Correlations between typological features predict their geo-spatial patterning

Deepthi Gopal (Uppsala), **Henri Kauhanen** (Konstanz), **Christopher Kitching** (Manchester), **Tobias Galla** (Manchester), **Ricardo Bermúdez-Otero** (Manchester)

LAGB Annual Meeting 2022, 13 September

Roadmap

Background

Distance & distribution

The geo-spatial properties of linguistic features

Modelling distributions of individual features

Rates of change & stability

Our model (2021)

Correlations between features

Typological observations & word-order features

Hypotheses

Empirical testing

Our model (2022)

Outlook, etc.

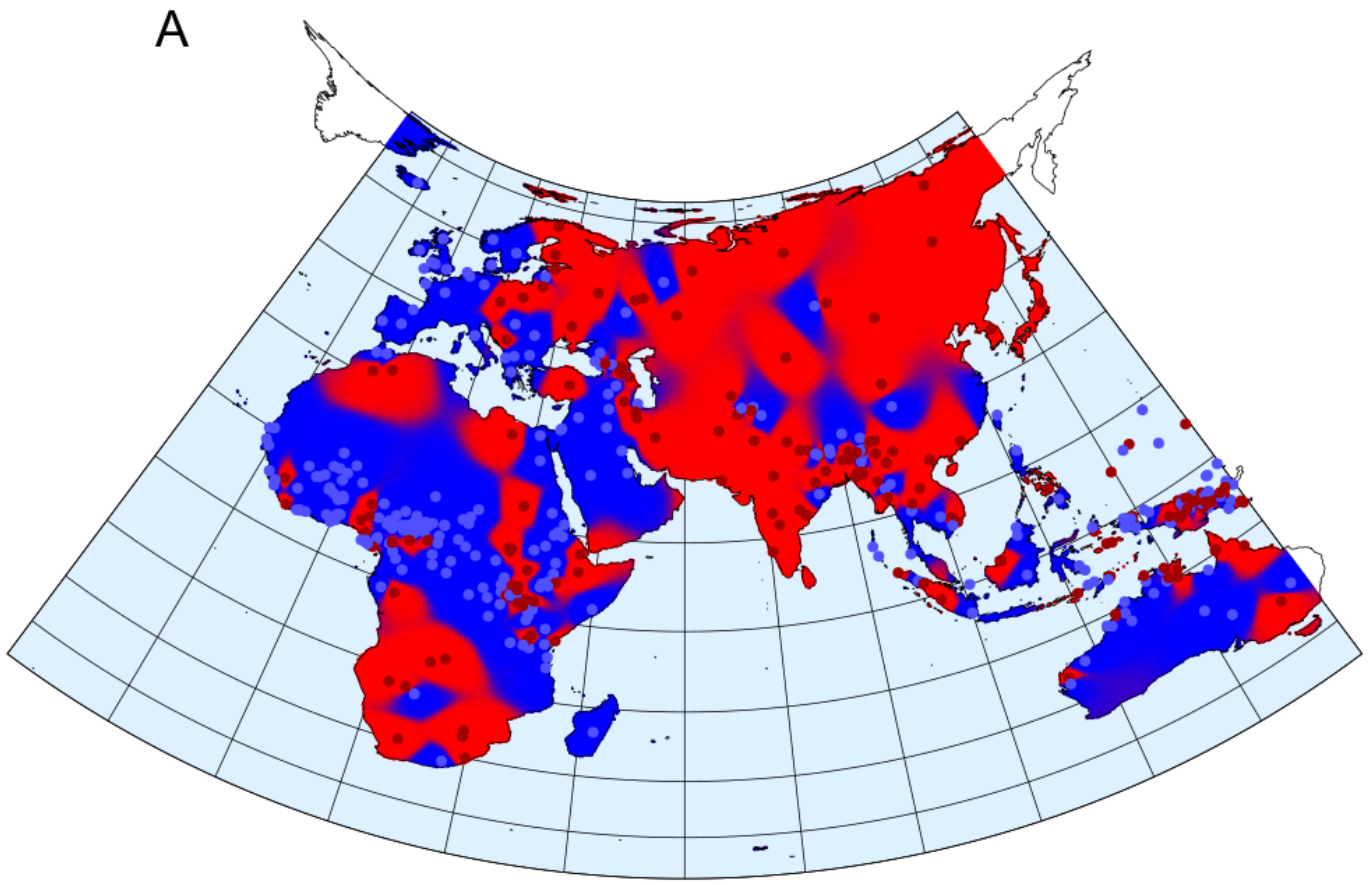
Background

Distance & distribution

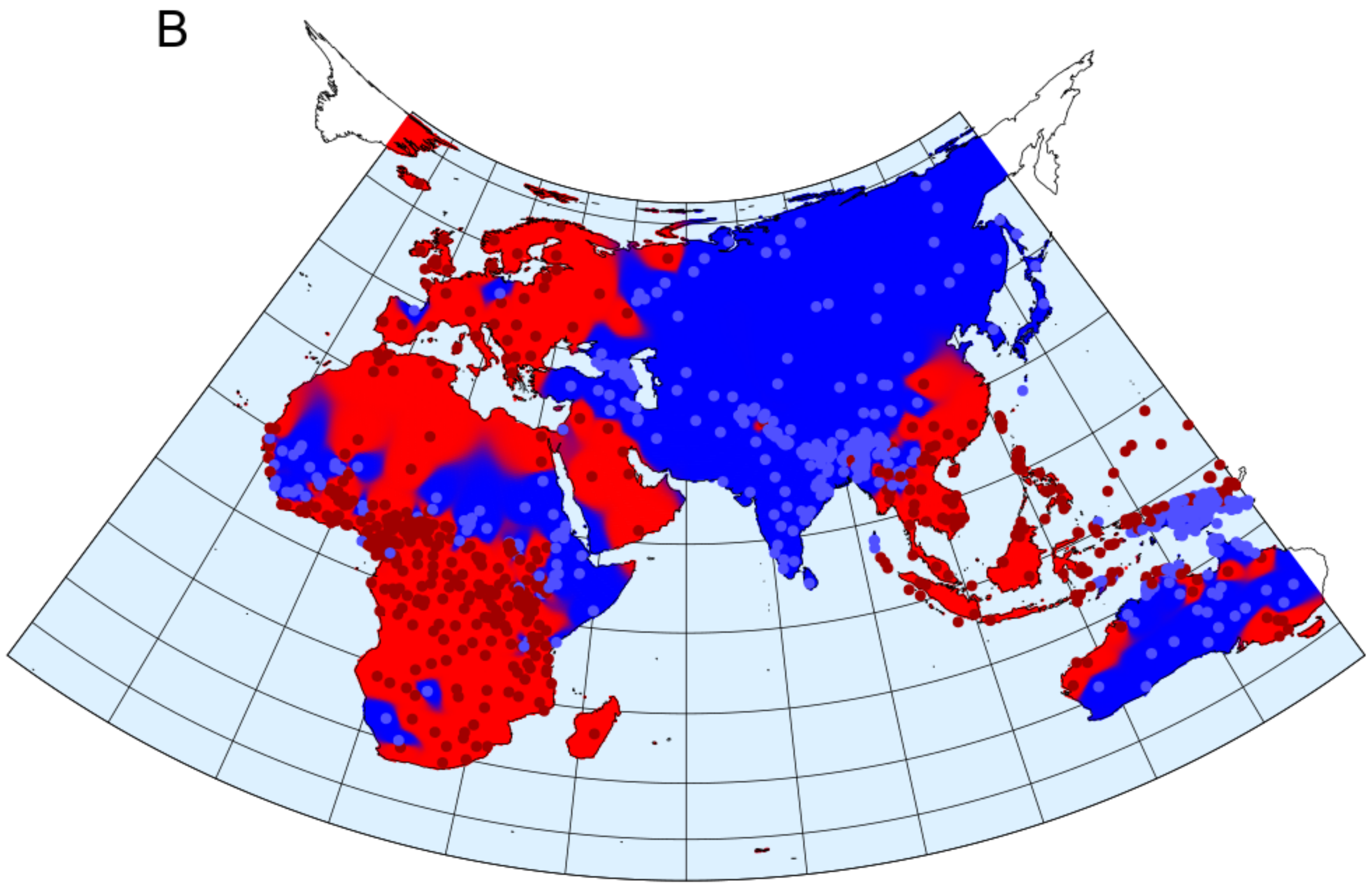
- **Question.** *Does the linguistic distance between A and B depend on the geographical distance between them?*
  - There exists an *intuition* that neighbours are likely to be similar, languages that are very far apart less so.
  - This intuition has two major underliers:
    - **Phylogeny** — neighbours are more likely than non-neighbours to share a common origin;
    - **Contact** — neighbours are more likely to converge to one another over time than non-neighbours.
    - Not trivial to separate out real-world results of one or the other.
  - If true, then true with respect to both *individual properties* and *sets of properties*.
    - **Any** distance metric we define requires the latter (is defined *over* some set of features …)
- Assume (!) that there is some kind of licit decomposition of *linguistic distance* into *features*. **Question.** *Does the proposed geographical dependence of ling. dist. care which features we're talking about?*
  - **viz.** *how different are different features in susceptibility to* [change]*?*

Background

The geo-spatial properties of linguistic features



definite article (**yes** **no**), WALS 37A

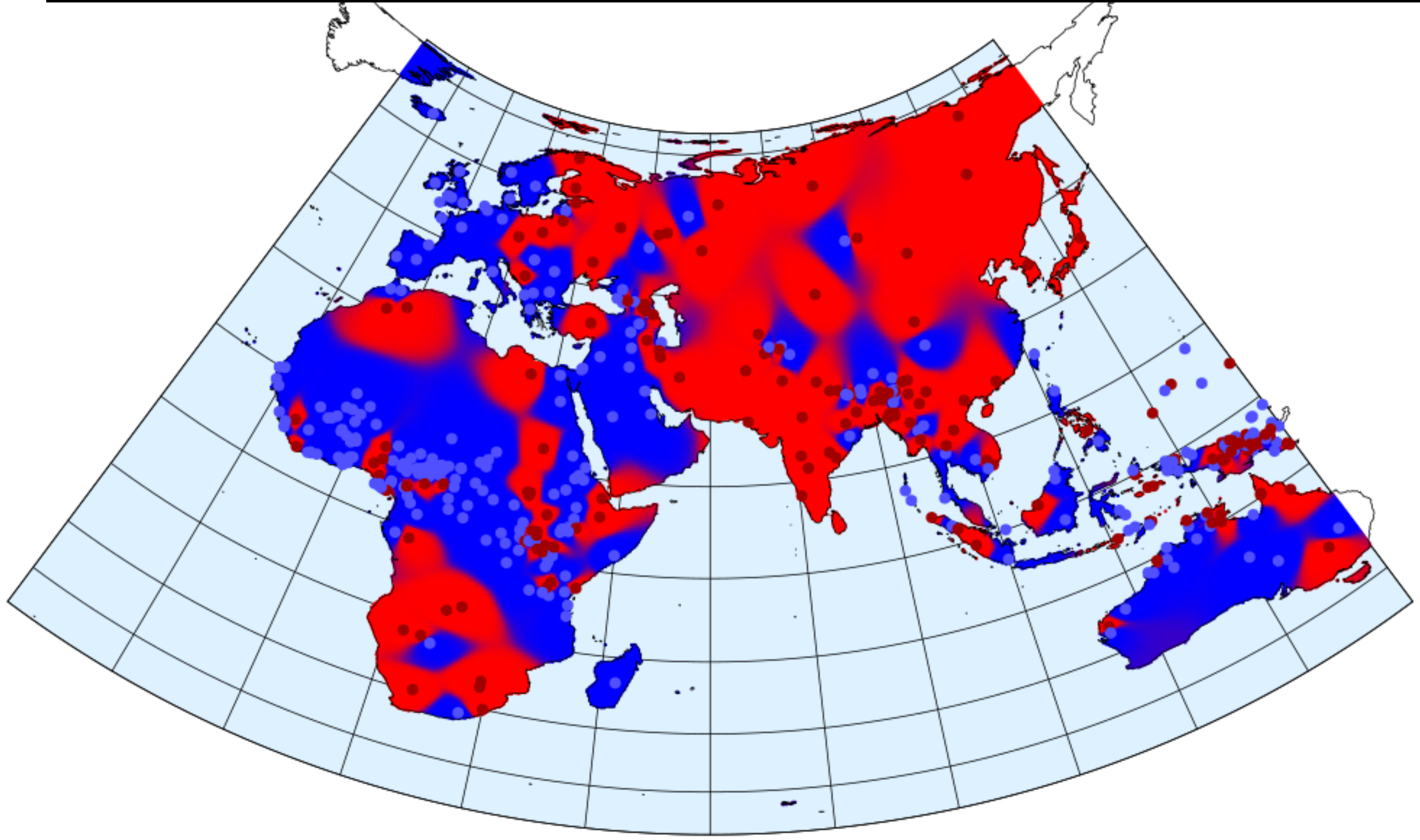


basic word order (**OV** **VO**), WALS 83A

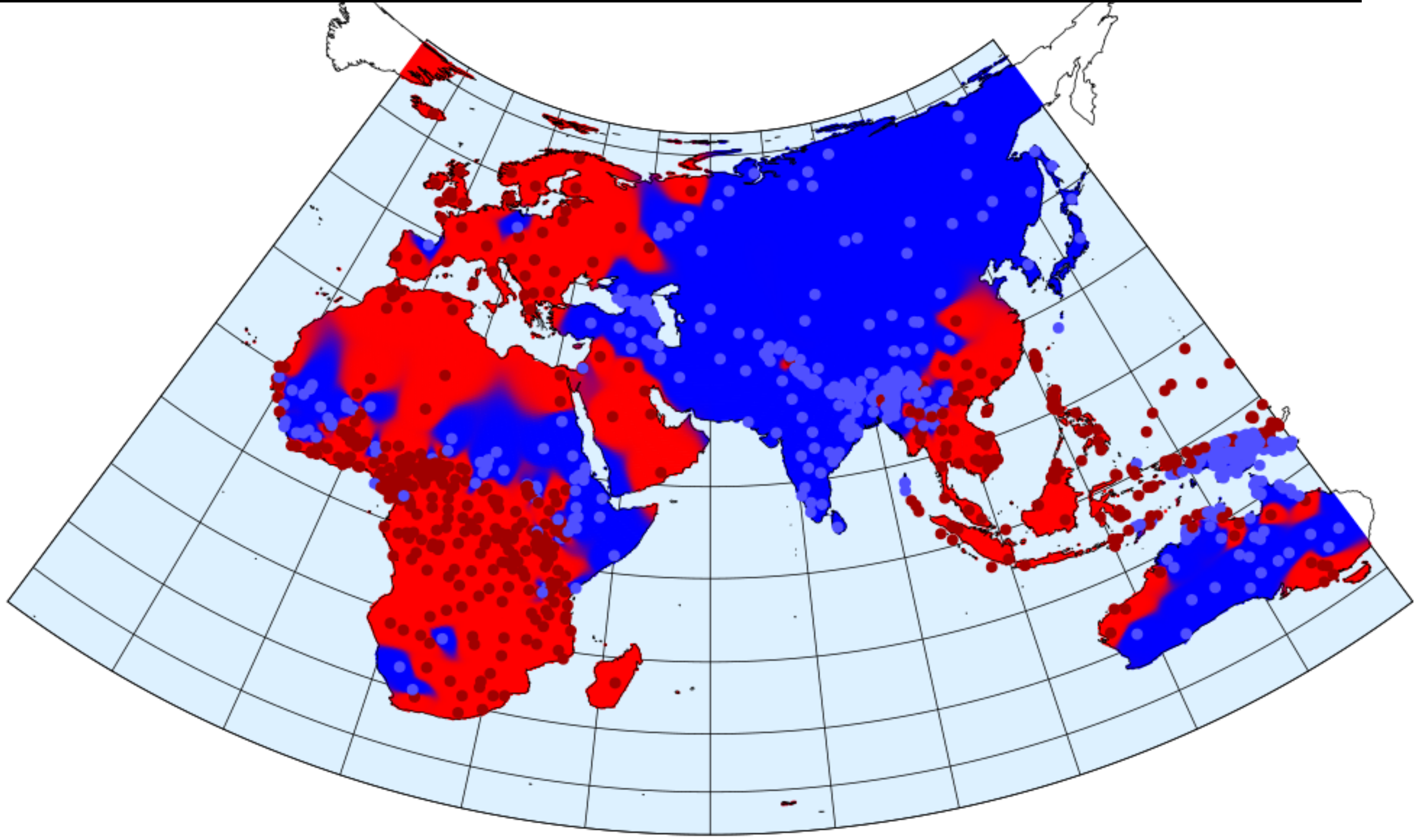
Background

The geo-spatial properties of linguistic features

Individual features show different kinds of spatial patterns. **Question.** *Why?*



definite article (**yes** **no**), WALS 37A



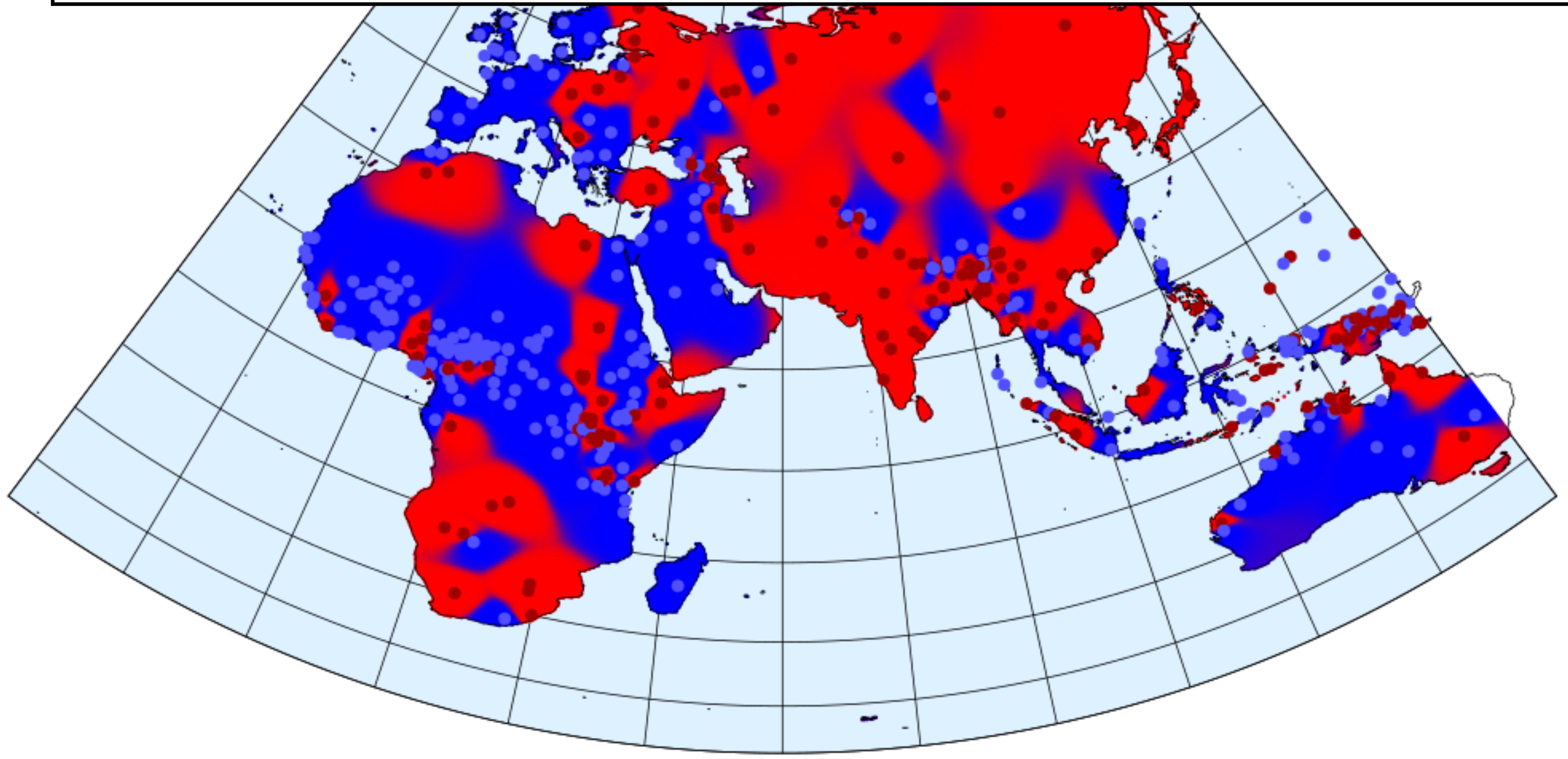
basic word order (**OV** **VO**), WALS 83A

Background

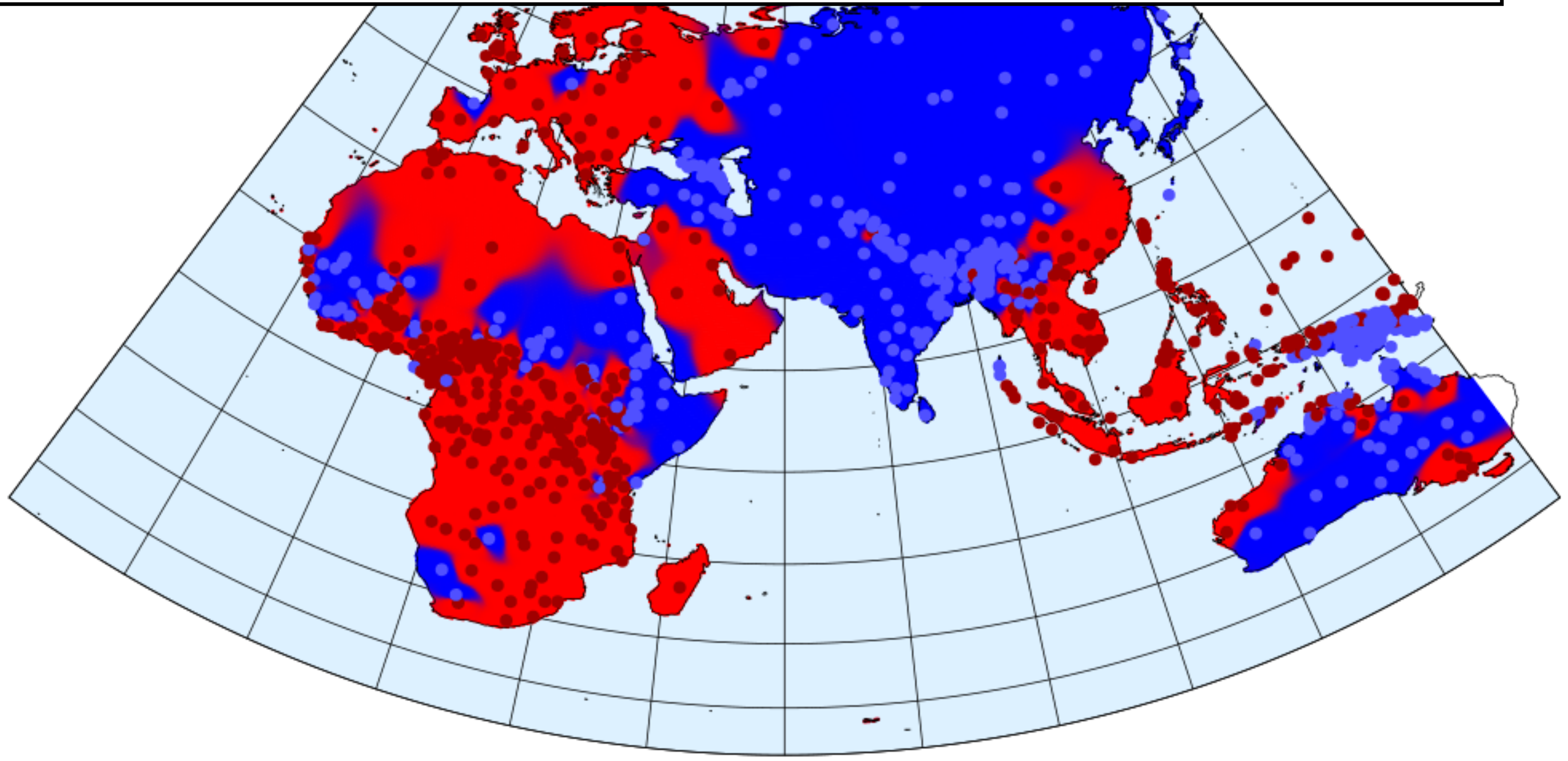
The geo-spatial properties of linguistic features

Individual features show different kinds of spatial patterns. **Question.** *Why?*

Two randomly-chosen neighbours A and B are **more likely to agree** on basic word order than on def. art.



definite article (**yes** **no**), WALS 37A



basic word order (**OV** **VO**), WALS 83A

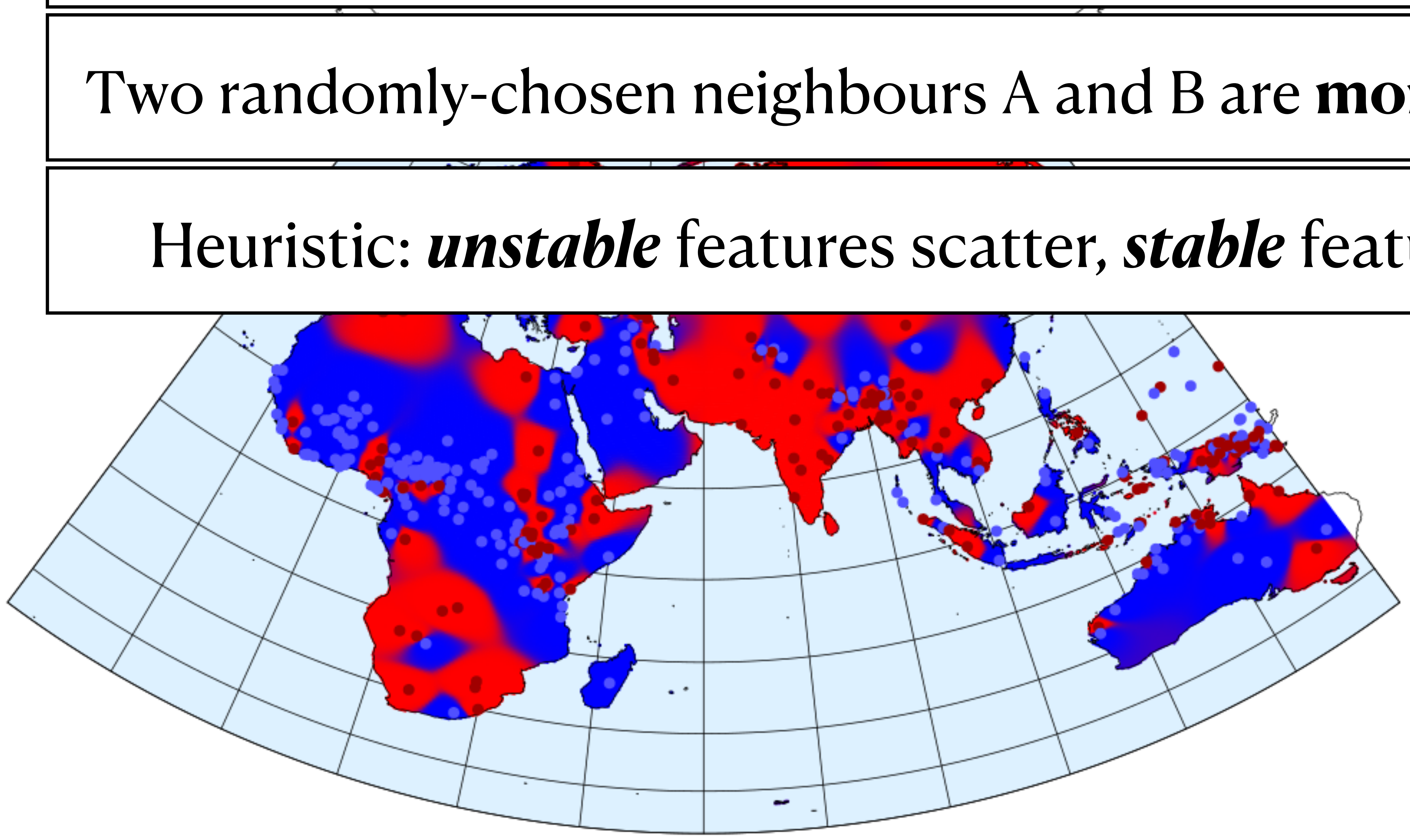
Background

The geo-spatial properties of linguistic features

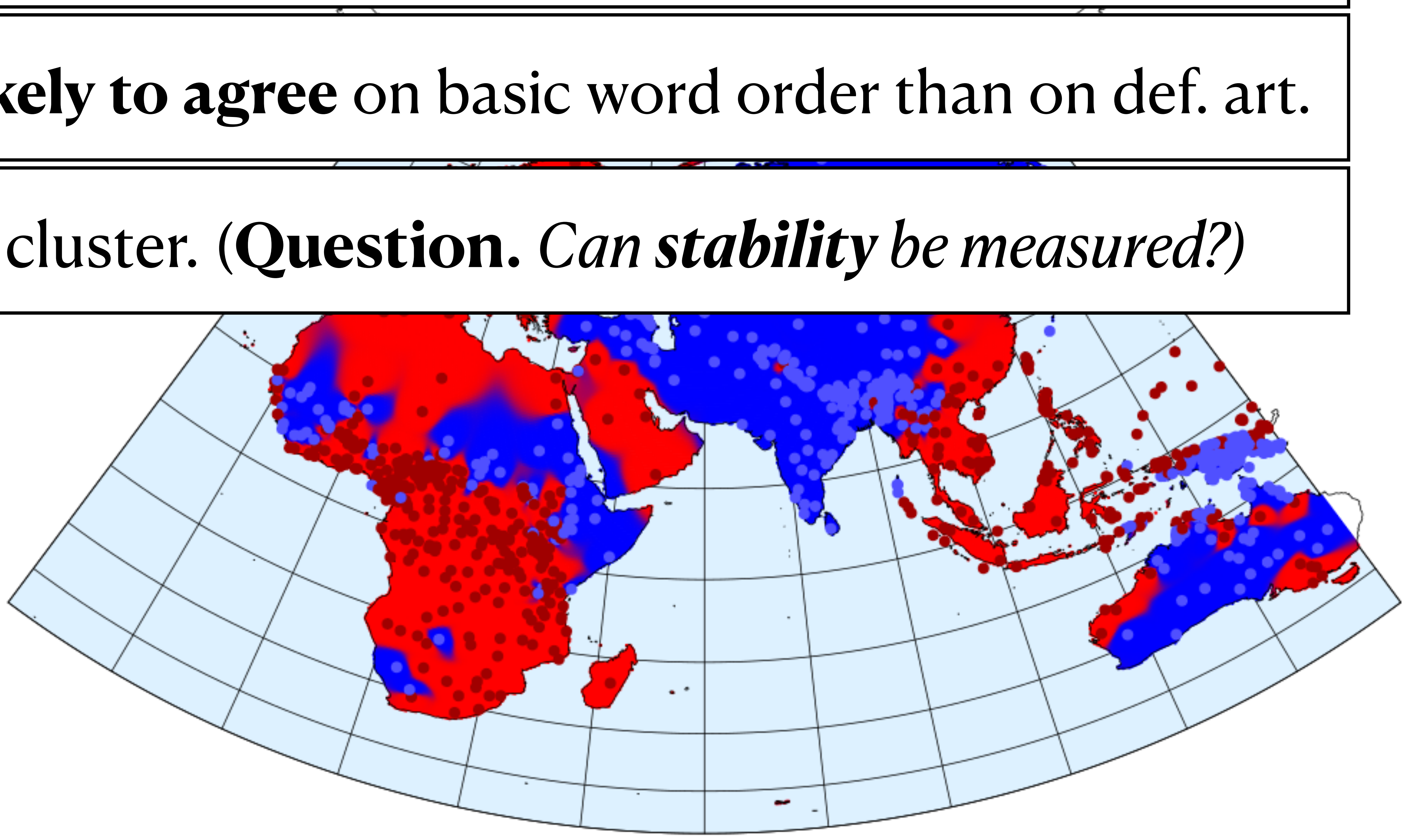
Individual features show different kinds of spatial patterns. **Question.** *Why?*

Two randomly-chosen neighbours A and B are **more likely to agree** on basic word order than on def. art.

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. (**Question.** *Can **stability** be measured?*)



definite article (**yes** **no**), WALS 37A



basic word order (**OV** **VO**), WALS 83A

Background

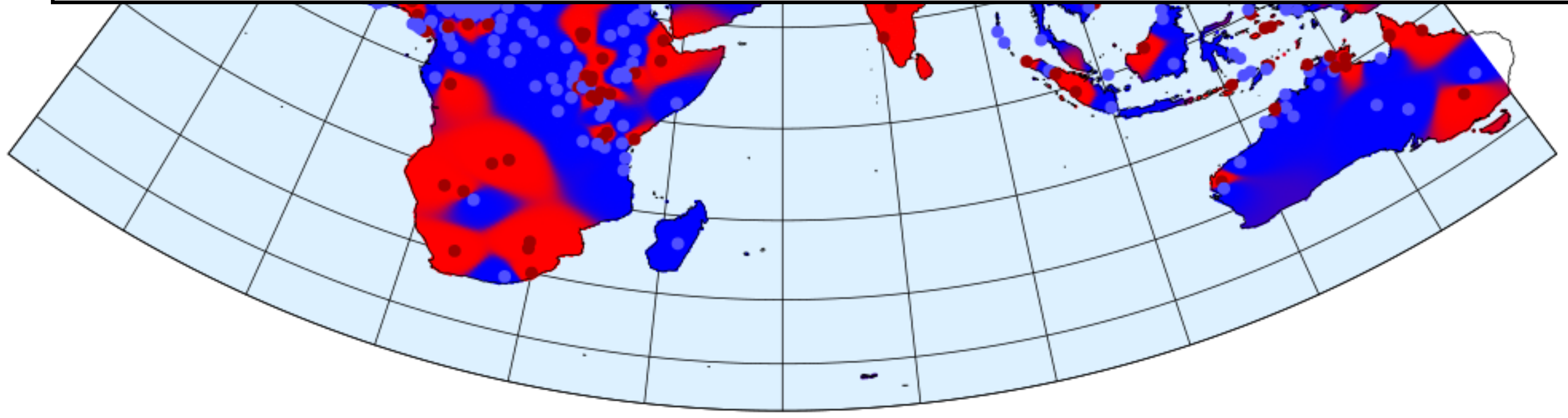
The geo-spatial properties of linguistic features

Individual features show different kinds of spatial patterns. **Question.** *Why?*

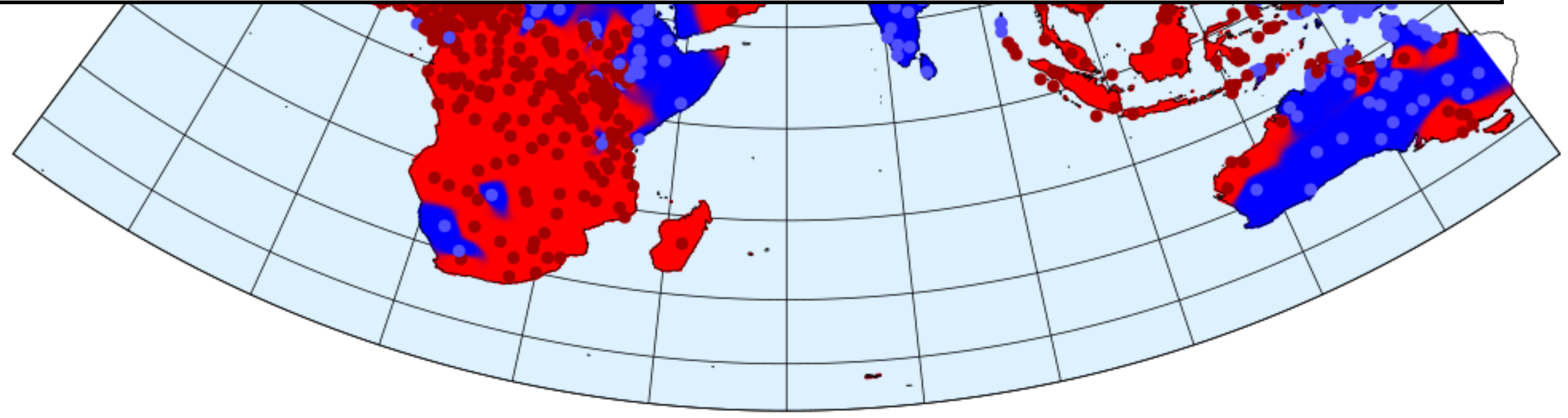
Two randomly-chosen neighbours A and B are **more likely to agree** on basic word order than on def. art.

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. (**Question.** *Can **stability** be measured?*)

Features don’t exist in isolation. **Question.** *How do we account for this non-independence?*



definite article (**yes** **no**), WALS 37A



basic word order (**OV** **VO**), WALS 83A

Background

The geo-spatial properties of linguistic features

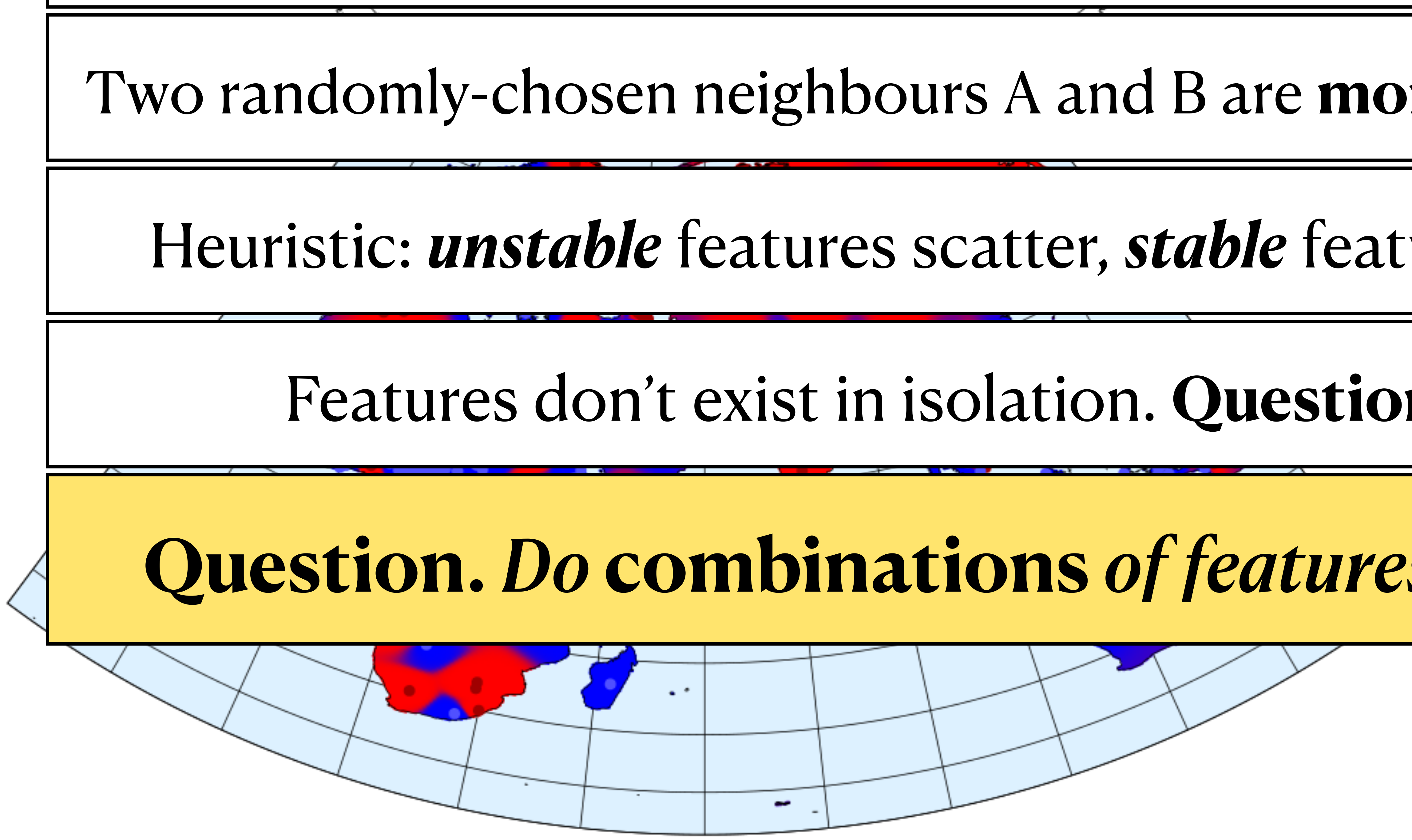
Individual features show different kinds of spatial patterns. **Question.** *Why?*

Two randomly-chosen neighbours A and B are **more likely to agree** on basic word order than on def. art.

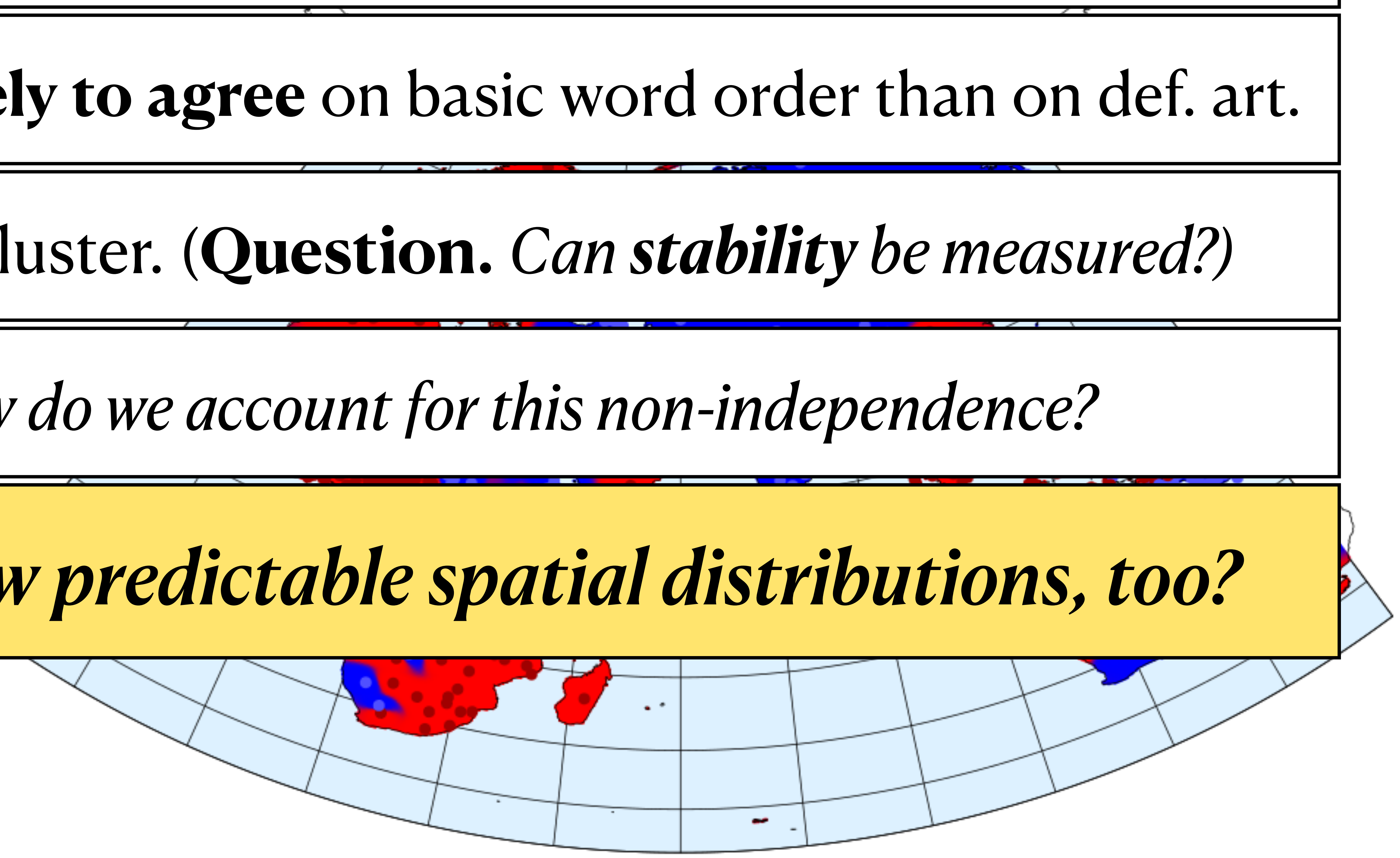
Heuristic: ***unstable*** features scatter, ***stable*** features cluster. (**Question.** *Can **stability** be measured?*)

Features don't exist in isolation. **Question.** *How do we account for this non-independence?*

Question. ***Do combinations of features show predictable spatial distributions, too?***



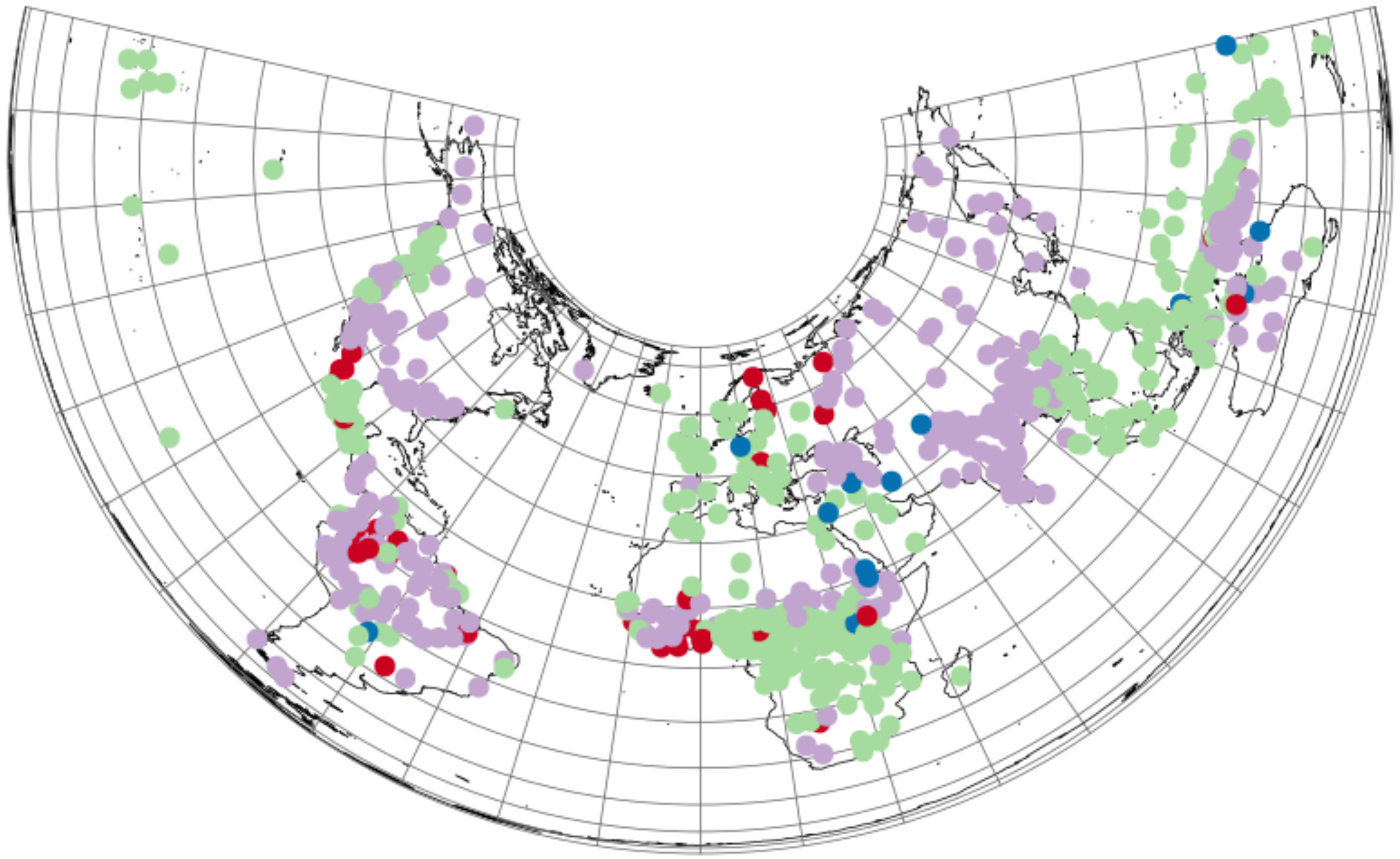
definite article (yes no), WALS 37A



basic word order (OV VO), WALS 83A

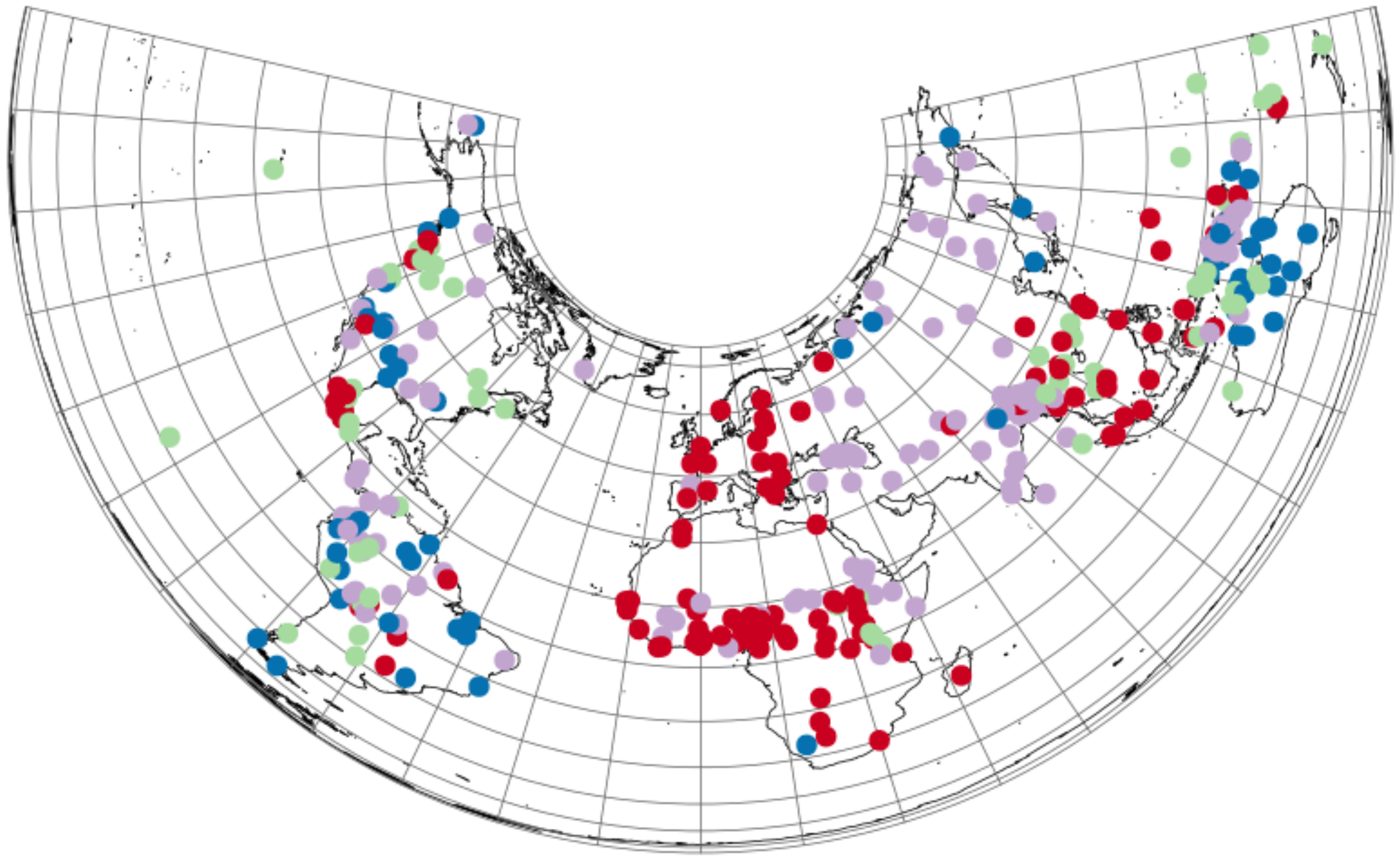
Background

The geo-spatial properties of linguistic features



basic word order (WALS 83A) x adposition order (WALS 85A)

VO, postpositions
VO, prepositions
OV, postpositions
OV, prepositions



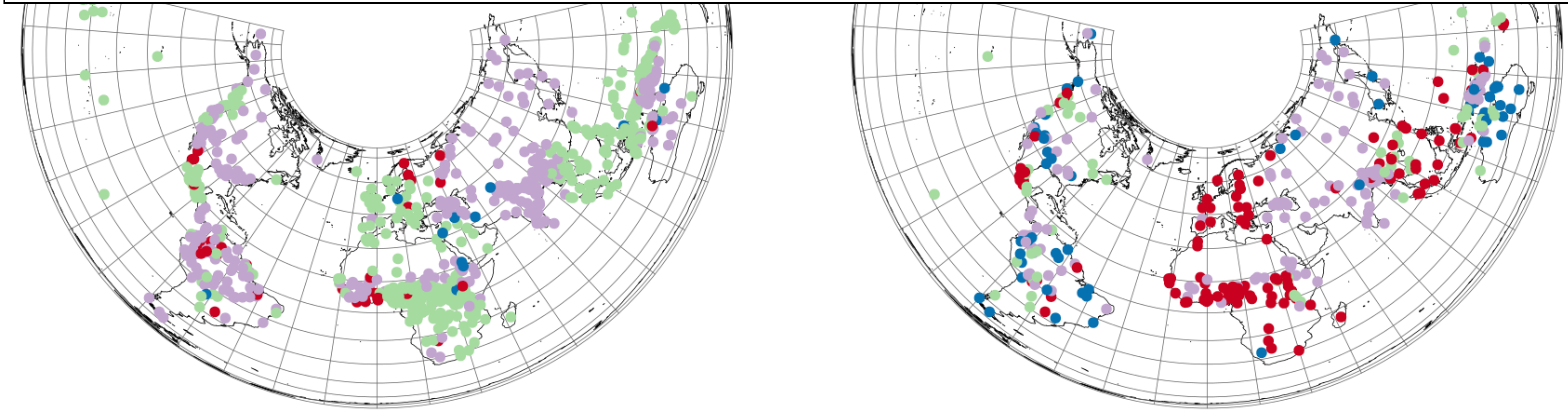
basic word order (WALS 83A) x stop voicing (WALS 04A)

VO, voicing
VO, no voicing
OV, voicing
OV, no voicing

Background

The geo-spatial properties of linguistic features

Harder to interpret than the simple case (therefore motivating defining some kind of metric ...)



basic word order (WALS 83A) x adposition order (WALS 85A)

VO, postpositions
VO, prepositions
OV, postpositions
OV, prepositions

basic word order (WALS 83A) x stop voicing (WALS 04A)

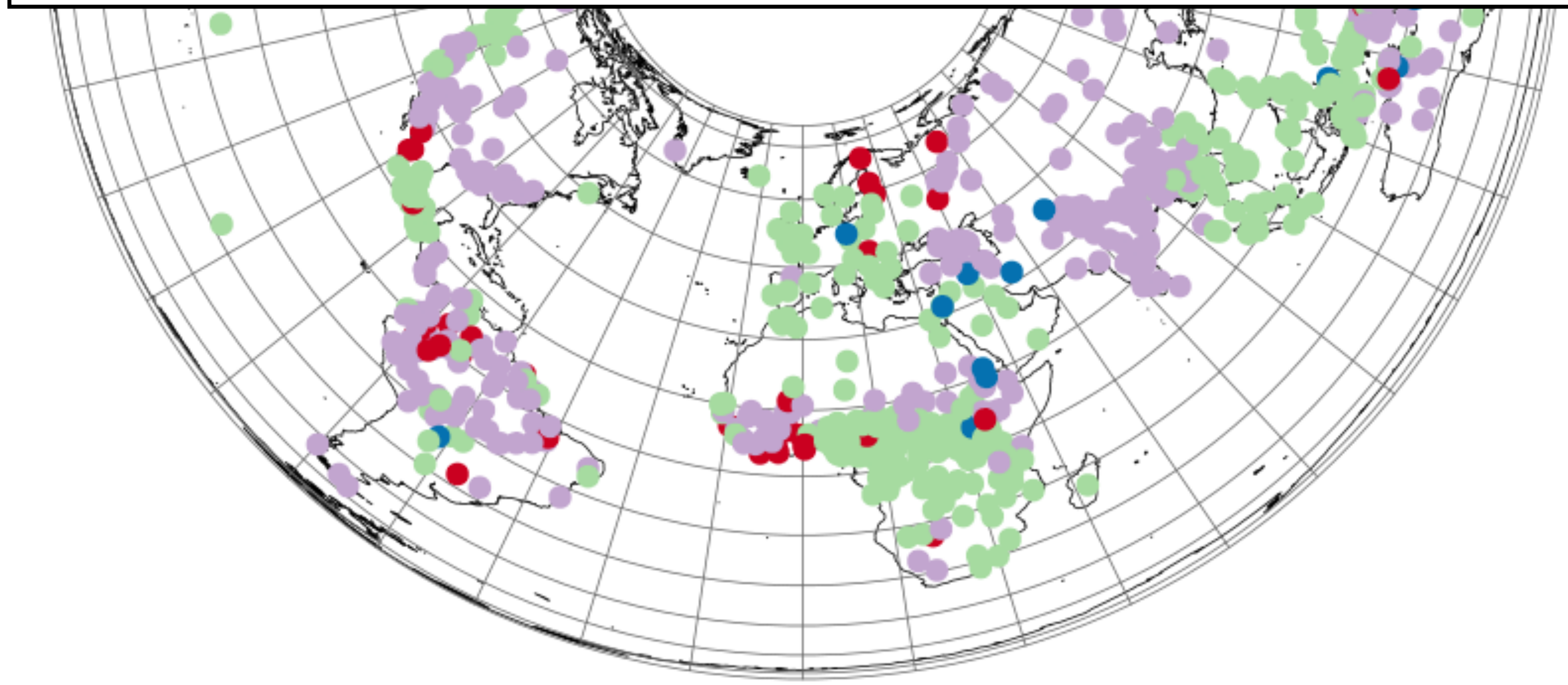
VO, voicing
VO, no voicing
OV, voicing
OV, no voicing

Background

The geo-spatial properties of linguistic features

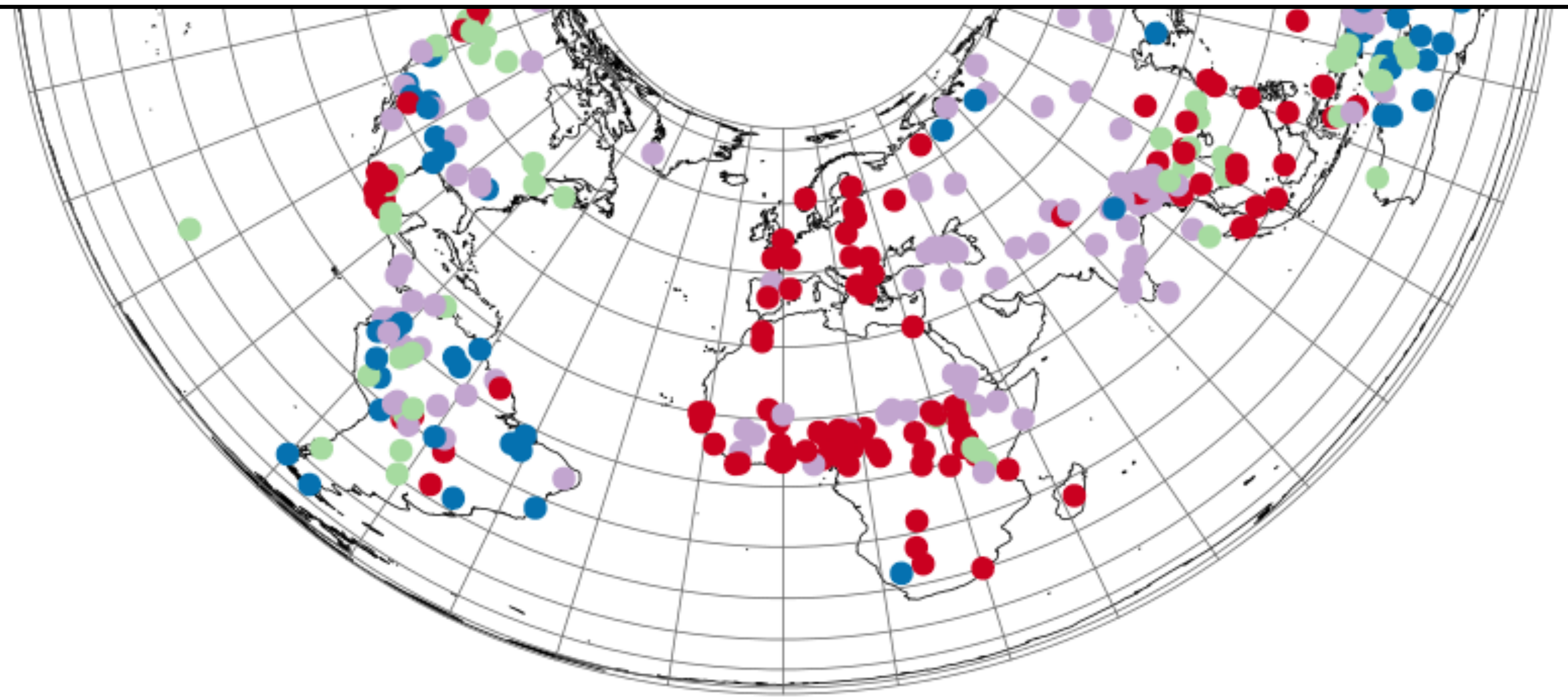
Harder to interpret than the simple case (therefore motivating defining some kind of metric ...)

Some incentive to claim that distribution of underlyingly-related features ≠ distribution of underlyingly-independent features.



basic word order (WALS 83A) x adposition order (WALS 85A)

VO, postpositions
VO, prepositions
OV, postpositions
OV, prepositions



basic word order (WALS 83A) x stop voicing (WALS 04A)

VO, voicing
VO, no voicing
OV, voicing
OV, no voicing

Modelling distributions of individual features

Rates of change & stability

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. **Question.** *Can **stability** be measured?*

Modelling distributions of individual features

Rates of change & stability

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. **Question.** *Can **stability** be measured?*

- Beginning with the **one-feature** case: considered individually, different typological features change on different timescales.
  - This has given rise to quite a bit of work on **rate-of-change estimation**, mostly in the typological tradition.
    - E.g. Maslova (2004), Wichmann and Holman (2009), Greenhill, Atkinson, Meade, and Gray (2010), Dediu (2011), Dediu and Cysouw (2013), Greenhill et al. (2017).
      - Also our own work: Kauhanen et al. (2021), on which more shortly.
  - Most of these: **discard spatial interactions between contiguous lgs.**
    - Heuristic of this kind of work: stable features are conserved within language families. For unstable features, there is within-family variation.

Modelling distributions of individual features

Rates of change & stability

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. **Question.** *Can **stability** be measured?*

'Stability estimation': usually, **discard spatial interactions between contiguous lgs.**

Modelling distributions of individual features

Rates of change & stability

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. **Question.** *Can **stability** be measured?*

'Stability estimation': usually, **discard spatial interactions between contiguous lgs.**

- Is this a good model of reality? **Problem(s, of which we are most concerned with this one). Perhaps some features are more prone to spatial interactions than others?** (and so discarding spatial info. distorts data).
    - **Susceptibility to (phylogenetic) change by descent mostly about L1; to change by contact mostly about L2.** Question then becomes whether we think these are homogeneous across features ...
        - Eg. uninterpretable (syntactic!) features — systematically L2-difficult, irrespective of L1 content? (Hawkins & Hatori 2006; Tsimpli & Dimitrakopoulou 2007)
        - Work arguing that 'simplifying' change emerges from wholesale L2 learning (Trudgill 2001, Walkden & Breitbarth 2019) — *vulnerability to simplifying change = vulnerability to spatial interactions?*
- Broader explicit point: if we think that the cognitive abilities involved in contact situations/L2 learning are a proper <u>subset</u> of those involved in child language acquisition, then we should be suspicious of the idea that spatial interactions aren't variable.

Modelling distributions of individual features

Our model (so far)

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. **Question.** *Can **stability** be measured?*

Conjecture. Some features are more prone to spatial interactions than others.

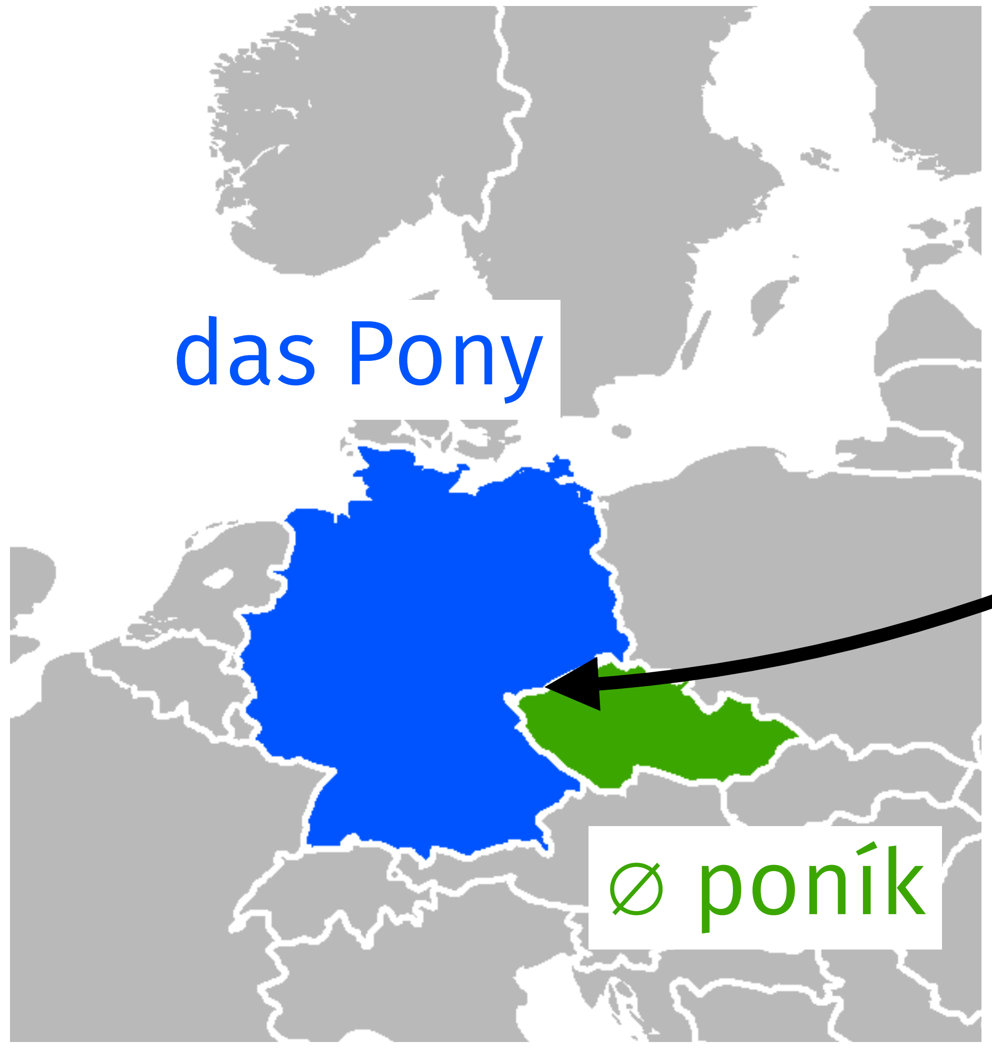
Modelling distributions of individual features

Our model (so far)

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. **Question.** *Can **stability** be measured?*

Conjecture. Some features are more prone to spatial interactions than others.

Quantify this? **Isogloss density** σ (probability that two neighbours disagree).



A trivial isogloss — boundary between opposing values of this binary feature, ± def. art.

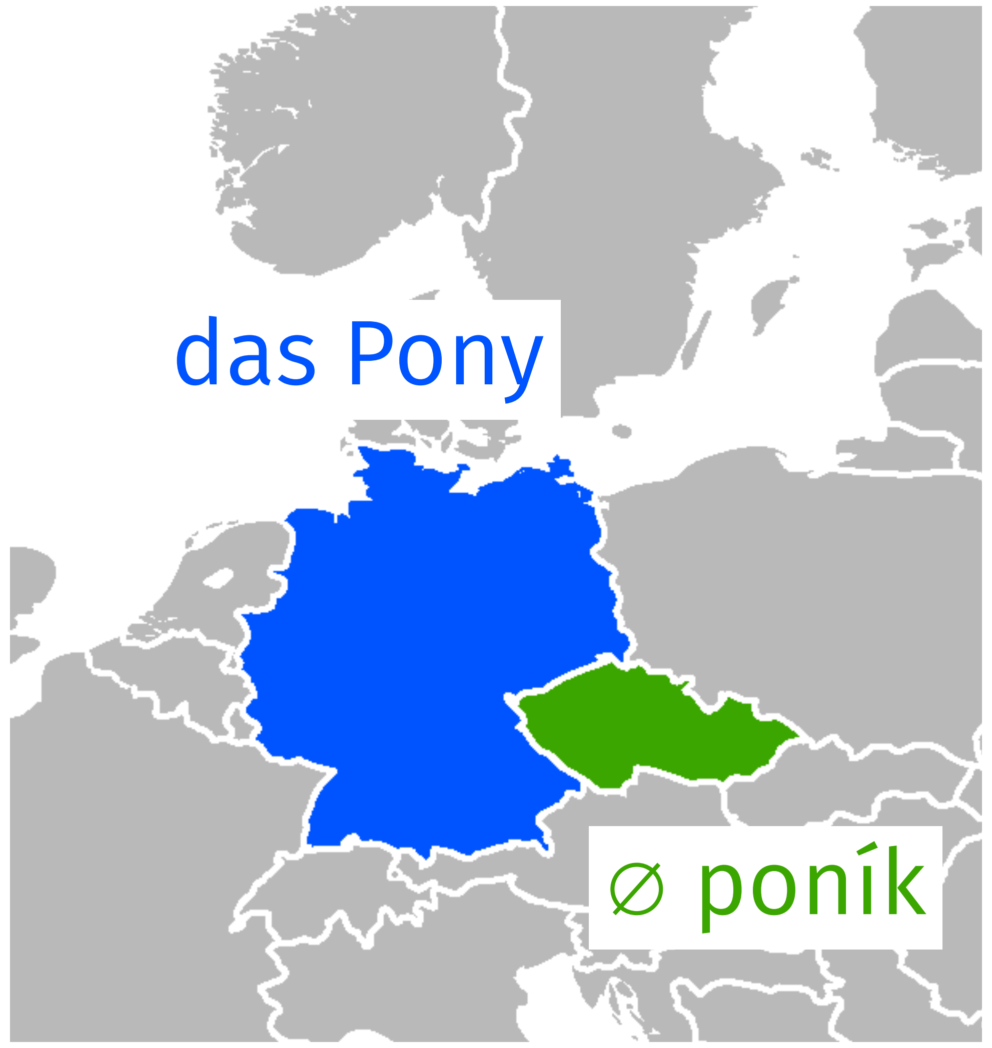
Modelling distributions of individual features

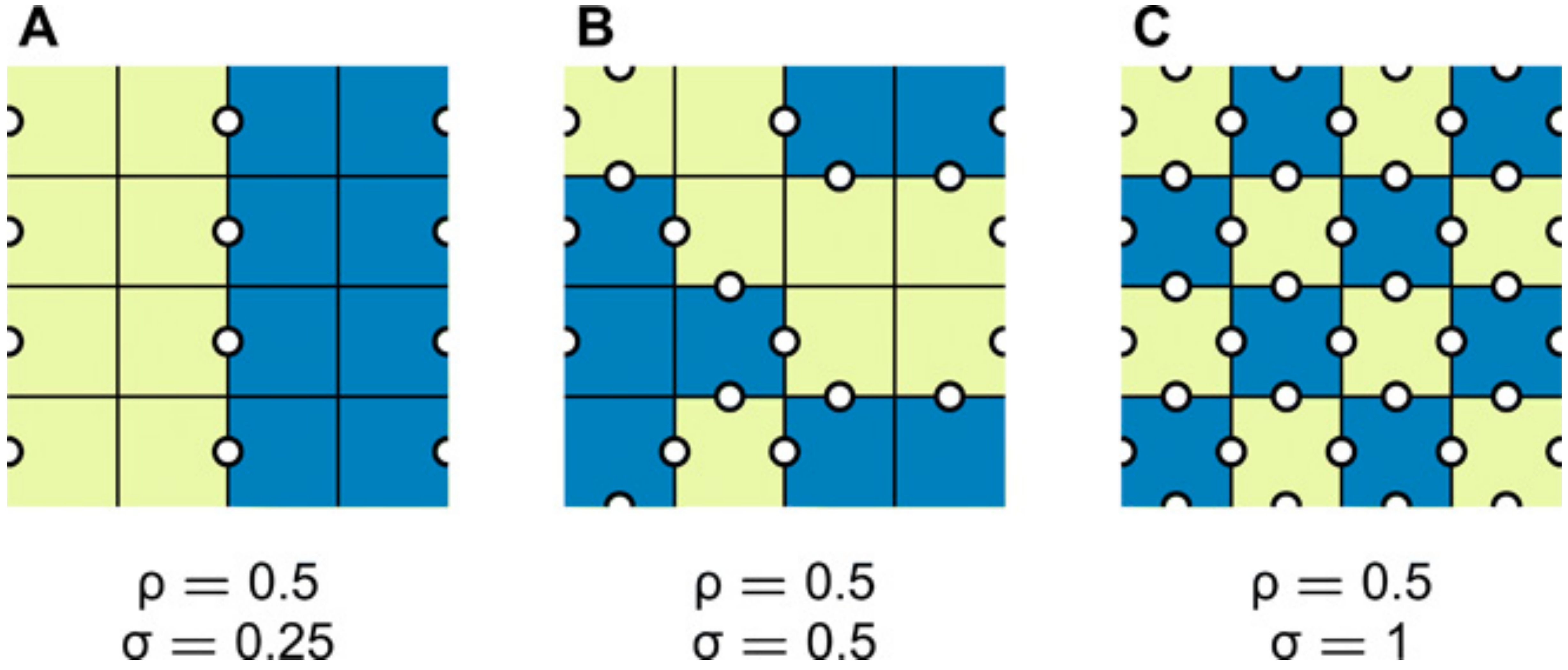
Our model (so far)

Heuristic: ***unstable*** features scatter, ***stable*** features cluster. **Question.** *Can **stability** be measured?*

Conjecture. Some features are more prone to spatial interactions than others.

Quantify this? **Isogloss density** σ (probability that two neighbours disagree).





For constant feature frequency $\varrho = 0.5$ (half the sites are blue, half are yellow), 3 different values of σ isogloss density — low when opposing values sort into extended domains, intermediate 'random', high when values are *preferentially* scattered.

Modelling distributions of individual features

Our model (so far)

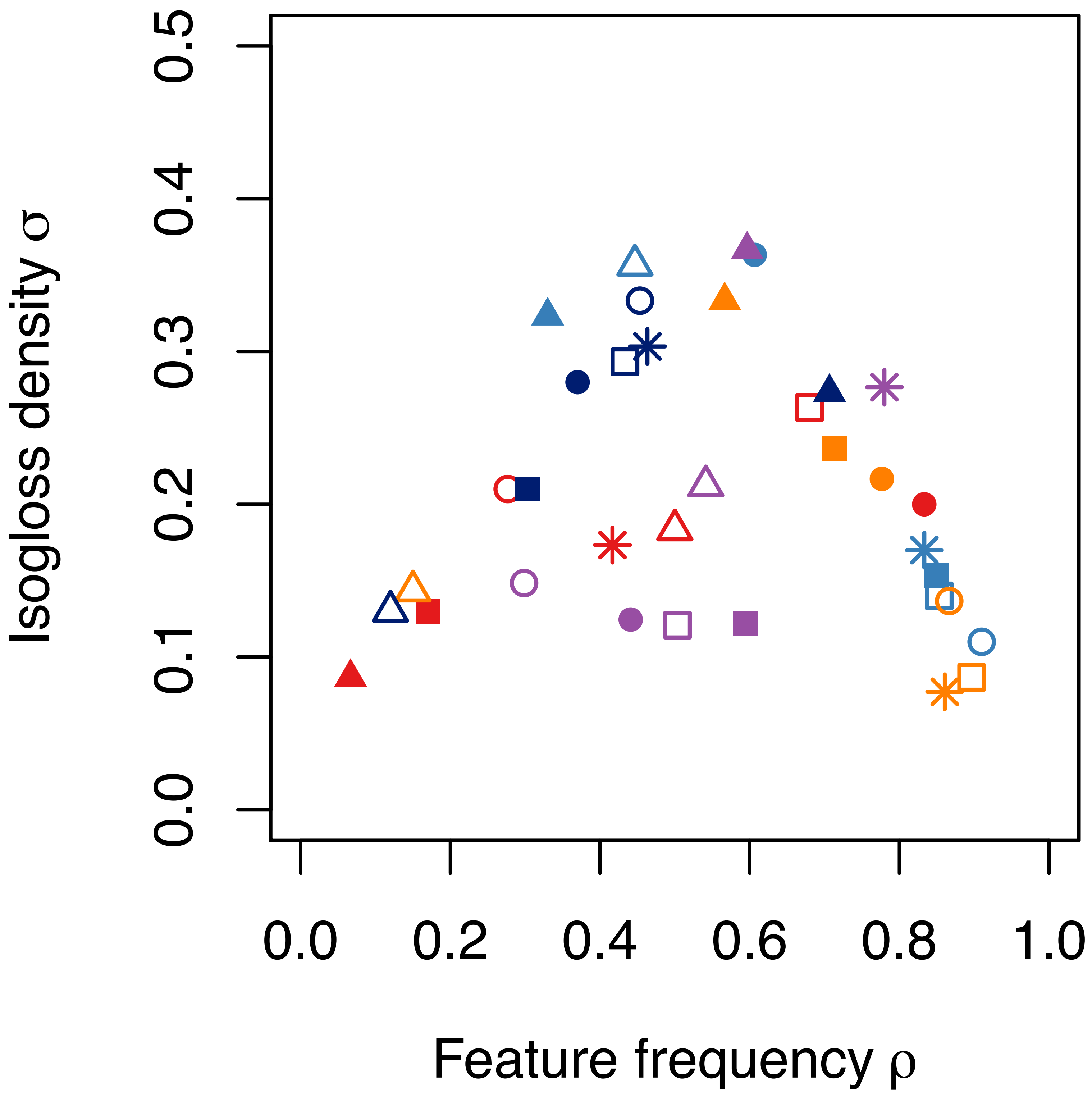
Isogloss density σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).

1. voicing contrast
2. uvular consonants
3. glottalized consonants
4. lateral consonants
5. velar nasal
6. front rounded vowels
7. tone
8. inflectional morphology
9. productive reduplication
10. plural
11. definite article
12. indefinite article
13. gender distinctions in independent personal pronouns
14. adpositions
15. ordinal numerals
16. possessive affixes
17. tense–aspect inflection
18. morphological second-person imperative
19. inflectional optative
20. grammatical evidentials
21. question particle
22. verbal person marking
23. order of subject and verb is SV
24. order of object and verb is OV
25. order of genitive and noun is GenN
26. order of adjective and noun is AdjN
27. order of numeral and noun is NumN
28. order of degree word and adjective is DegAdj
29. preverbal negative morpheme
30. postverbal negative morpheme
31. passive construction
32. shared encoding of nominal and locational predication
33. zero copula for predicate nominals
34. *hand* and *arm* identical
35. *hand* and *finger(s)* identical

Modelling distributions of individual features

Our model (so far)

Isogloss density σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).

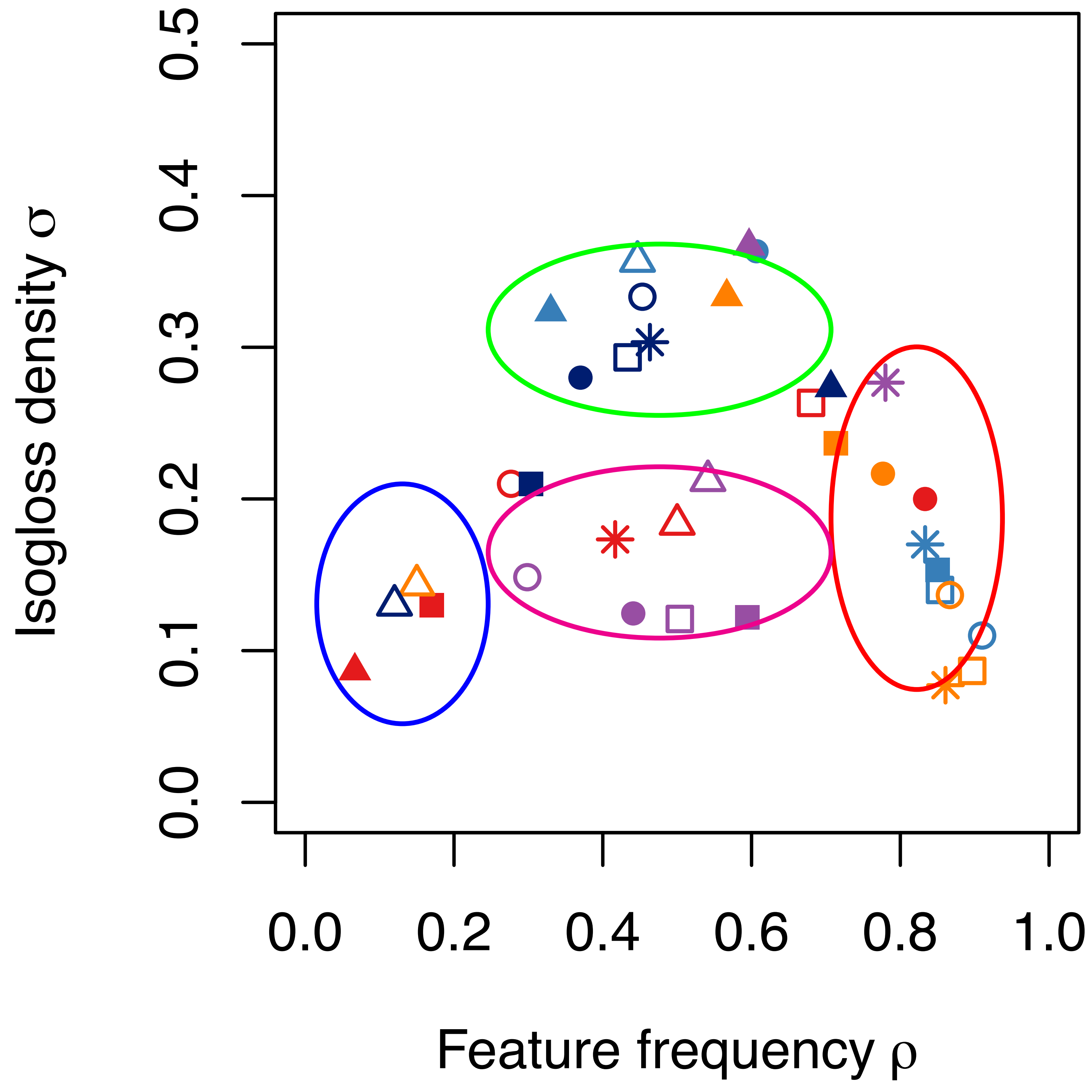


- Greenberg (1978): we should be able to capture this variability entirely in terms of **ingress** and **egress** probabilities,
  - **ingress:** p. of a language acquiring feature *f*
  - **egress**: p. of a language losing feature *f*
- *low ingress, low egress*
  low ingress, high egress
  high ingress, low egress
  high ingress, high egress
- Old and previously untested intuition …

Modelling distributions of individual features

Our model (so far)

Isogloss density σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



- Greenberg (1978): we should be able to capture this variability entirely in terms of **ingress** and **egress** probabilities,
  - **ingress:** p. of a language acquiring feature *f*
  - **egress**: p. of a language losing feature *f*
- *low ingress, low egress* — **low-scattering** (not very susceptible to change)
  low ingress, high egress — **rare**, low scattering (universally absent)
  high ingress, low egress — **common**, low scattering (universal)
  high ingress, high egress — **high-scattering** (susceptible to all change)
- Old and previously untested intuition …

Modelling distributions of individual features

Our model (so far)

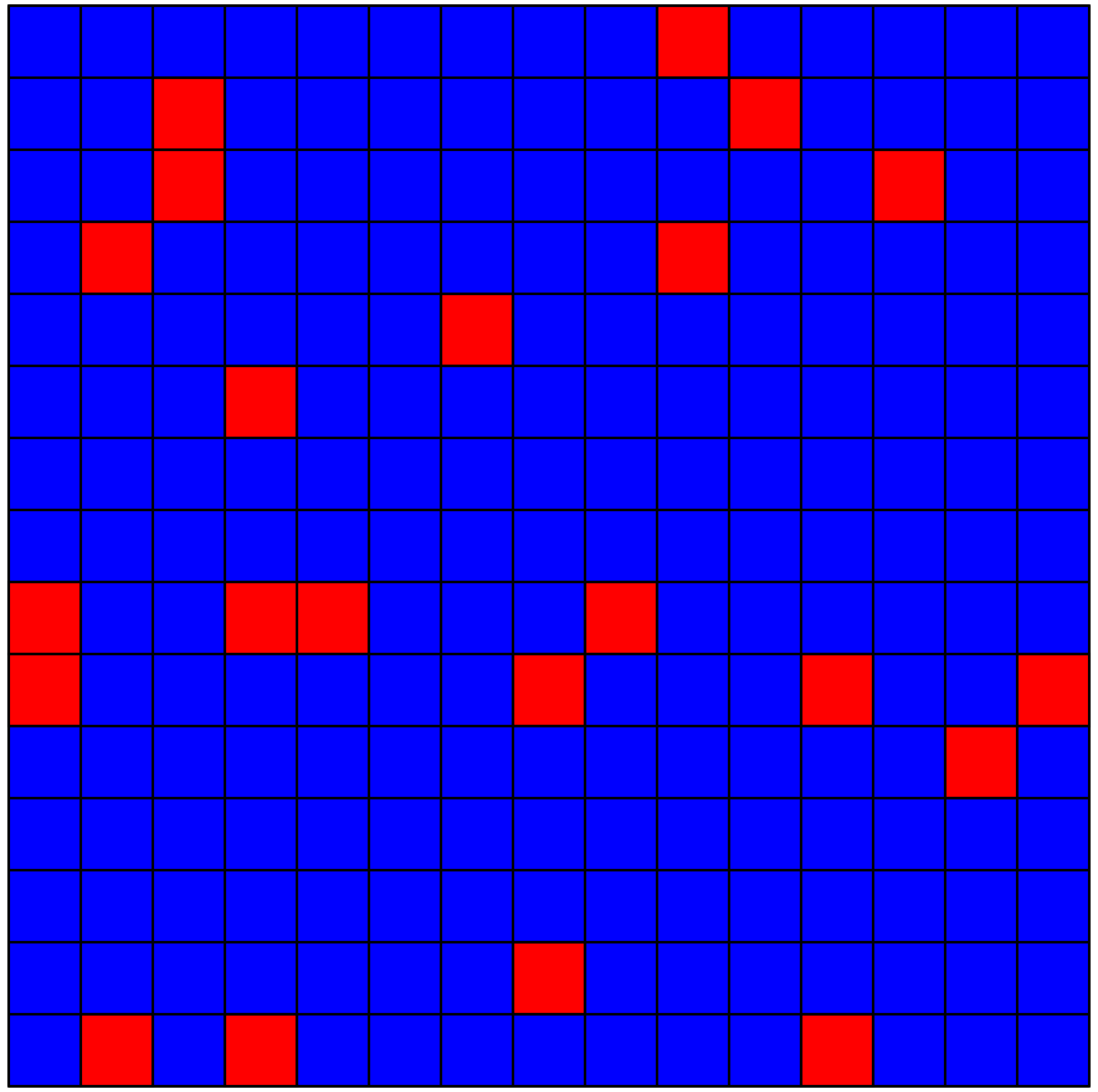
Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).

- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



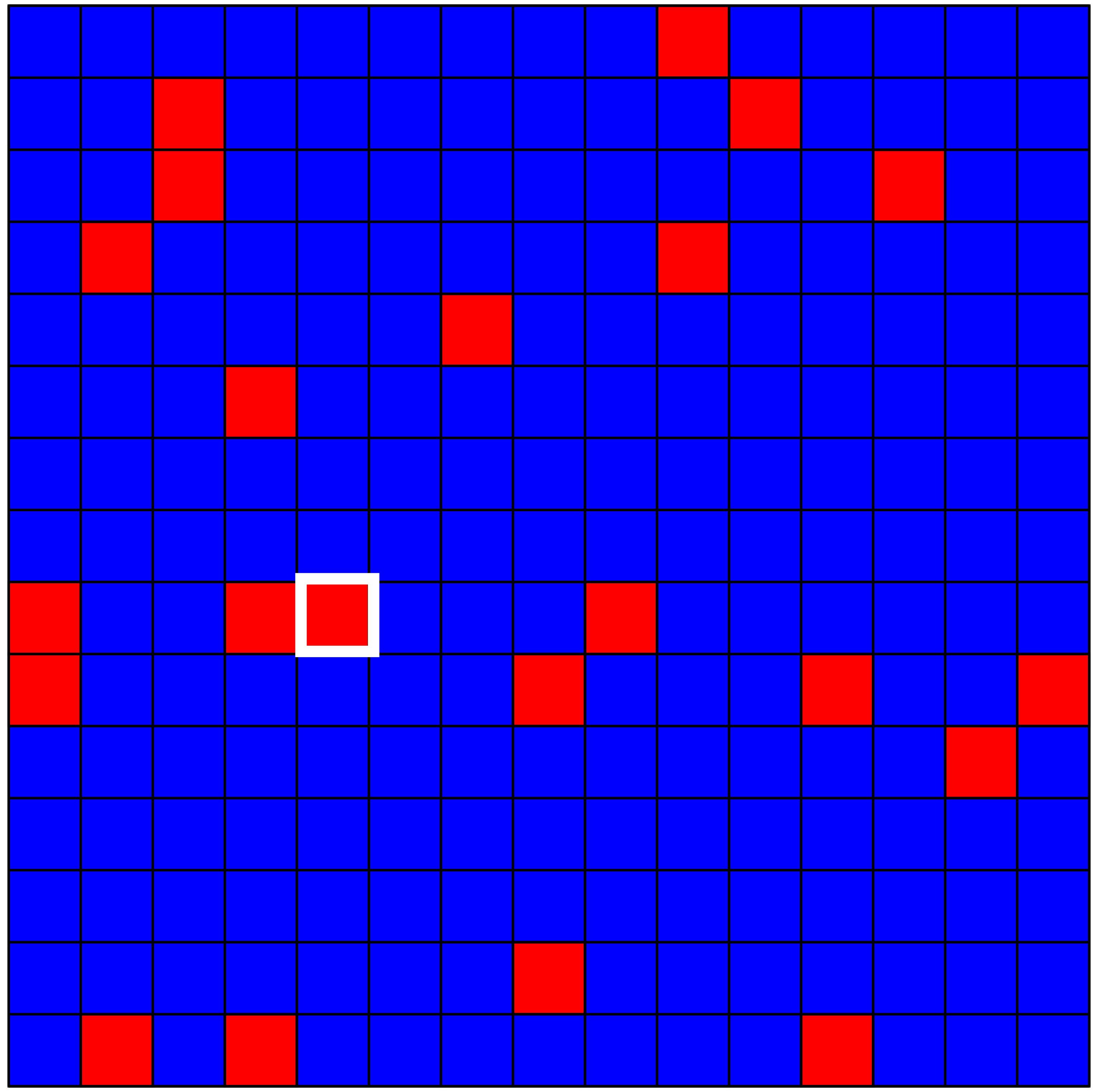
- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

Think about a grid of cells — a regular lattice with periodic boundary conditions. Each cell has one of two feature values — blue, red.

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



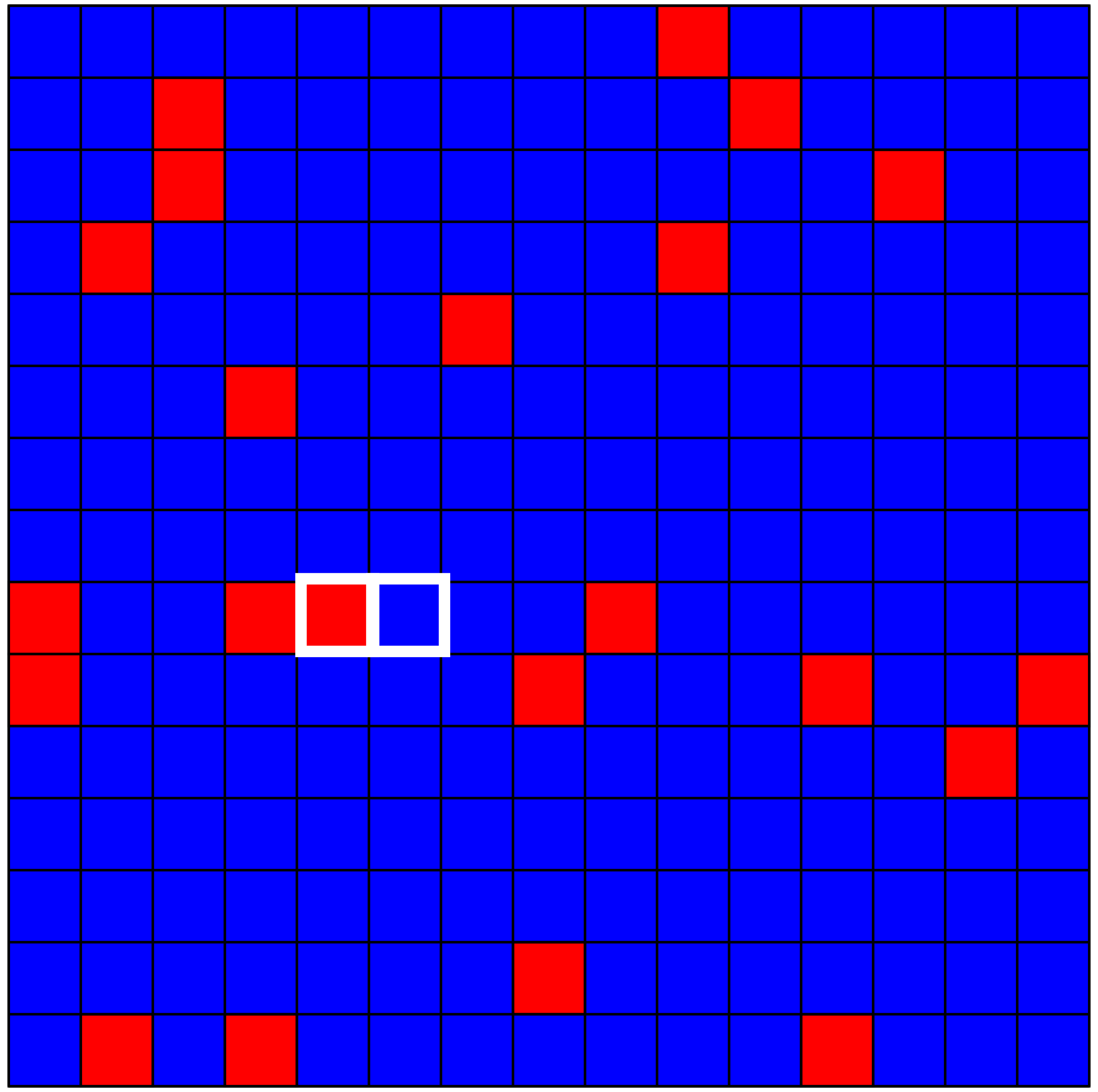
- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - ‘Greenbergian’ egress, ingress;
  - Classical “voter model” for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

‘contact dynamic’: with probability *q*, pick a random site

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



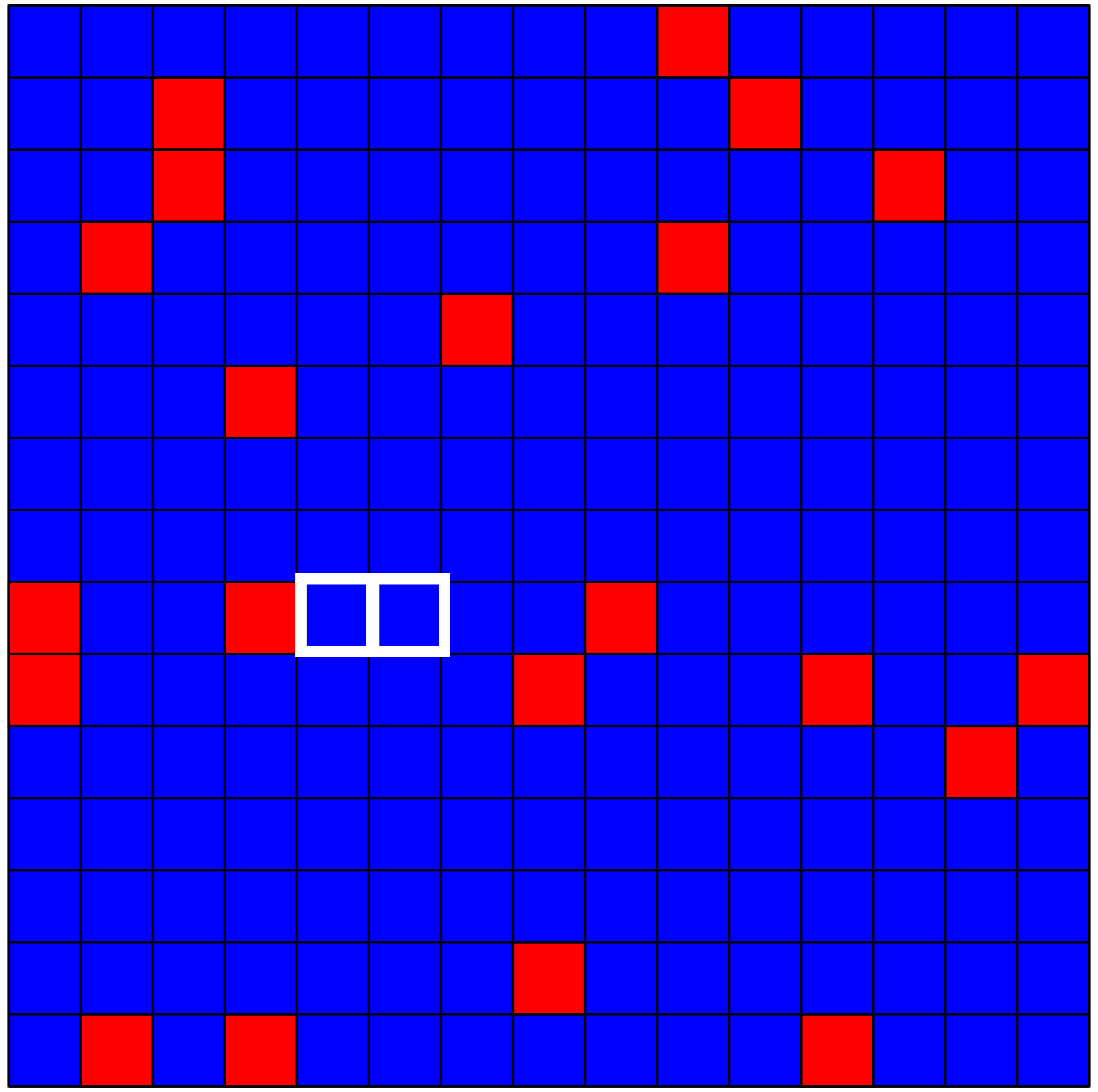
- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - ‘Greenbergian’ egress, ingress;
  - Classical “voter model” for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

‘contact dynamic’: with probability *q*,
pick a random site
pick a neighbour

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



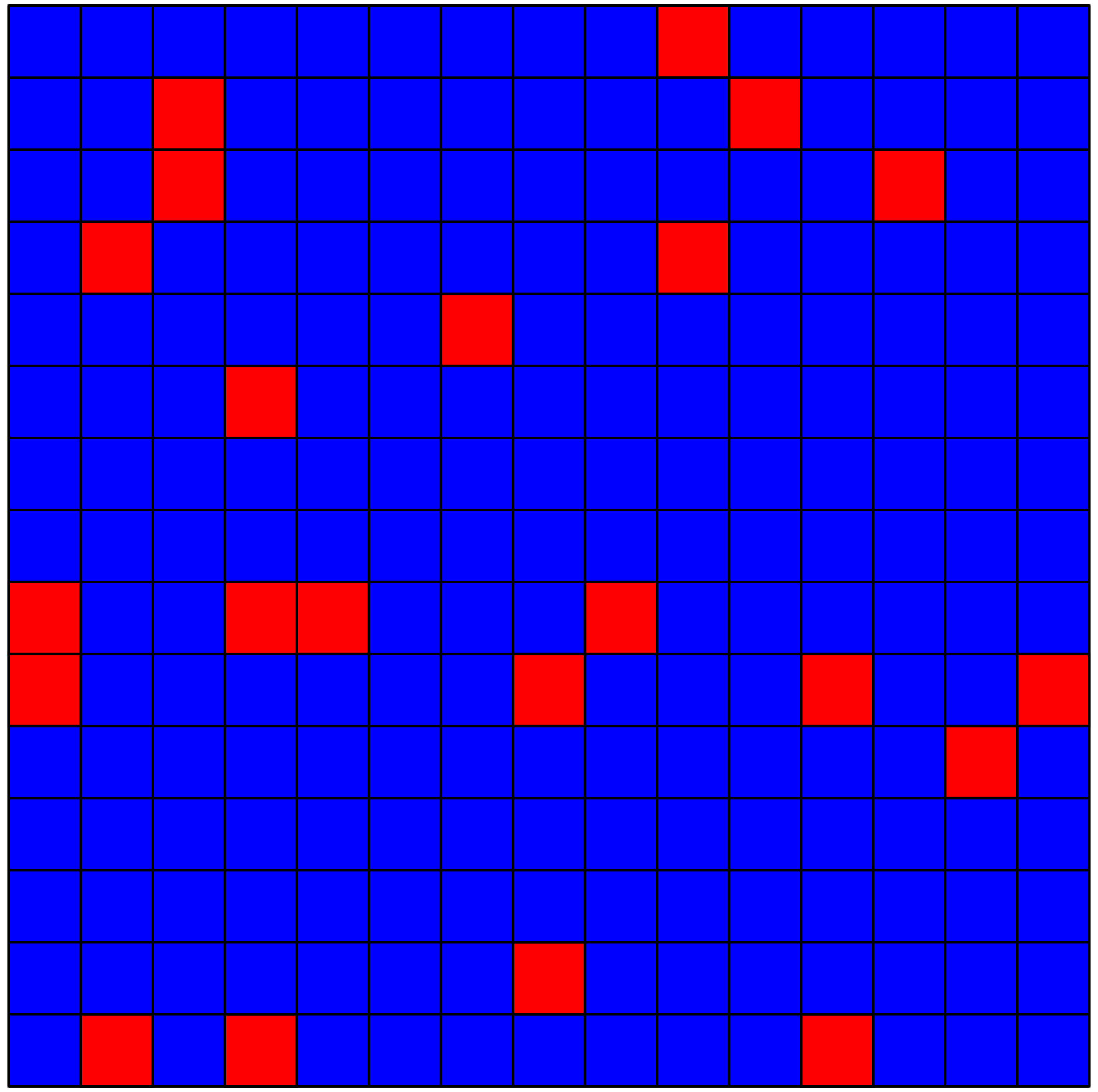
- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture ...
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

'contact dynamic': with probability *q*,
pick a random site
pick a neighbour
copy the neighbour's value

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

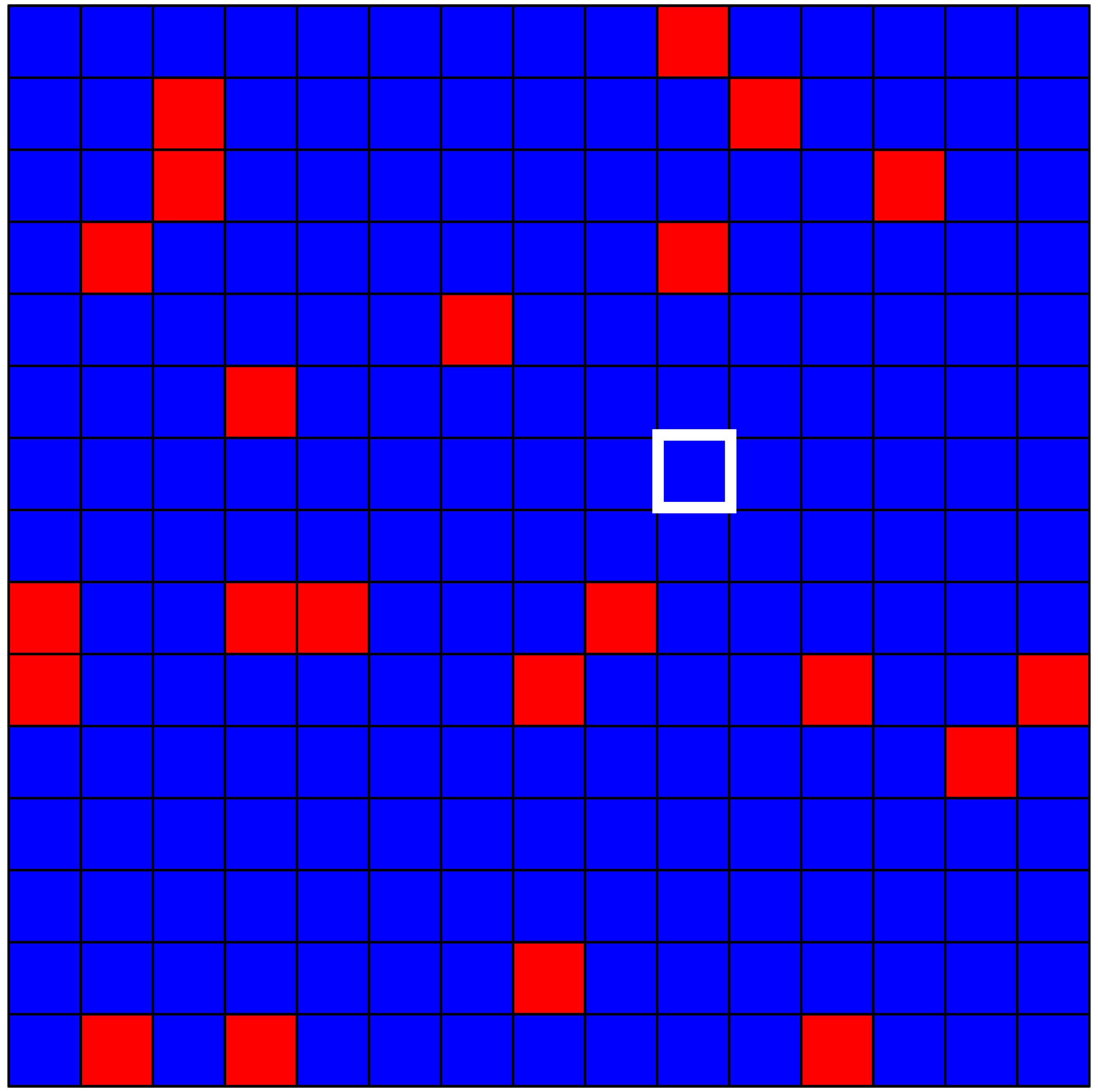
'contact dynamic': with probability q,
pick a random site
pick a neighbour
copy the neighbour's value

or

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

'contact dynamic': with probability q,
pick a random site
pick a neighbour
copy the neighbour's value

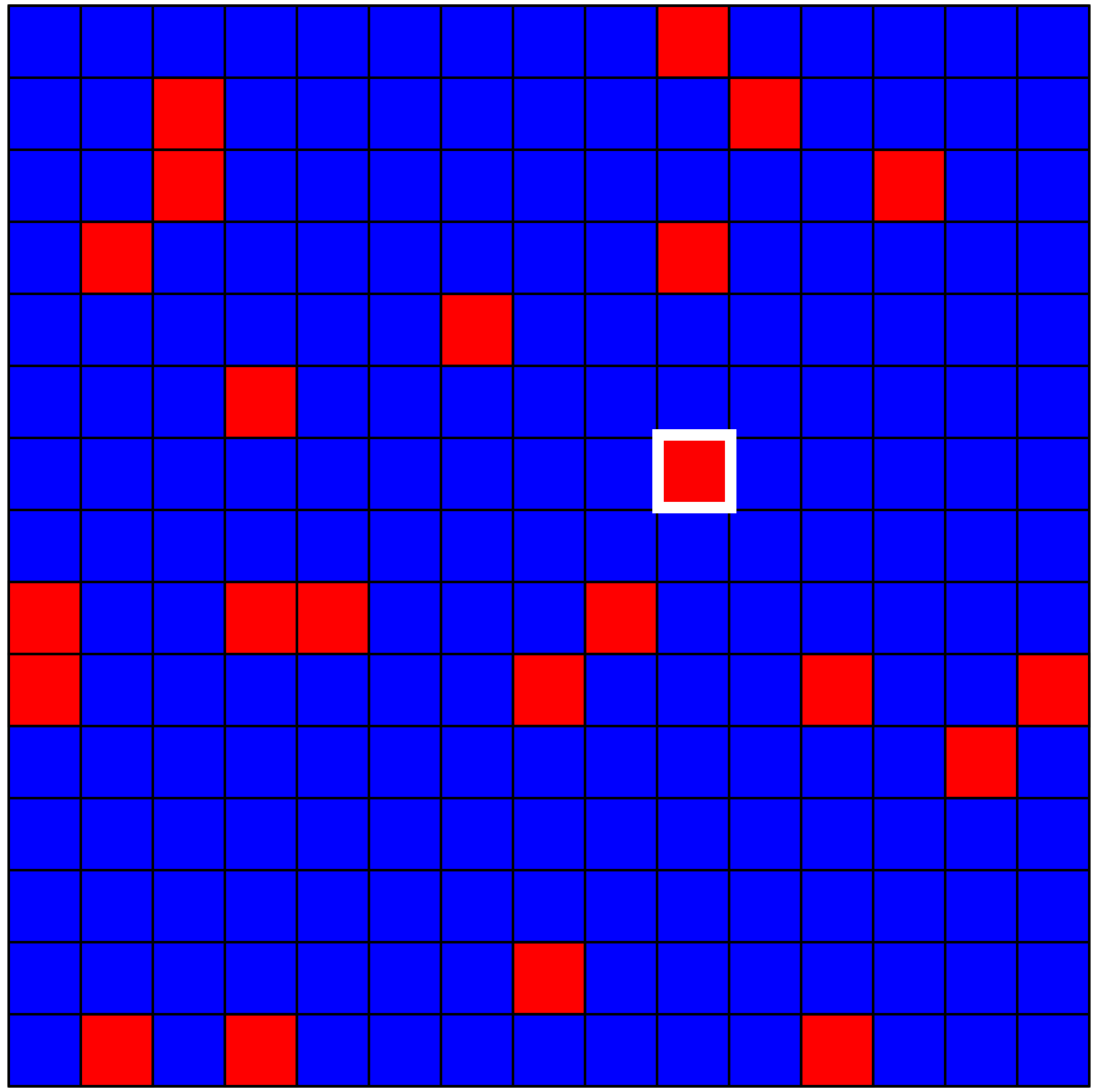
or

'genetic dynamic': with probability $1-q$,
pick a random site

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture ...
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

'contact dynamic': with probability q,
pick a random site
pick a neighbour
copy the neighbour's value

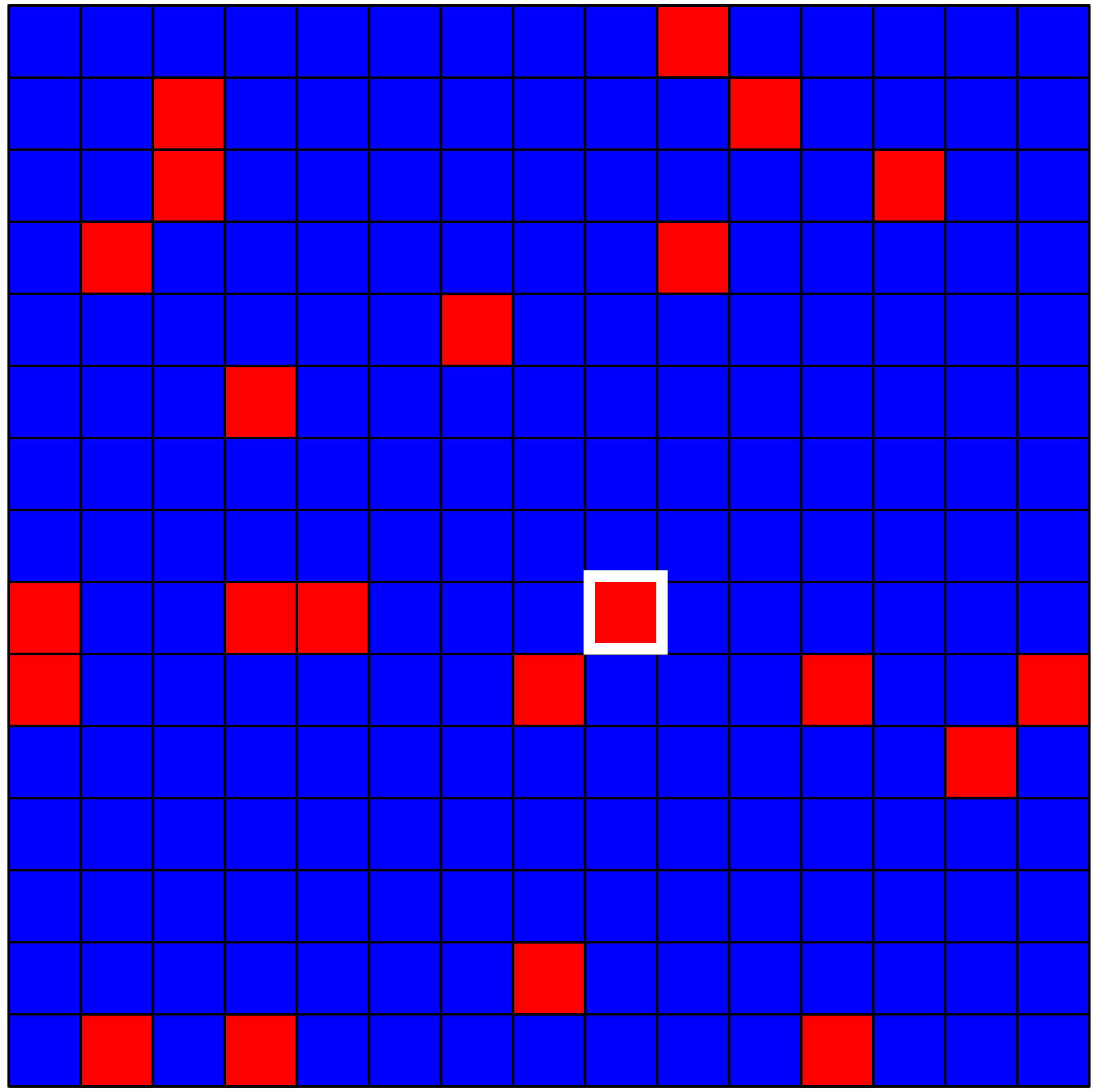
or

'genetic dynamic': with probability $1-q$,
pick a random site
with probability p_i (ingress) innovate

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - ‘Greenbergian’ egress, ingress;
  - Classical “voter model” for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

‘contact dynamic’: with probability q,
pick a random site
pick a neighbour
copy the neighbour’s value

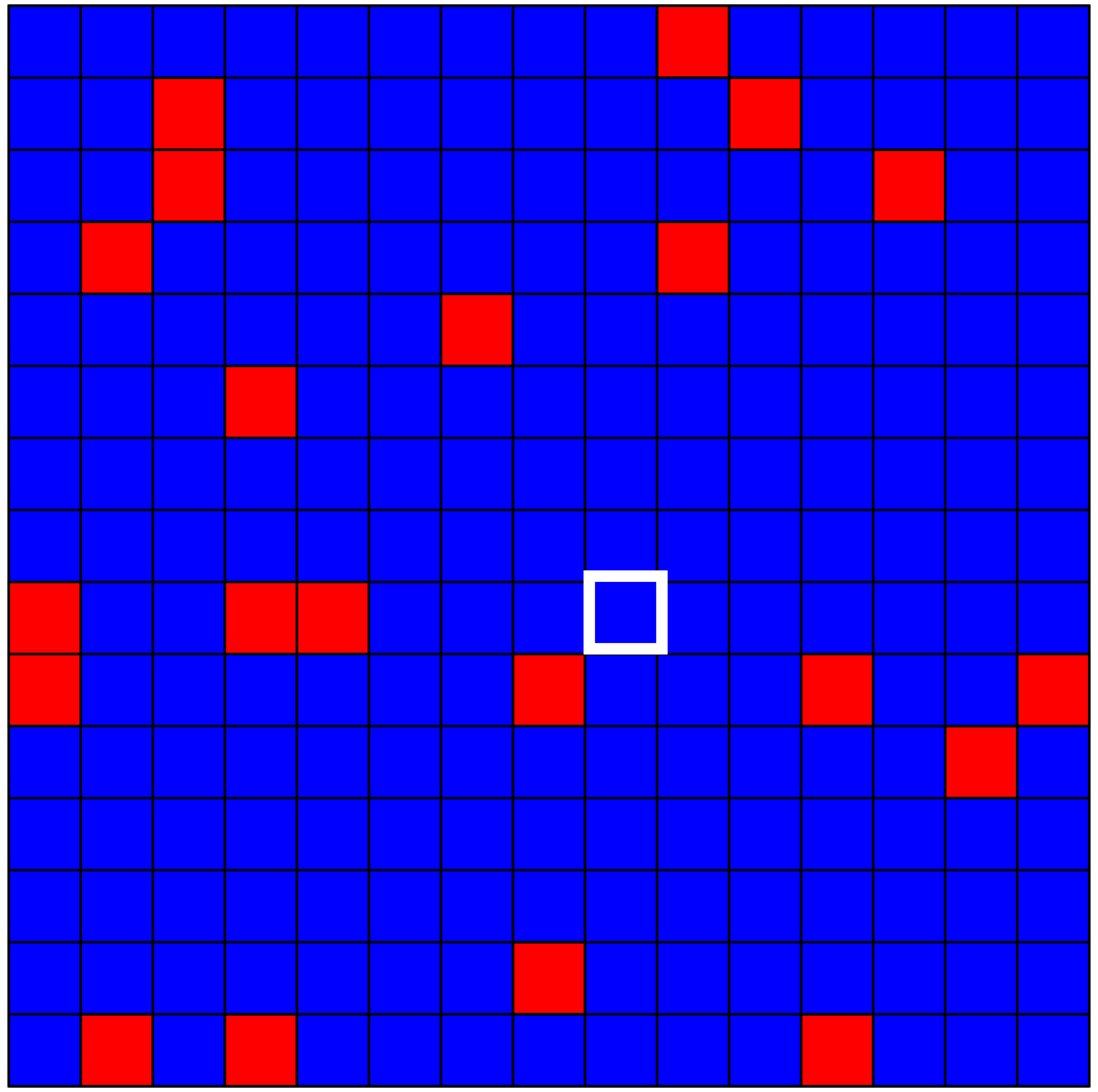
or

‘genetic dynamic’: with probability 1-q,
pick a random site
with probability p_i (ingress) innovate **or**

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

'contact dynamic': with probability q,
pick a random site
pick a neighbour
copy the neighbour's value

or

'genetic dynamic': with probability $1-q$,
pick a random site
with probability p_i (ingress) innovate **or**
with probability p_e (egress) lose feature

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).

Steady-state feature frequency (probability of feature) is

$$\rho = \frac{p_i}{p_i + p_e} = \frac{1}{1 + p_e/p_i}$$

Steady-state isogloss density is

$$\sigma = h(\tau)\rho(1-\rho)$$

with

$$h(\tau) = \frac{(1+\tau)\pi}{K\left(\frac{1}{1+\tau}\right)} - 2\tau,$$

where $K(\cdot)$ is the complete elliptic integral of the first kind and

$$\tau = \frac{(1-q)(p_i+p_e)}{q}.$$

- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

This turns out to have an analytic solution — the values of ϱ and σ at the stationary state can be solved as soon as p_i, p_e and q are known.

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).

Steady-state feature frequency (probability of feature) is

$$\rho = \frac{p_i}{p_i + p_e} = \frac{1}{1 + p_e/p_i}$$

Steady-state isogloss density is

$$\sigma = h(\tau)\rho(1-\rho)$$

with

$$h(\tau) = \frac{(1+\tau)\pi}{K\left(\frac{1}{1+\tau}\right)} - 2\tau,$$

where $K(\cdot)$ is the complete elliptic integral of the first kind and

$$\tau = \frac{(1-q)(p_i+p_e)}{q}.$$

- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture ...
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

This turns out to have an analytic solution — the values of ϱ and σ at the stationary state can be solved as soon as p_i, p_e and q are known.

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).

Steady-state feature frequency (probability of feature) is

Solution is a function of [something] τ, which gives us an single overall parameter we can use to talk about the stability of a feature.

Steady

with

$$(\tau) = \frac{(1+\tau)\pi}{K\left(\frac{1}{1+\tau}\right)} - 2\tau,$$

where $K(\cdot)$ is the complete elliptic integral of the first kind and

$$\tau = \frac{(1-q)(p_i + p_e)}{q}.$$

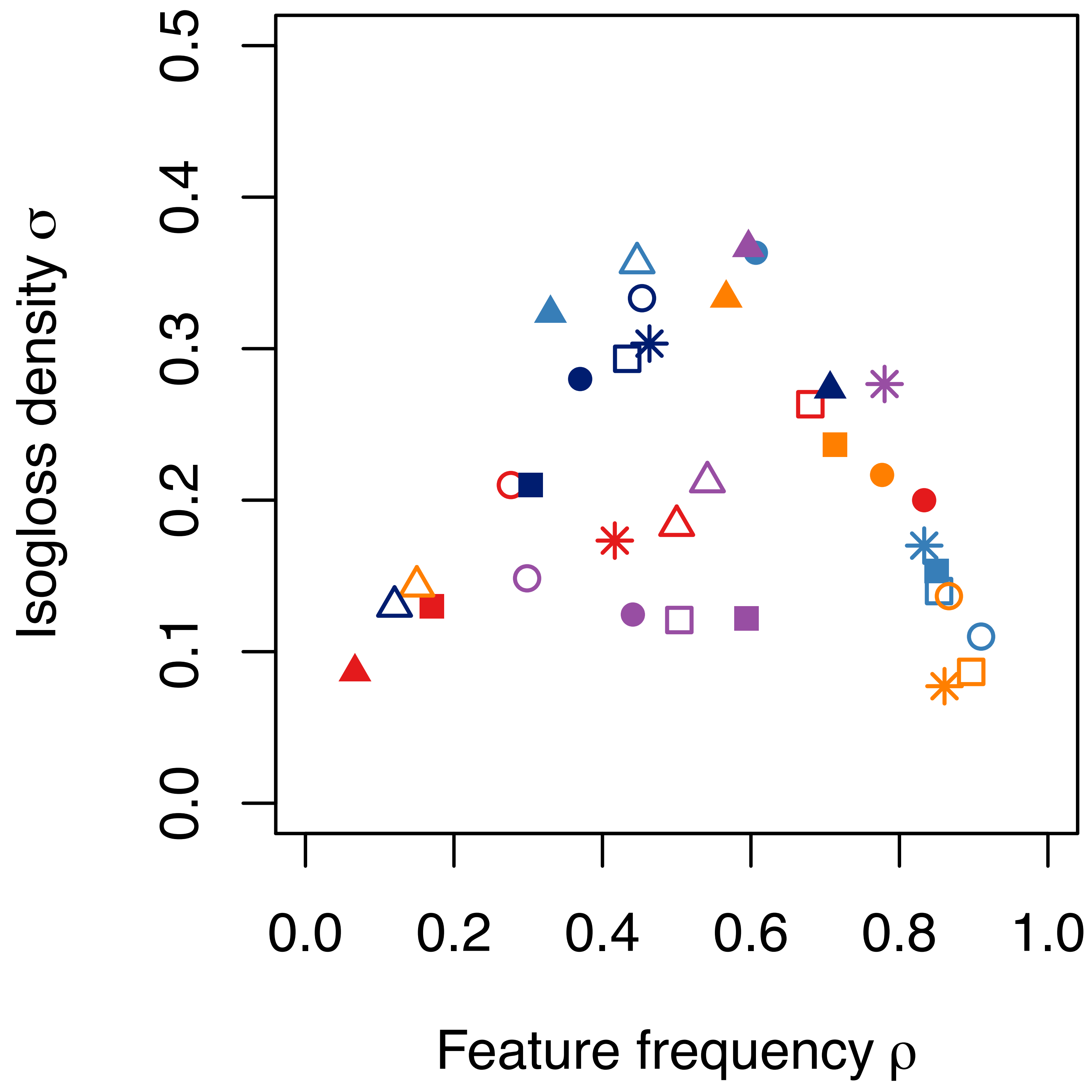
- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - 'Greenbergian' egress, ingress;
  - Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

This turns out to have an analytic solution — the values of ϱ and σ at the stationary state can be solved as soon as p_i, p_e and q are known.

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).



- Kauhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …
  - ‘Greenbergian’ egress, ingress;
  - Classical “voter model” for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

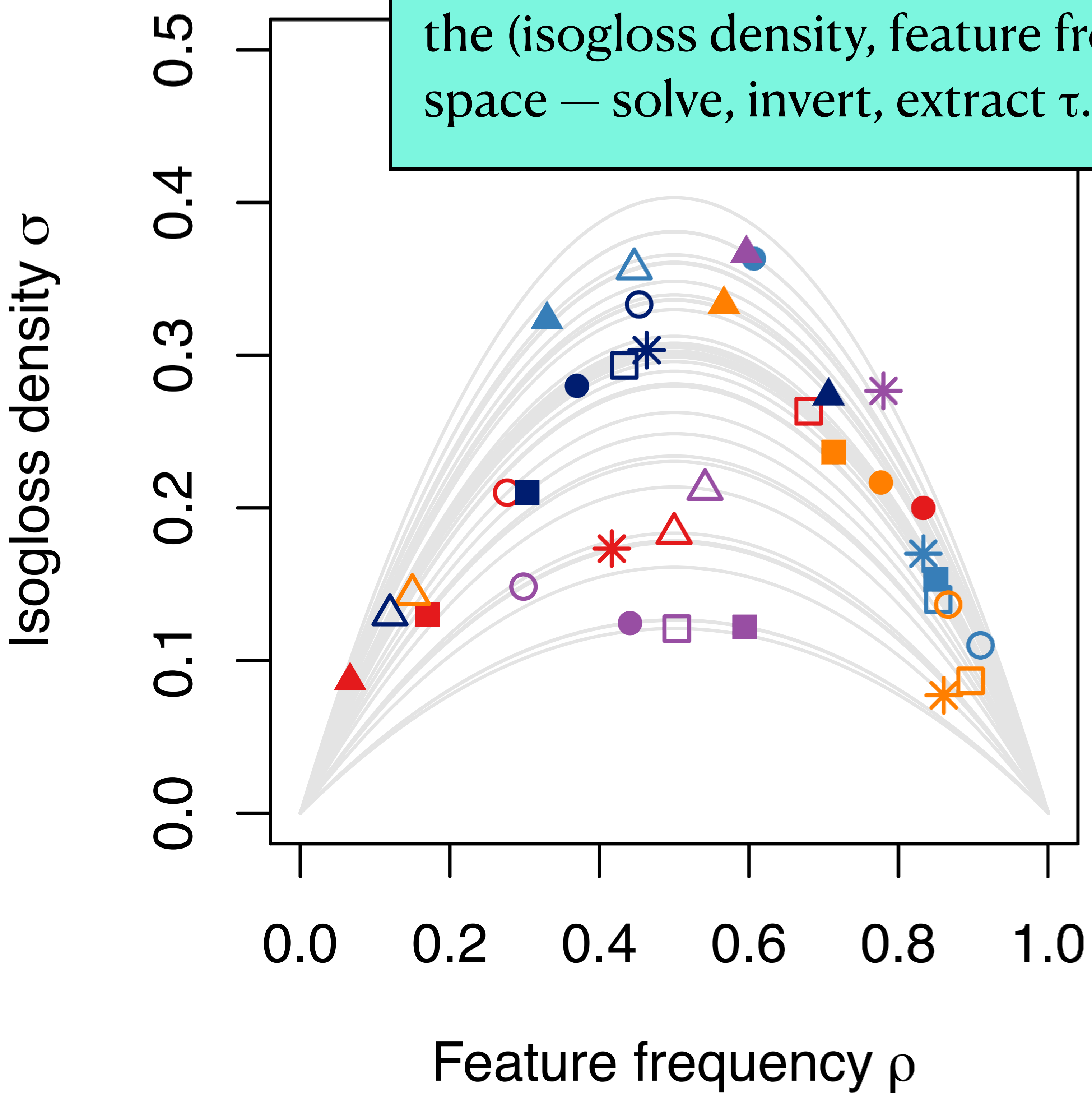
This turns out to have an analytic solution — the values of ρ and σ at the stationary state can be solved as soon as p_i, p_e and q are known.

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).

Each feature lies on a unique parabola in the (isogloss density, feature frequency) space — solve, invert, extract τ.



uhanen et al. (2021): **formalise Greenberg**, adding an explicit **spatial dynamic** into the picture …

- 'Greenbergian' egress, ingress;
- Classical "voter model" for spatial interactions (see Castellano, Fortunato, & Loreto, 2009, for an overview).

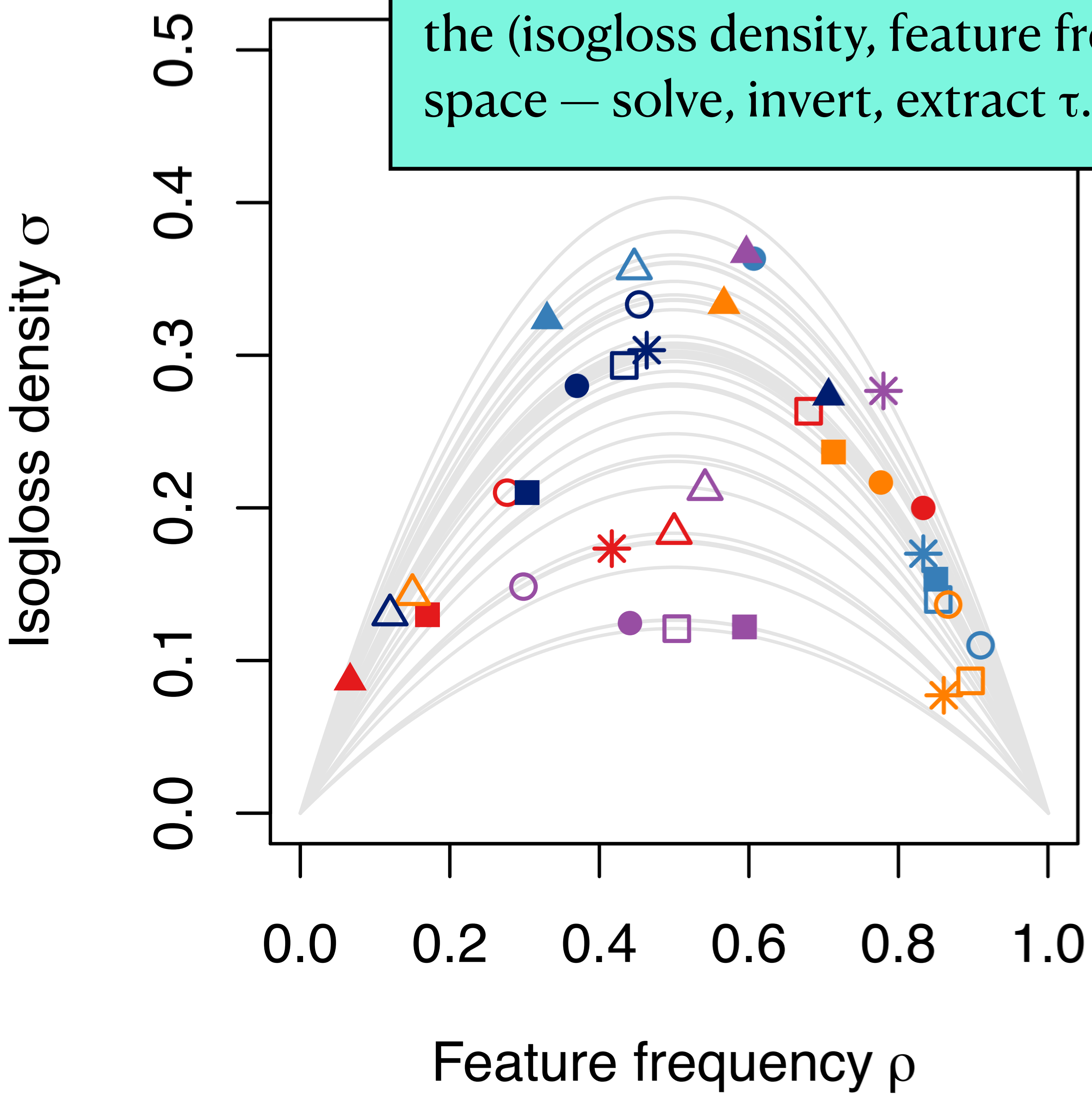
This turns out to have an analytic solution — the values of ϱ and σ at the stationary state can be solved as soon as p_i, p_e and q are known.

Modelling distributions of individual features

Our model (so far)

Density of reactive interfaces σ (probability that two neighbours disagree) for a subset of the WALS data (35 features).

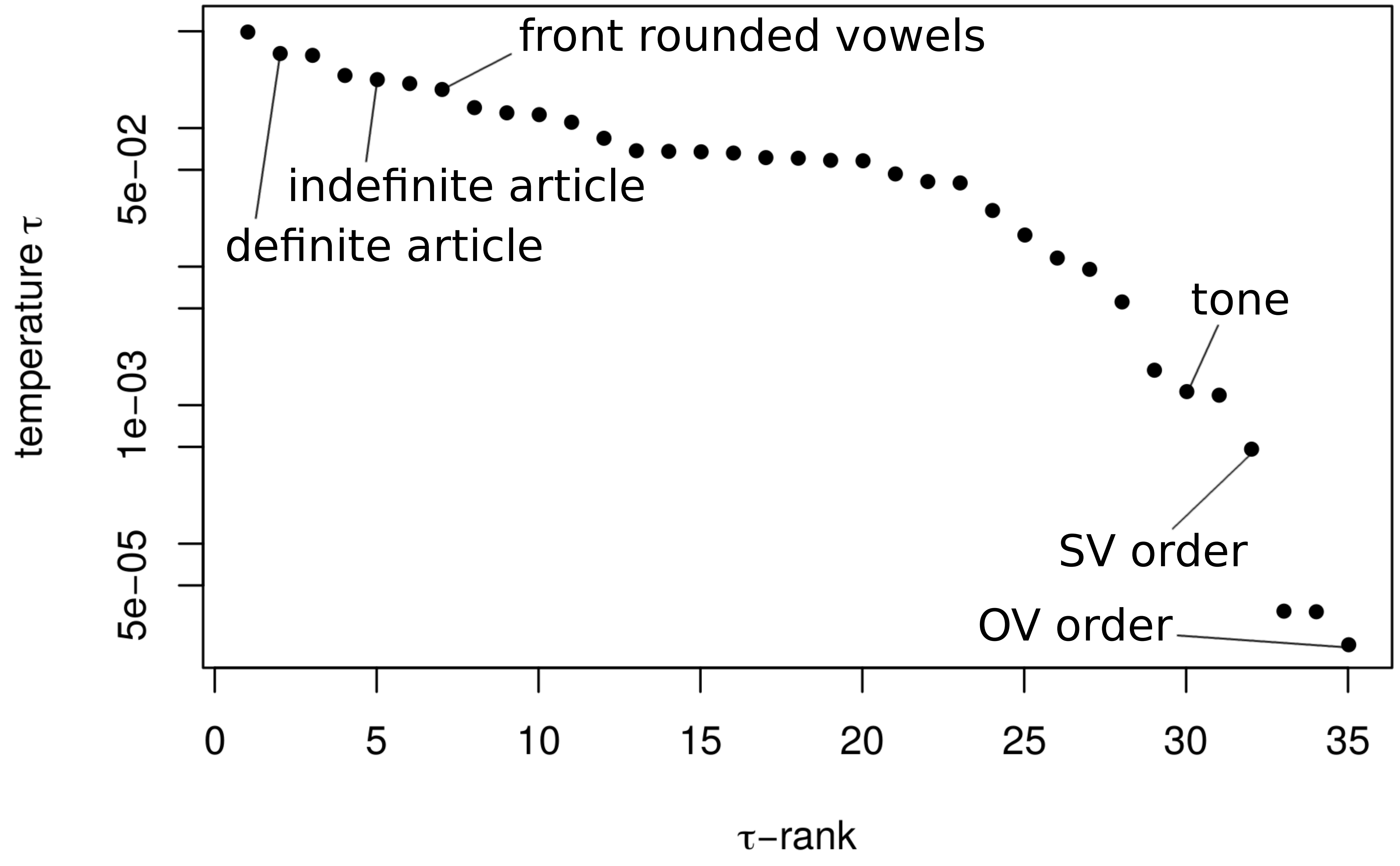
Each feature lies on a unique parabola in the (isogloss density, feature frequency) space — solve, invert, extract τ.



uh
expli

- ‘G
- Cl
  Ca

This
and
and



Correlations between features

Typological observations

The story so far. (We think) spatial distributions of individual features emerge from properties that we can treat as inherent to each feature — probabilities of *egress* and *ingress* ⊆ parameter denoting overall feature stability.

- **But,** we have talked about 'features' as though they operate independently.
  - Not very plausible for a number of reasons — long history of implicational universals that involve multiple typological features of the type discussed here, tendency of syntactic properties to cluster in a 'macroparameterish' way ...
  - Recent work in parametric comparison eg. Guardiano & Longobardi 2016, Ceolin et al. 2020: goes beyond much prev. literature in discarding *redundant* values where there are obvious interdependencies between parameters, but not always clear* how we capture *non-redundant* statistical correlations, hierarchical structure, etc.
    * to me
- From our point of view, to really capture this (and in order to make predictions about *distance*, implicitly involving more than 1 feature) there are two key tasks.
  - **empirically** test whether 'preferred' and 'dispreferred' combinations of features have predictable geographies.
  - **extend** the preceding model to the case of non-independent features.

Correlations between features

Word-order features

Do combinations of features have associated geo-spatial patterning?

Correlations between features

Word-order features

Do combinations of features have associated geo-spatial patterning?

- **Proof of concept for this talk**: word order features as in WALS, Dryer (2013), etc.
  - Well-known and well-established as a paradigmatic example of *typologists'* features of this type that are strongly interdependent — certain combinations of features are disproportionately likely to be over- or under-represented.
    - As, of course, in various of the Greenberg (1968) universals ...
      1. "In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object."
      2. "In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes."
      3. "Languages with dominant VSO order are always prepositional."
      4. "With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional."
      5. "If a language has dominant SOV order and the genitive follows the governing noun, then the adjective likewise follows the noun."
      6. "All languages with dominant VSO order have SVO as an alternative or as the only alternative basic order."
  - **And** in the long history of both typological work on word-order & syntactic work on head-directionality, harmony ...

Correlations between features

Word-order features

Do combinations of features have associated geo-spatial patterning?

- **Proof of concept for this talk**: word order features as in WALS, Dryer (2013), etc.
  - Well-known and well-established as a paradigmatic example of *typologists'* feature
    interdependent — certain combinations of features are disproportionately likely t
    - As, of course, in various of the Greenberg (1968) universals ...
      1. "In declarative sentences with nominal subject and object, the dominant order is
         the object."
      2. "In languages with prepositions, the genitive almost always follows the governing
         almost always precedes."
      3. "Languages with dominant VSO order are always prepositional."
      4. "With overwhelmingly greater than chance frequency, languages with normal SO
      5. "If a language has dominant SOV order and the genitive follows the governing noun, then the adjective likewise follows the noun."
      6. "All languages with dominant VSO order have SVO as an alternative or as the only alternative basic order."
  - **And** in the long history of both typological work on word-order & syntactic work on head-directionality, harmony ...

81 Order of Subject, Object and Verb
82 Order of Subject and Verb
83 Order of Object and Verb
84 Order of Object, Oblique, and Verb
85 Order of Adposition and Noun Phrase
86 Order of Genitive and Noun
87 Order of Adjective and Noun
88 Order of Demonstrative and Noun
89 Order of Numeral and Noun
90 Order of Relative Clause and Noun

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(*Sandwich Conjecture*)**

- Head-directionality (taken as a composite property) is fairly *phylogenetically* stable. It is also a canonical example of typologists' harmony (Dryer 1992): all head-complement order tends to match the order of V and O within a given language.
  - Cases in which headedness-related properties **don't match phylogenetic predictions** tend to be attributed in the literature to **contact effects**: **Indic**, which is more rigidly OV than predicted due to Dravidian contact (Ledgeway & Roberts 2017); **Iranian**, where Persian is prepositional, Adj-N, and has head-initial relative clauses, but retains OV order and pre-head quantifiers (…etc…) plausibly due to Turkic (Harris & Campbell 1995).
    - Obvious idea: can we claim that Persian headedness is messy because it can 'see' lots of OV? (Not new.)
- **Prediction. The environments of 'dispreferred types' (messy macroparameters …) are more varied than 'default'.**

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. (***Sandwich Conjecture***)

- Is all this testable? **Question.** How do we measure 'diversity of geographical environment'? **Further question.** Is there a measure of the inherent correlatedness of individual features that it's worth thinking about (data-up, rather than theory-down)?

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. ***(Sandwich Conjecture)***

- Measure of correlation between variables: ϕ (see eg. Jäger & Wahle 2021 for more on this in a linguistic context), which for two features f_1, f_2:

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(***Sandwich Conjecture***)**

- Measure of correlation between variables: ϕ (see eg. Jäger & Wahle 2021 for more on this in a linguistic context), which for two features f_1, f_2:

	$f_1 = 0$	$f_1 = 1$
$f_2 = 0$	?	?
$f_2 = 1$	?	?

count observations, highest if most of the observations fall along the diagonal (feature values match often).

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(Sandwich Conjecture)**

- **Measure of the 'amount of stuff' in the geographical environment.** Slightly more challenging!
  - **Intuition.** What we are searching for is a measure of *neighbourhood variability* = *entropy*: for a selected language (individual cell), we want to know whether its nearest neighbours are relatively **homogeneous** (low-entropy), or relatively **heterogeneous** (high-entropy).
    - For each language *v*, we can use the information-theoretic (Shannon) entropy

$$H_v(i) = -\sum_{j \in I} p_v(j) \log \left(p_v(j) \right) ,$$

where *p* gives the probability of the *j*th type in the neighbourhood of *v*.

- But we're really looking for a property of a 'type' (combination of features) — the *mean* entropy averaged over all languages of that type.

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(*Sandwich Conjecture*)**

- **Measure of the 'amount of stuff' in the geographical environment.** Slightly more challenging!
  - **Intuition.** What we are searching for is a measure of *neighbourhood variability* = *entropy*: for a selected language (individual cell), we want to know whether its nearest neighbours are relatively **homogeneous** (low-entropy), or relatively **heterogeneous** (high-entropy).
    - For each language *v*, we can use the information-theoretic (Shannon) entropy

Assuming that for each combination of features we then have a quantity *sandwichness* D = entropy of dispreferred types - entropy of preferred types, D > 0 supports the Sandwich Conjecture ...

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(*Sandwich Conjecture*)**

Assuming that for each combination of features we then have a quantity *sandwichness* D = entropy of dispreferred types - entropy of preferred types, D > 0 supports the Sandwich Conjecture ...

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(**Sandwich Conjecture**)**

Assuming that for each combination of features we then have a quantity *sandwichness* D = entropy of dispreferred types - entropy of preferred types, D > 0 supports the Sandwich Conjecture …

- **Issue.** What if one type is vastly overrepresented? If the frequencies of the different types are very dissimilar, then even a random distribution of types over languages is not guaranteed to give D = 0 (more frequent types are more likely to be surrounded by themselves, so their neighbourhood entropies can be expected to be slightly lower).
- One brute-force solution: carry out a permutation test by repeatedly recalculating D over randomly-generated sets of languages. This gives us an idea of what kinds of values of D to expect under the assumption that *types* are just randomly "thrown" onto the set of languages (and then something against which to compare our empirical D).

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(*Sandwich Conjecture*)**

Assuming that for each combination of features we then have a quantity *sandwichness* D = entropy of dispreferred types - entropy of preferred types, D > 0 supports the Sandwich Conjecture …

1. Calculate D from the original dataset.
2. Permute s, the function that assigns 'types' to languages.
3. Calculate D from the permuted dataset.
4. Repeat 2 and 3 many times.

Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(*Sandwich Conjecture*)**

Assuming that for each combination of features we then have a quantity *sandwichness* D = entropy of dispreferred types - entropy of preferred types, D > 0 supports the Sandwich Conjecture ...

- **Quick illustration.** The full result of this procedure for 3 WALS features: 83A, OV vs. VO, 85A, prepositions vs. postpositions, & 4A, obstruent voicing contrast.

Correlations between features

Hypothesis testing

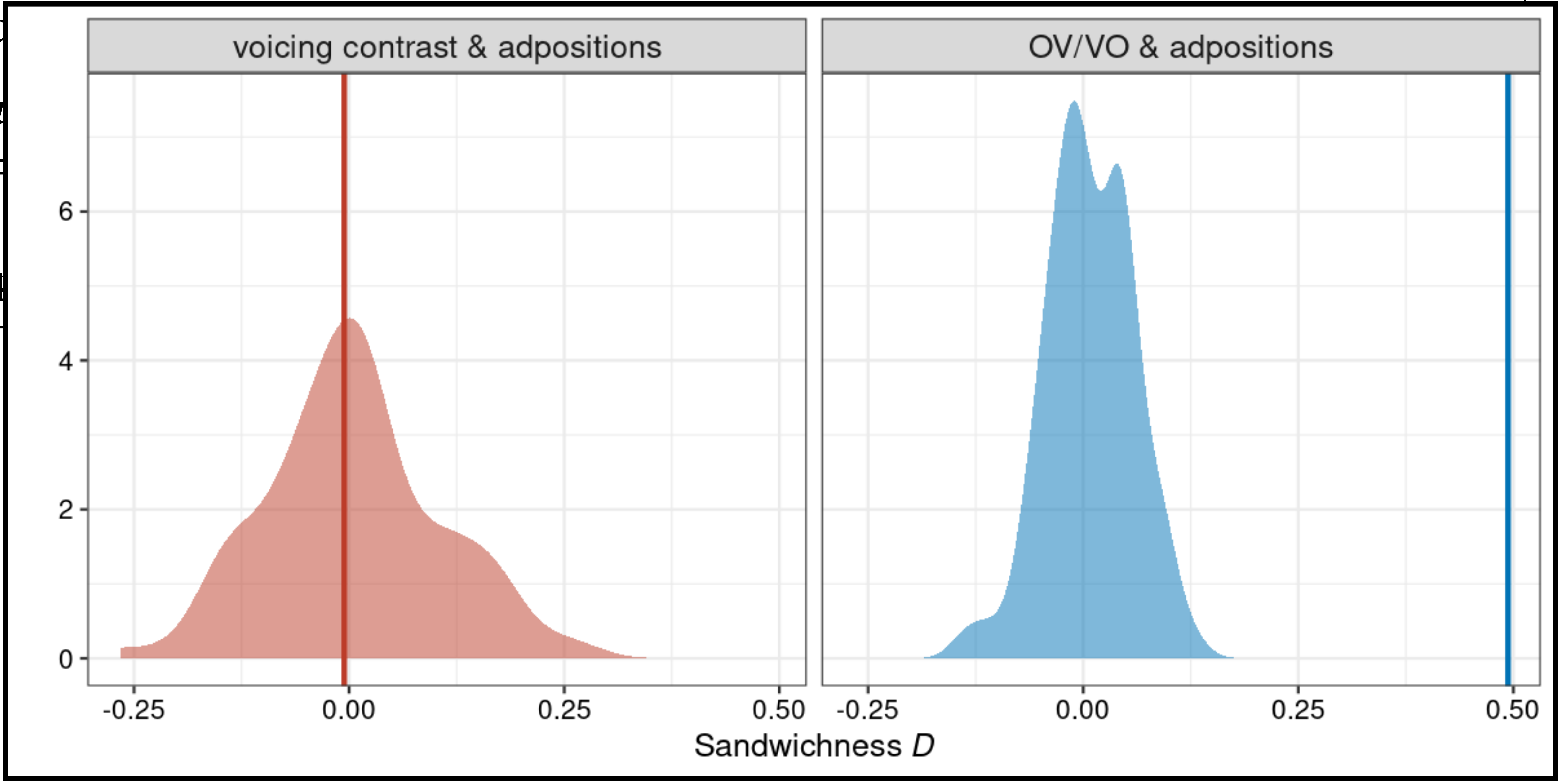
Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a c

Dispreferred 'types' should tend to be s

Assuming that for each combination
entropy of p

- **Quick illustration.** The full result of this procedure for 3 WALS features: 83A, OV vs. VO, 85A, prepositions vs. postpositions, & 4A, obstruent voicing contrast.



Correlations between features

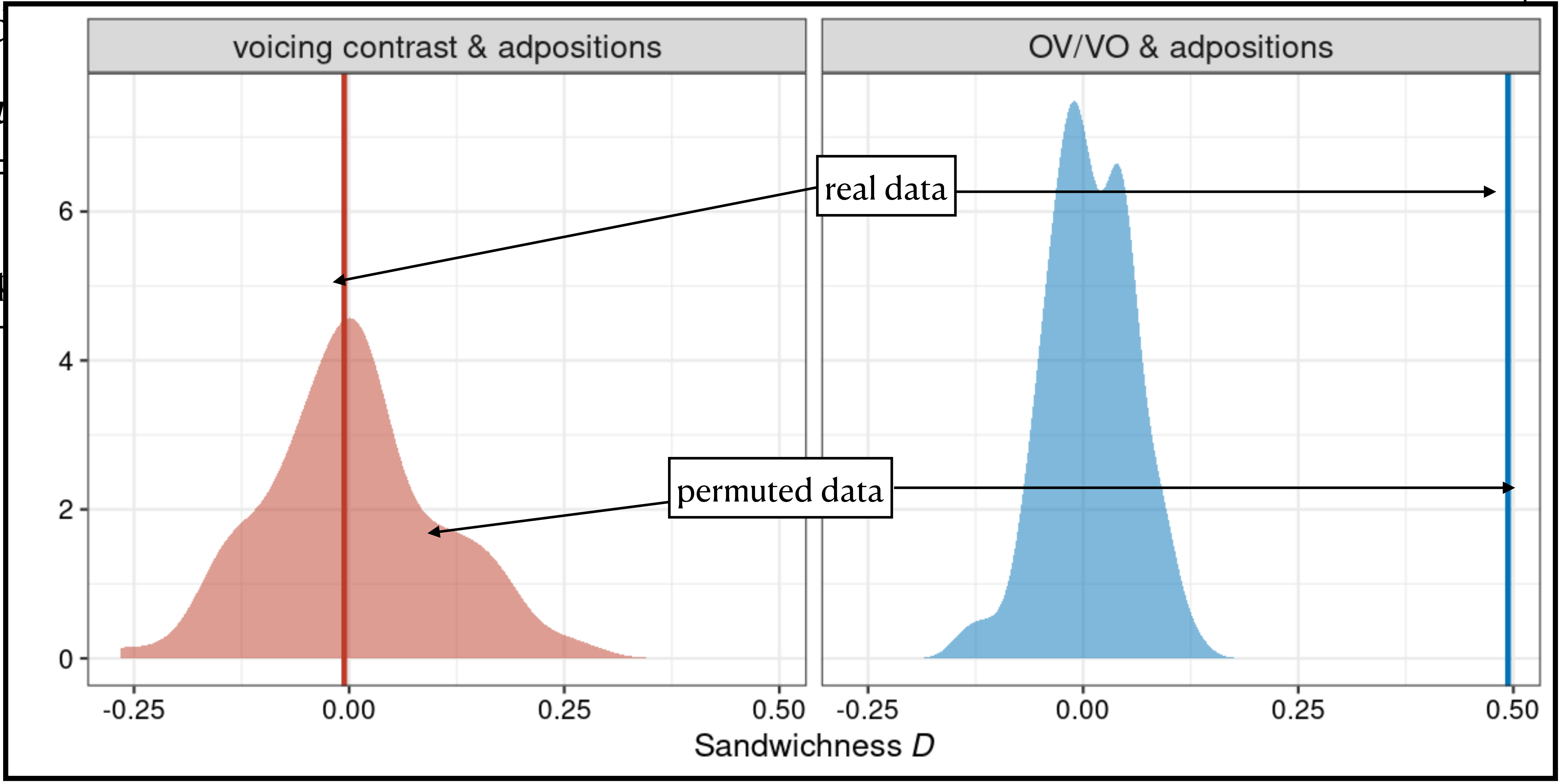
Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a c

Dispreferred 'types' should tend to be st

Assuming that for each combination
entropy of p



- **Quick illustration.** The full result of this procedure for 3 WALS features: 83A, OV vs. VO, 85A, prepositions vs. postpositions, & 4A, obstruent voicing contrast.

Correlations between features

Hypothesis testing

Do combinations of f … **o-**

Intuition: the stability of a c…

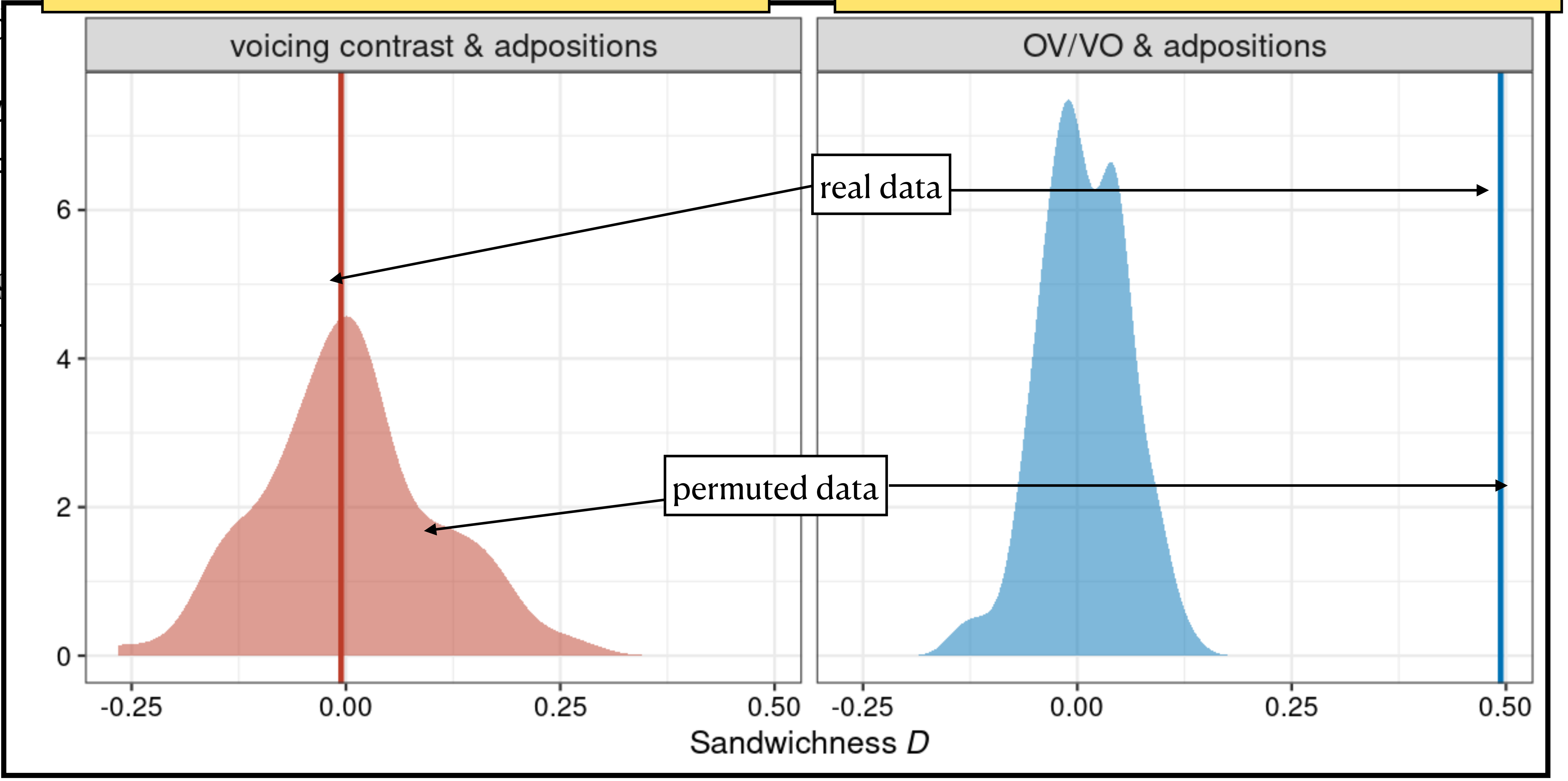
Dispreferred 'types' should tend to be st…

Assuming that for each combination… entropy of p…

- **Quick illustration.** The full result of this procedure for 3 WALS features: 83A, OV vs. VO, 85A, prepositions vs. postpositions, & 4A, obstruent voicing contrast.

empirical line looks exactly like the median of the randomised distribution — these 'types' can't be distinguished from the 'random universe'.

empirical line several standard deviations away from the random distribution — these 'types' can be distinguished from the 'random universe'.



Correlations between features

Hypothesis testing

Do combinations of features have associated geo-spatial patterning?

Intuition: the stability of a dispreferred type can be enhanced in certain configurations of contact.

Dispreferred 'types' should tend to be surrounded by a greater variety of types than preferred 'types'. **(*Sandwich Conjecture*)**

Assuming that for each combination of features we then have a quantity *sandwichness* D = entropy of dispreferred types - entropy of preferred types, D > 0 supports the Sandwich Conjecture …

- **Larger illustration.** The result of this procedure for all the WALS word-order features vs. 4A, obstruent voicing contrast, looking at *z-score* D (empirical - null) only.

Correlations between features

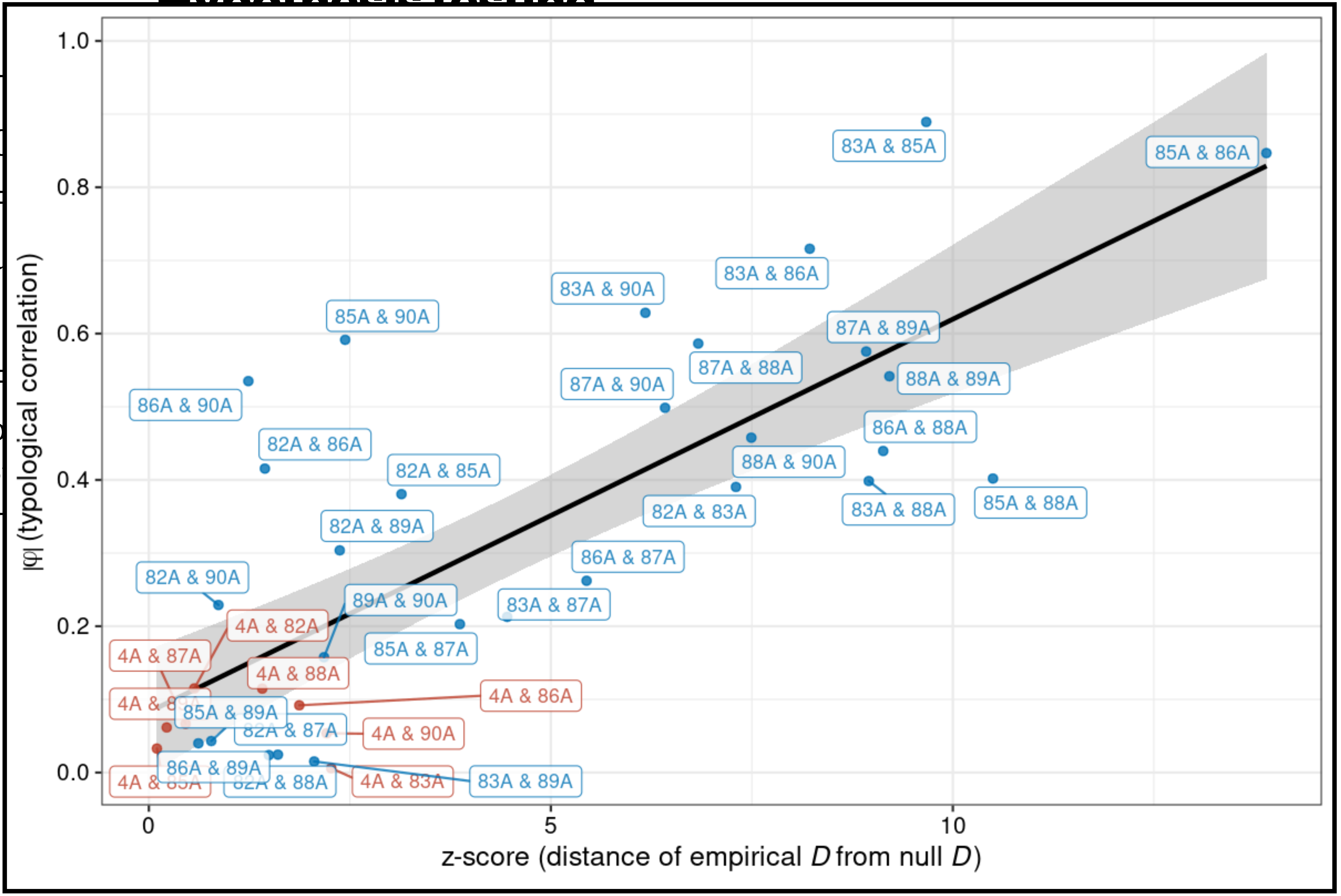
Hypothesis testing

Do combinations c

Intuition: the stability of a

Dispreferred 'types' should tend to be

Assuming that for each combinatio
entropy o



- **Larger illustration.** The result of this procedure for all the WALS word-order features vs. 4A, obstruent voicing contrast, looking at *z-score* D (empirical - null) only.

Correlations between features

Hypothesis testing

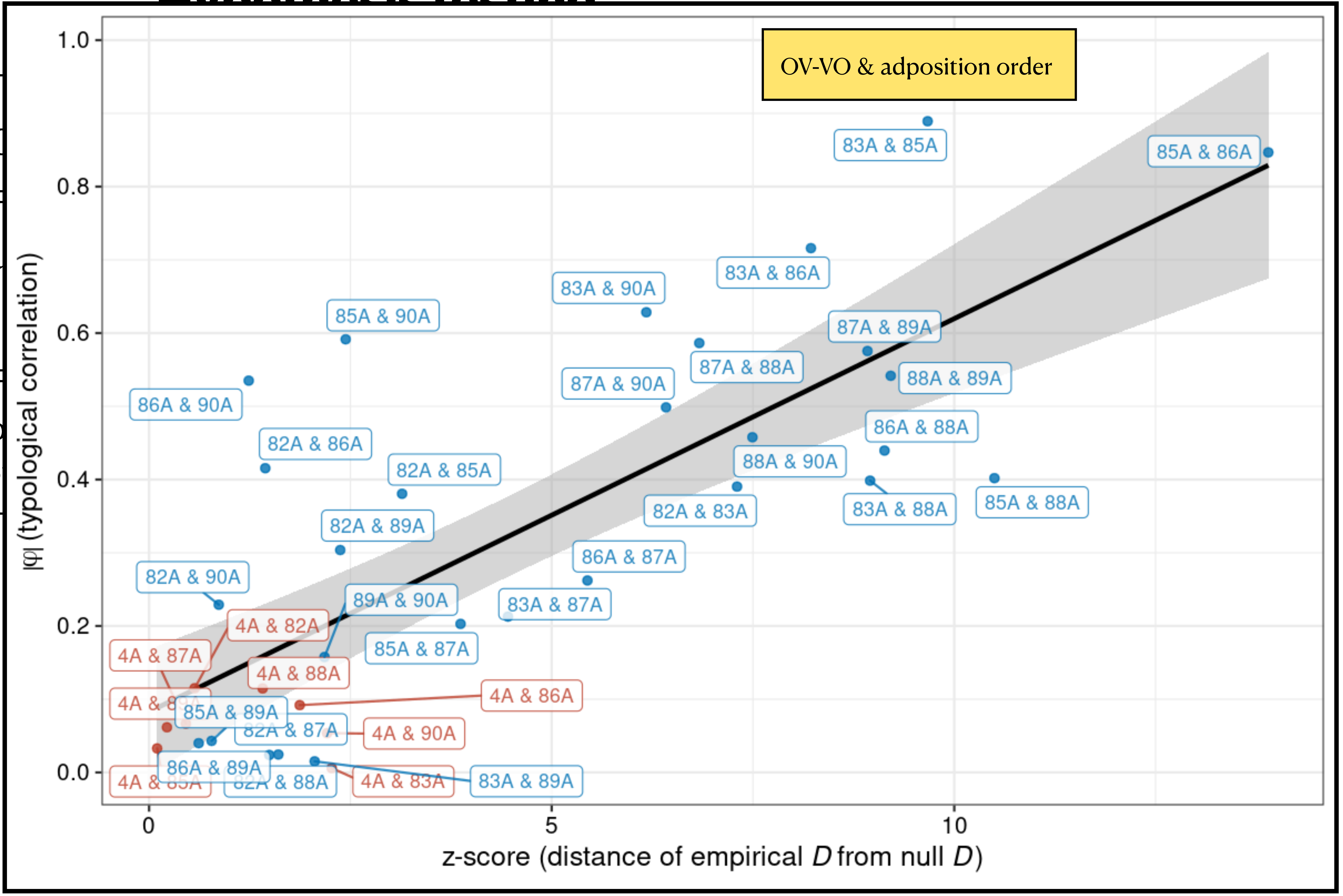
Do combinations o

Intuition: the stability of a

Dispreferred 'types' should tend to be

Assuming that for each combinatio
entropy o

- **Larger illustration.** The result of this procedure for all the WALS word-order features vs. 4A, obstruent voicing contrast, looking at *z-score* D (empirical - null) only.



Conclusions & outlook

The dynamics of multiple-feature interactions

- **Essential point of this talk.** It's nice to be able to frame typological facts that 'everyone knows' in ways that allow us to think about **the emergent properties of simpler dynamics.**
  - **Empirical spatial distributions have surprisingly predictable relationships to intuitions about actual linguistic properties.**
  - **'Horizontal' dynamics really matter …**

Conclusions & outlook

- For Kauhanen, Gopal, Galla, & Bermúdez-Otero (2021, Science Advances):

- For technical details, refs., etc. email me deepthi.gopal@lingfil.uu.se
- Thanks!