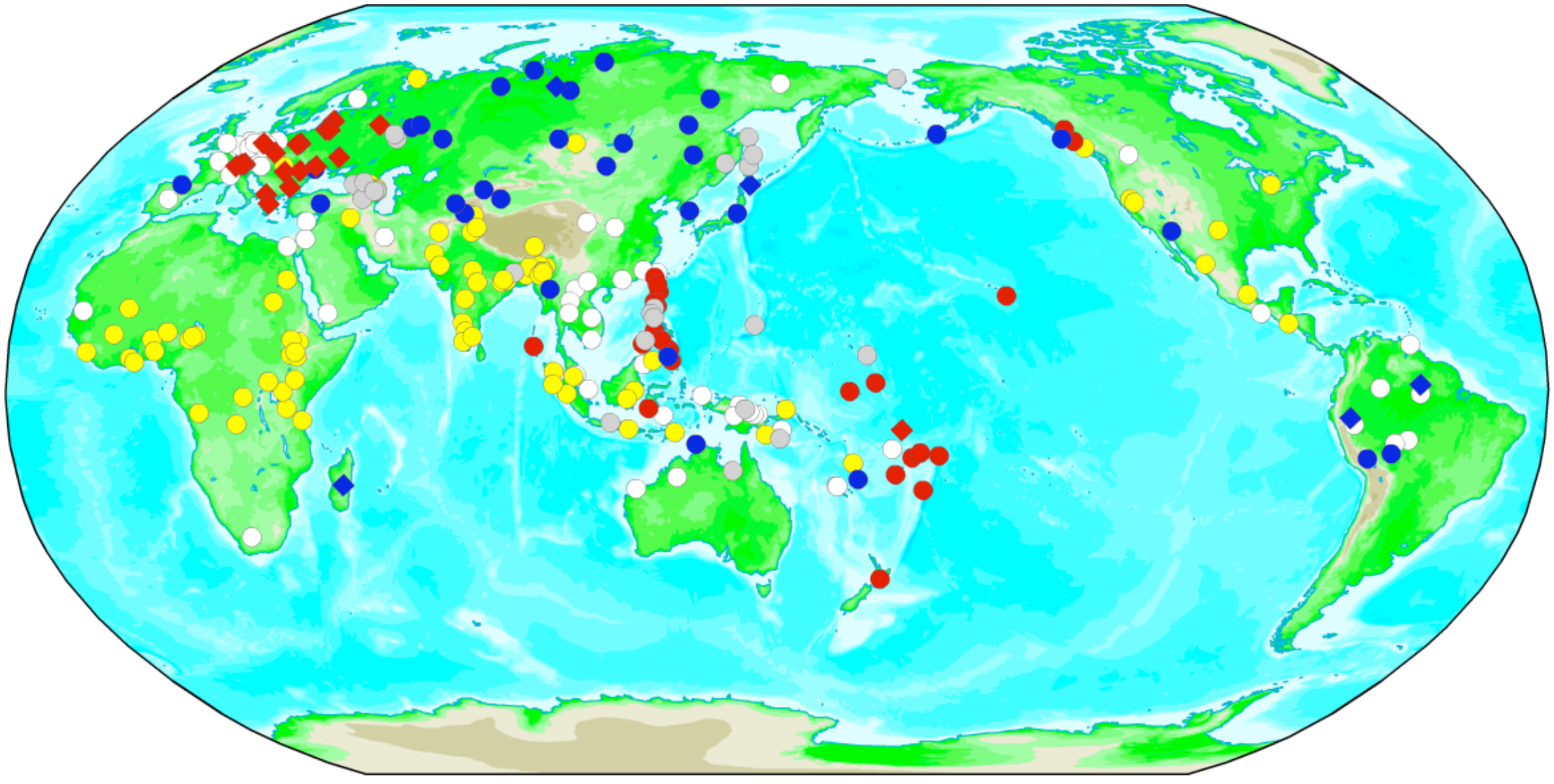
Preparing WALS for quantitative analyses

Michael Cysouw & Mihai Albu
Max Planck Institute for Evolutionary Anthropology

Problems in the WALS data:

- Types consisting of dissimilar languages
- Independent features combined in one map
- Dependencies between maps
- No relative similarities available

Map 54: Distributive numerals



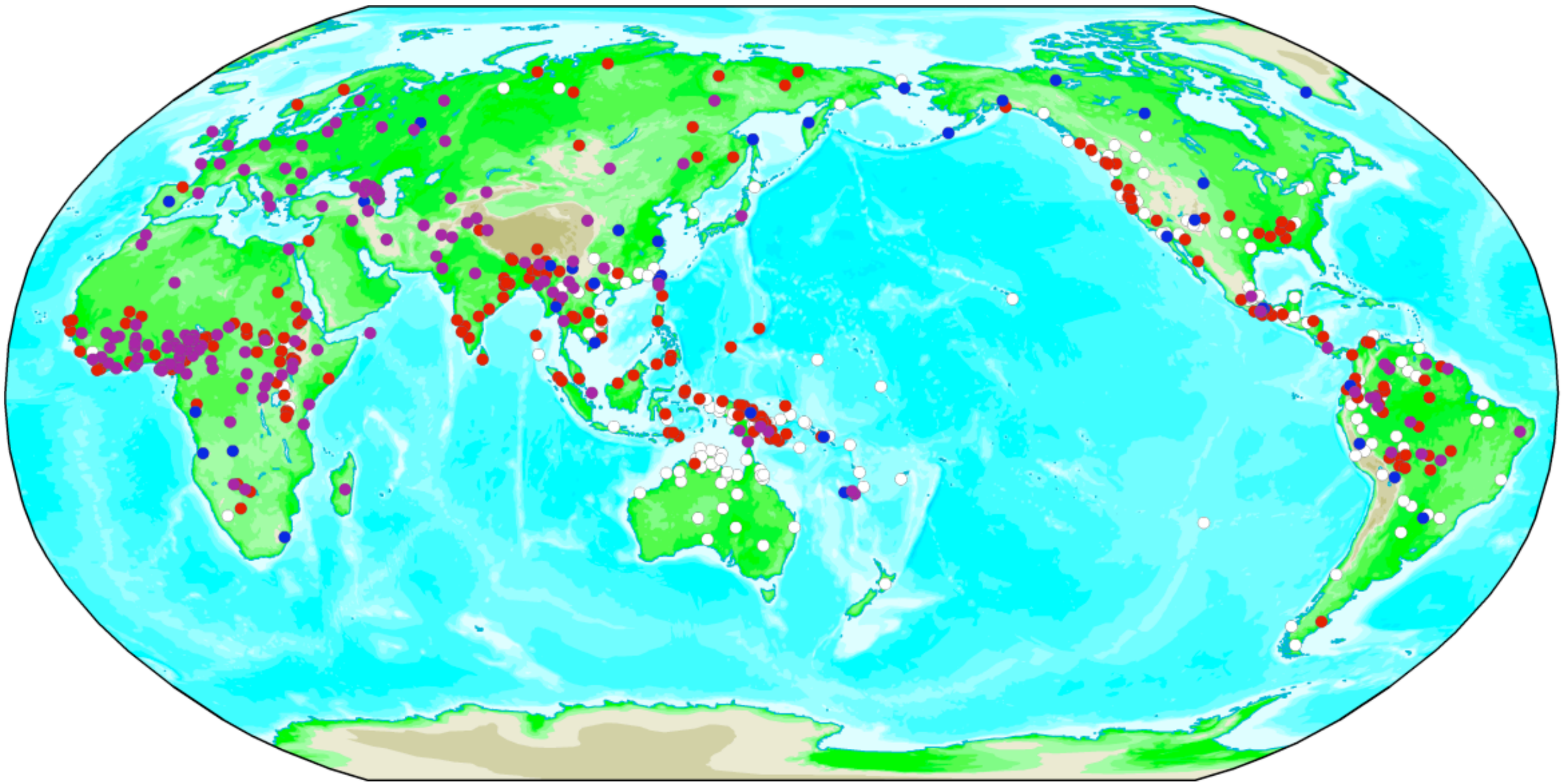
1. No distributive numerals [62]
2. Marked by reduplication [84]
3. Marked by prefix [23]
4. Marked by suffix [32]
5. Marked by preceding word [21]
6. Marked by following word [5]
7. Marked by mixed or other strategies [23]

Solution:
recode all as different types

Problems in the WALS data:

- Types consisting of dissimilar languages
- Independent features are depicted in one map
- Dependencies between maps
- No relative similarities available

Map 4: Voicing in Plosives and Fricatives



1. No voicing constrast [181]
2. In plosives alone [189]
3. In fricatives alone [38]
4. In both plosives and fricatives [158]

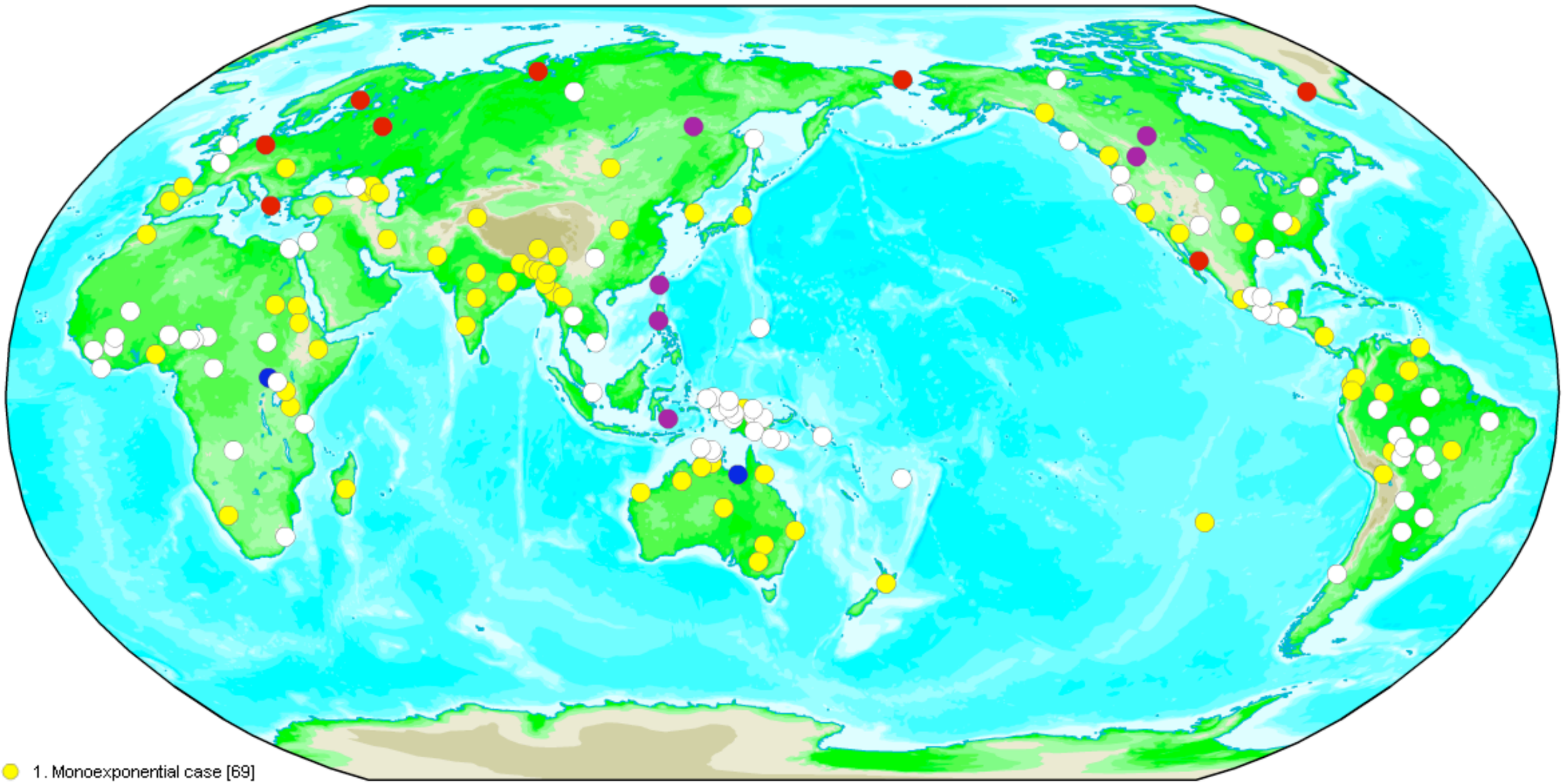
Solution:
split, and disregard the original

Problems in the WALS data:

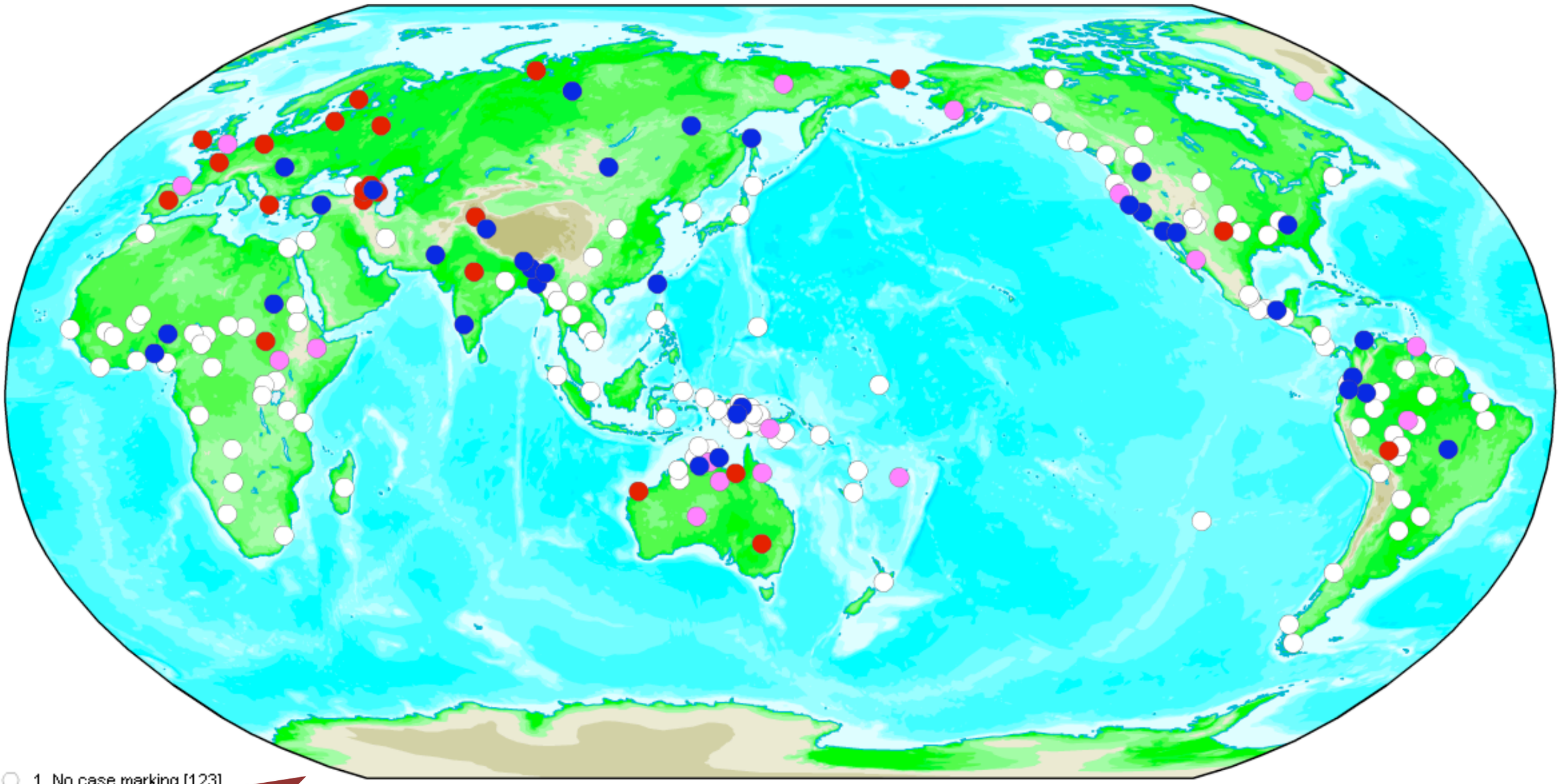
- Types consisting of dissimilar languages
- Independent features are depicted in one map
- Dependencies between maps
- No relative similarities available

Problem:
(Apparently) identical values

Map 21: Exponence of Selected Inflectional Formatives



Map 28: Case Syncretism



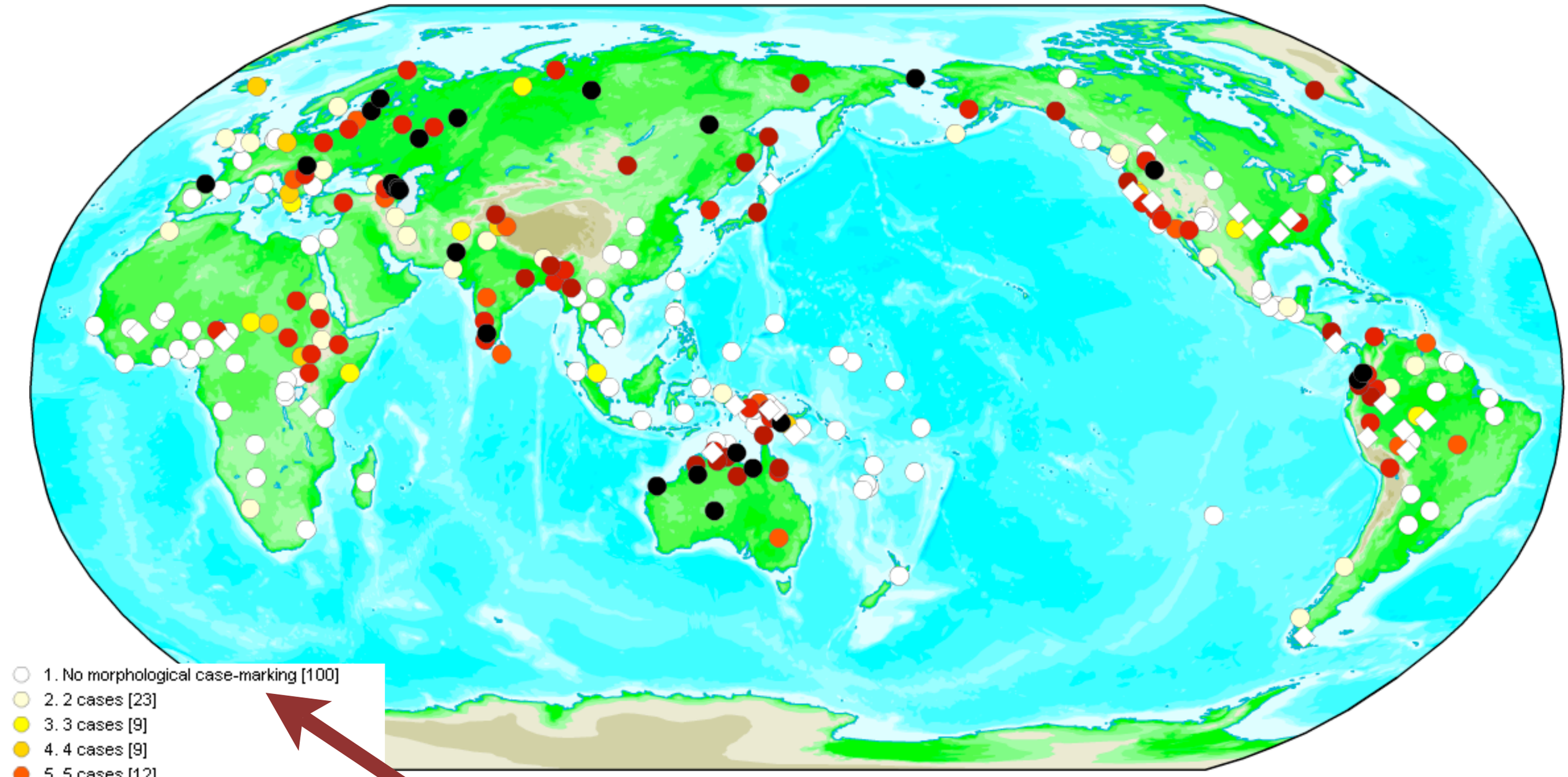
1. No case marking [123]
2. Core cases only [18]
3. Core and non-core [22]
4. No syncretism [34]

Different definitions and interpretation

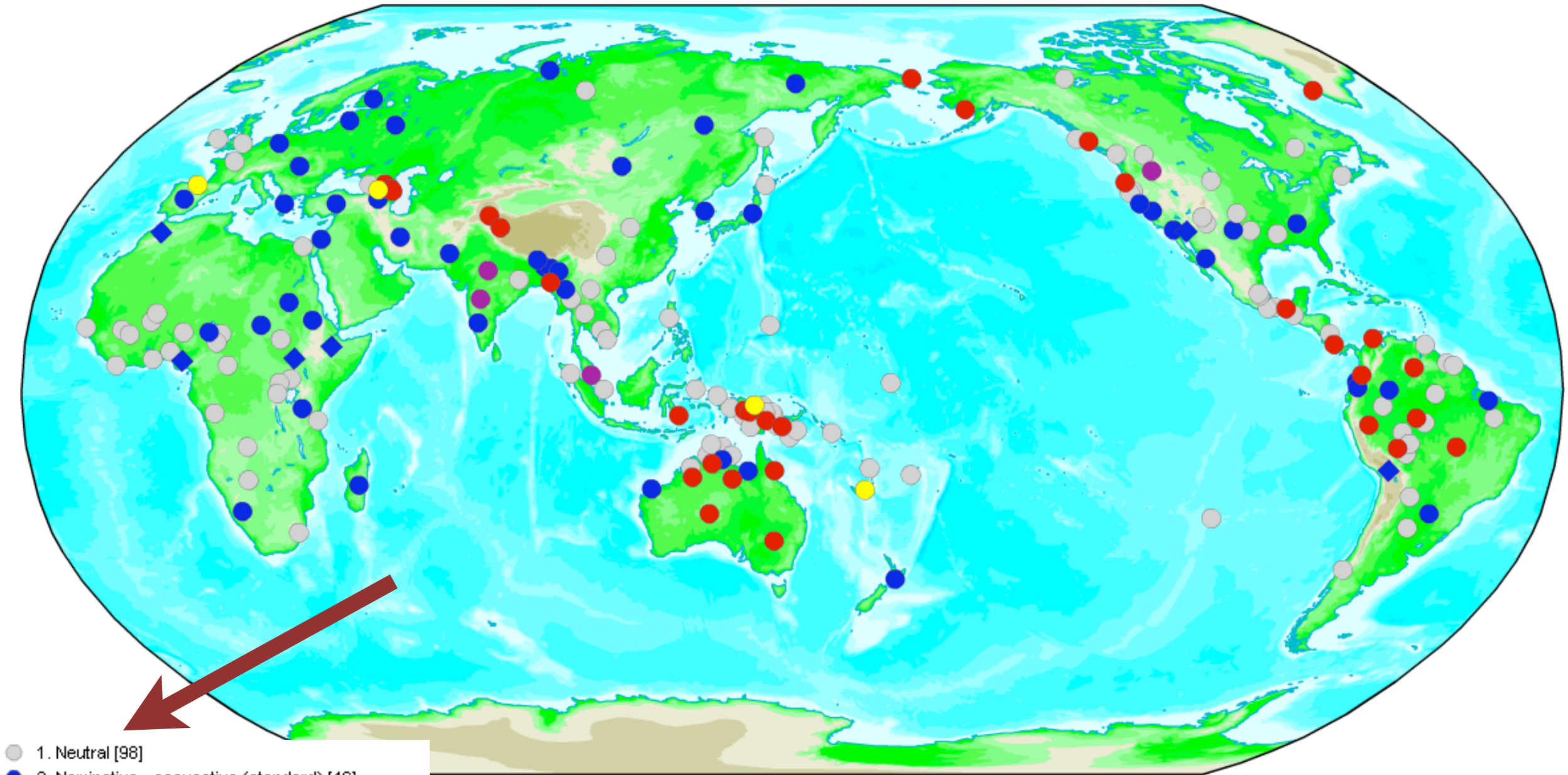
	no case (map 28)	case (map 28)
no case (map 21)	48	8
case (map 21)	28	44

Problem:
Covert Dependencies

Map 49: Number of Cases



Map 98: Alignment of case marking of full noun phrases



Marking of full noun phrases

	neutral alignment	non-neutral alignment
no case distinctions	62	12
case distinctions	17	77

Possible solutions:

Make sets of interdependent features

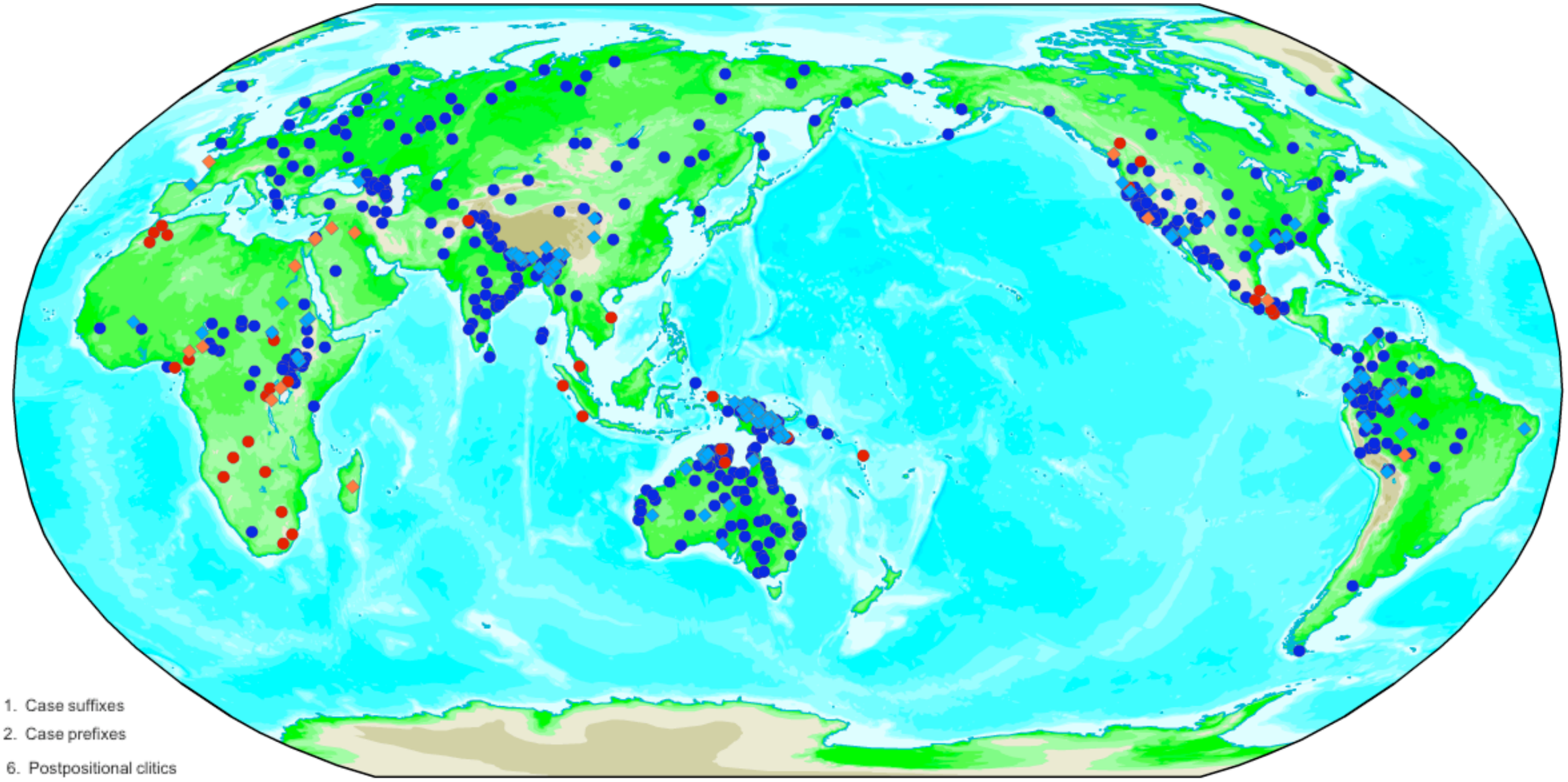
a) Choose maximally one out of each set

b) Combine dependent features into one larger feature with very many values

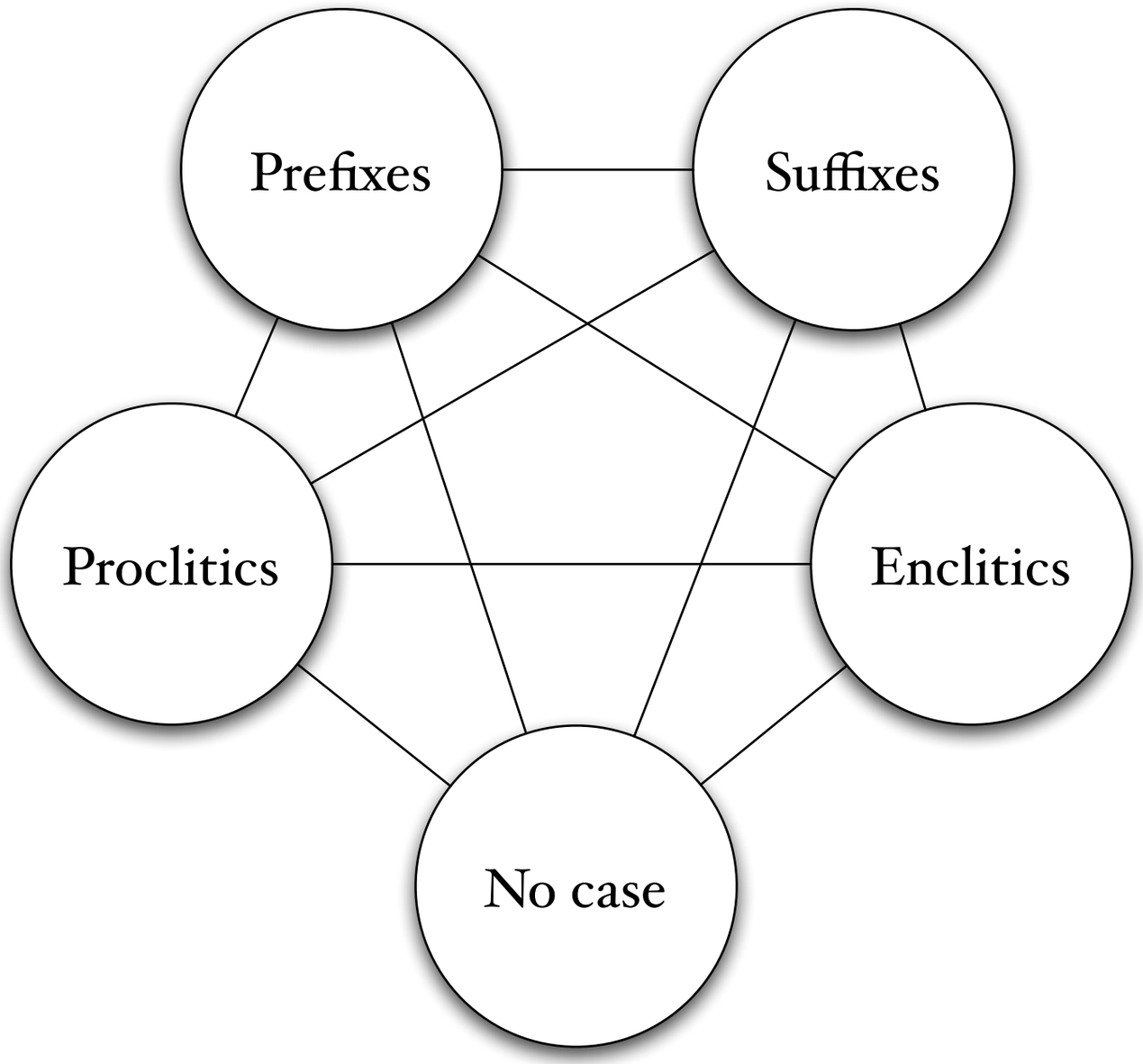
Problems in the WALS data:

- Types consisting of dissimilar languages
- Independent features are depicted in one map
- Dependencies between maps
- No relative similarities available

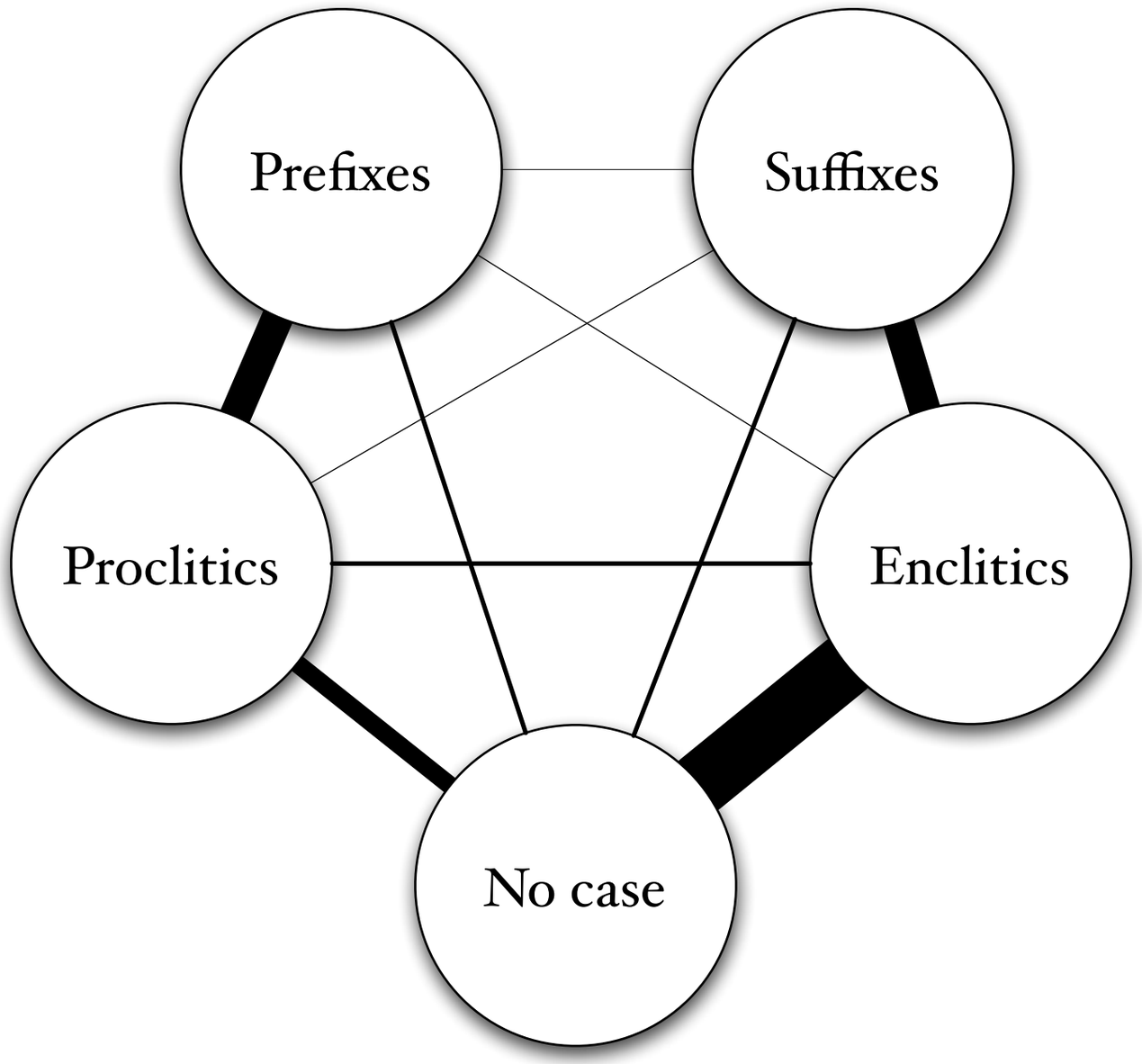
Map 51: Position of Case affixes (selection)



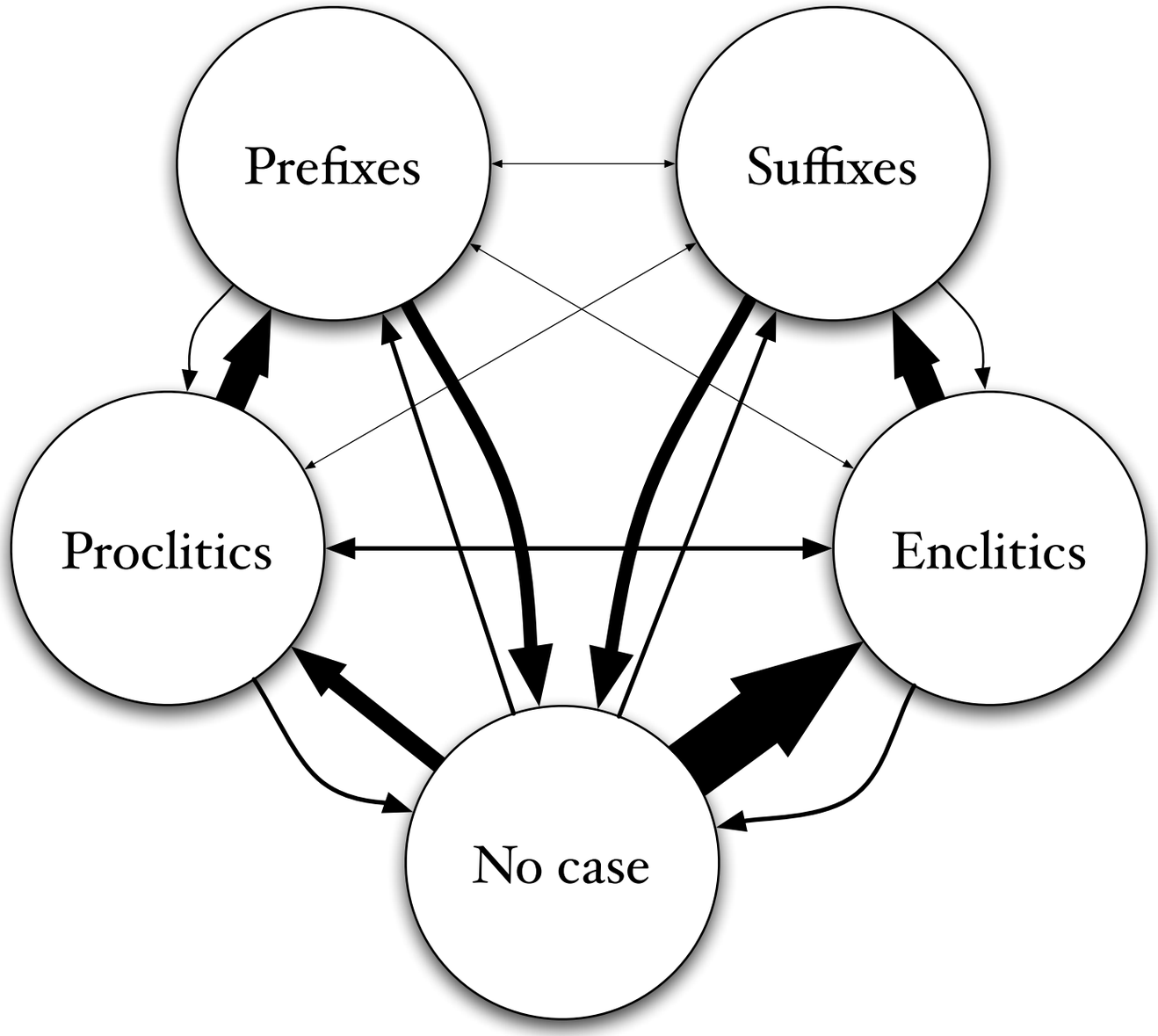
Undifferentiated typology



Similarities



Transitional probabilities



Solution:
add tables with such information

Transitional probabilities

(relative values; higher values are less likely)

From:

To:

	no case	proclitics	enclitics	prefixes	suffixes
no case	0	6	6	4	4
proclitics	3	0	5	8	10
enclitics	1	5	0	10	8
prefixes	6	2	10	0	10
suffixes	6	10	2	10	0

No symmetry
(6 ≠ 1, 8 ≠ 2, etc.)

From:

To:

	no case	proclitics	enclitics	prefixes	suffixes
no case	0	6	6	4	4
proclitics	3	0	5	8	10
enclitics	1	5	0	10	8
prefixes	6	2	10	0	10
suffixes	6	10	2	10	0

No mathematical transitivity
(3 + 5 ≠ 1)

From:

To:

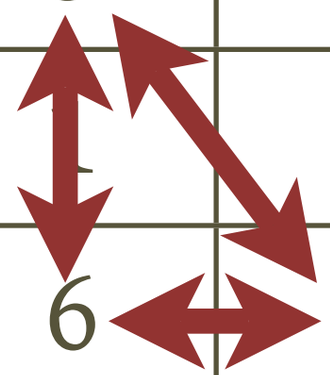
	no case	proclitics	enclitics	prefixes	suffixes
no case	0	6	6	4	4
proclitics	3	0	5	8	10
enclitics	1	5	0	10	8
prefixes	6	2	10	0	10
suffixes	6	10	2	10	0

No triangular inequality
(3 + 2 < 6)

From:

To:

	no case	proclitics	enclitics	prefixes	suffixes
no case	0	6	6	4	4
proclitics	3	0	5	8	10
enclitics	3	5	0	10	8
prefixes	6	2	10	0	10
suffixes	6	10	2	10	0



But:

- Application for such tables are still to be build
- We are working on that!
  - hierarchical clustering (‘trees’)
  - statistical testing
  - multidimensional scaling

The End