Case alignment across the lexicon

Balthasar Bickel
Alena Witzlack-Makarevich
Taras Zakharko
(U. of Leipzig)

Syntax of the World’s Languages IV
Lyon 23-26 September 2010

Alignment

- Both for descriptive and typological purposes, one of the standard characterizations of case (and other) systems is their **alignment**, i.e. which arguments receive the same case markers (incl. adpositions).

- Possible alignment types:
    - S=A=P - neutral,
    - S≠A≠P - tripartite,
    - S=A≠P - accusative,
    - S=P≠A - ergative,
    - S≠A=P - horizontal (double-oblique)

Alignment

- E.g. **Chechen** (Nakh-Daghestanian), *non-continuous aspect*:

S

so	*ohw-v-uzh-u.*
1s**NOM**	down-V-fall-PRS

'I fall.'

A **P**

as	*wazh*	*b-u'-u.*
1s**ERG**	apple(B).**NOM**	B-eat-PRS

'I eat apples.'

→ **S=P≠A** - ergative alignment

Alignment splits

- Known complications: a single language can have multiple alignments (*splits*), conditioned by:

- tense or aspect (e.g. past vs. non-past),
- clause type (e.g. finite vs. non-finite),
- referential properties of arguments (e.g. SAP vs. non-SAP),

  etc.

- Solution: alignment type is established for individual subsystems

Alignment splits: an example

E.g. Chechen *continuous aspect*

S

so	*ohw-v-uzh-ush*	*v-u.*
1s**NOM**	down-V-fall-CVB.SIM	V-eat-PRS

‘I’m falling.’

A **P**

so	*wazh-sh*	*b-u’-ush*	*v-u.*
1s**NOM**	apple(B).**NOM**-PL	B-eat-CVB.SIM	V-eat-PRS

‘I’m eating apples.’

→ **S=A=P** - neutral alignment

- Alignment type of case marking is established for individual subsystems:

- non-continuous: S=P≠A (ergative)
- continuous: S=A=P (neutral)

Alignment and predicates

- But there is a problem that goes far beyond this:

  in many languages, **lexical predicates vary in their case frames**, e.g.

  Chechen (Nakh-Daghestanian):

- one-argument predicates:

  $<S_{\text{NOM}}>$, $<S_{\text{DAT}}>$, $<S_{\text{ERG}}>$, $<S_{\text{ALL}}>$

- two-argument predicates:

  $<A_{\text{ERG}}\ P_{\text{NOM}}>$, $<A_{\text{DAT}}\ P_{\text{NOM}}>$, $<A_{\text{GEN}}\ P_{\text{NOM}}>$,

  $<A_{\text{NOM}}\ P_{\text{ALL}}>$, $<A_{\text{ERG}}\ P_{\text{LAT}}>$, etc.

- three-argument predicates:

  $<A_{\text{ERG}}\ T_{\text{NOM}}\ G_{\text{ALL}}>$, $<A_{\text{ERG}}\ T_{\text{NOM}}\ G_{\text{DAT}}>$, $<A_{\text{ERG}}\ T_{\text{LAT}}\ G_{\text{NOM}}>$,

  $<A_{\text{ERG}}\ T_{\text{INS}}\ G_{\text{ALL}}>$

Alignment and predicates

- Which arguments can be compared to determine the alignment type of a language system (e.g. case marking)?

Chechen non-continuous aspect

- *ohwad.ouzha* ‘fall down’ (SNOM)

so	ohw-v-uzh-u.
1s**NOM**	down-V-fall-PRS

‘I fall.’

- *d.aa* ‘eat’ (AERG PNOM)

as	wazh	b-u’-u.
1s**ERG**	apple(B).**NOM**	B-eat-PRS

‘I eat apples.’

→ **S=P≠A** - ergative alignment

Alignment and predicates

- Or some other one- and two-argument predicates:

- *jovxa d.aalla* ‘be hot’ (S_{DAT})

S		
suuna	jovxa	j-u.
1s**DAT**	hot	J-be.PRS

‘I’m hot.’

- *d.ieza* ‘love’ (A_{DAT} P_{NOM})

A	**P**	
suuna	Zaara	j-eez-a.
1s**DAT**	Zara(J).**NOM**	J-love-PRS

‘I love Zara.’

> **S=A≠P** - accusative alignment

Alignment and predicates

- Or yet another pair of one-argument and two-argument predicates:

- *jovxa d.aalla* ‘be hot’ (S_{DAT})

  S

suuna	jovxa	j-u.
1s**DAT**	hot	J-be.PRS

  ‘I’m hot.’

- *d.aa* ‘eat’ (A_{ERG} P_{NOM})

  A **P**

as	wazh-ash	b-u’-u.
1s**ERG**	apple(B).**NOM**-PL	B-eat-PRS

  ‘I eat apples.’

  > **S≠A≠P** - tripartite alignment

Alignment and predicates

- ➡ Alignment varies depending on which predicates (or generalized predicate classes) we compare!

two-argument → one-argument ↓	**A**$_{\text{ERG}}$ **P**$_{\text{NOM}}$ (*d.aa* ‘eat’)	**A**$_{\text{DAT}}$ **P**$_{\text{NOM}}$ (*d.ieza* ‘love’)	**A**$_{\text{ERG}}$ **P**$_{\text{DAT}}$ (*mohw tuoxa* ‘call’)	…
S$_{\text{NOM}}$ (*ohwad.ouzha* ‘fall down’)	**S=P≠A**	**S≠A≠P**	**S≠A≠P**	…
S$_{\text{DAT}}$ (*jovxa d.aalla* ‘be hot’)	**S≠A≠P**	**S=A≠P**	**S=P≠A**	…
…	…	…	…	…

Alignment and predicates

- The standard solution in the past has been to side-step the problem and concentrate on an *a priori* defined prototype or canon.

- The prototype or canon can be defined on the basis of
  - a gut feeling,
  - productivity,
  - token frequency,
  - type frequency,
  - a combination of criteria, e.g.
    - ✦ type frequency for one- and two-argument predicates and
    - ✦ a gut feeling for three-argument predicates (take ‘give’)

A priori defined prototype

- Chechen exhibits S=P≠A (e.g. if default is defined by type frequency)
- The rest are exceptions and can be ignored

two-argument → one-argument ↓	**AERG PNOM** **(default)** (*d.aa* ‘eat’)	**ADAT PNOM** (*d.ieza* ‘love’)	**AERG PDAT** (*mohw tuoxa* ‘call’)	…
SNOM (default) (*ohwad.ouzha* ‘fall down’)	**S=P≠A**	S≠A≠P Ignore!	S≠A≠P Ignore!	…
SDAT (*jovxa d.aalla* ‘be hot’)	S≠A≠P Ignore!	S=A≠P Ignore!	S=P≠A Ignore!	…
…	…	…	…	…

A priori defined prototype

- **Unsatisfactory:**
  - artificial reduction of the true variation,
  - arbitrary decisions on what counts as the prototype,
  - missing typologically interesting distributional patterns,
  - misinterpreting or simplifying patterns of language change (cf. Butt 2001 on the accusative to ergative shift in Indo-Aryan)
- **What is the alternative?**

Exhaustive alignment

➡ **Consider ALL predicates: “exhaustive alignment”**

- For this, define macro-argument roles S, A, P, T, and G exclusively by semantic entailment criteria (e.g. Dowty 1991 or Primus 1999),
  - e.g. contributing properties of proto-A:
    - volitional involvement in the event or state;
    - sentience (and/or perception);
    - causing an event or change of state in another participant, etc.
- Compare the marking of macro-argument roles of individual lexical predicate classes:
  - reference to lexical predicate classes captures any lexical and language-specific idiosyncrasies, including semantic specializations (e.g. as experiencers)

Alignment and predicates

➡ Consider ALL predicates: “exhaustive alignment”

- This can be done for all pairs (or triples if the comparison is extended to three-argument verbs).
- e.g. for Chechen

two-argument → one-argument ↓	**AERG PNOM (default)** (*d.aa* ‘eat’)	**ADAT PNOM** (*d.ieza* ‘love’)	**AERG PDAT** (*mohw tuoxa* ‘call’)	…
SNOM (default) (*ohwad.ouzha* ‘fall down’)	**S=P≠A**	S≠A=P Ignore!	S≠A≠P Ignore!	…
SDAT (*jovxa d.aalla* ‘be hot’)	S≠A≠P Ignore!	S=A≠P Ignore!	S=P≠A Ignore!	…
…	…	…	…	…

Default vs. non-default predicate classes

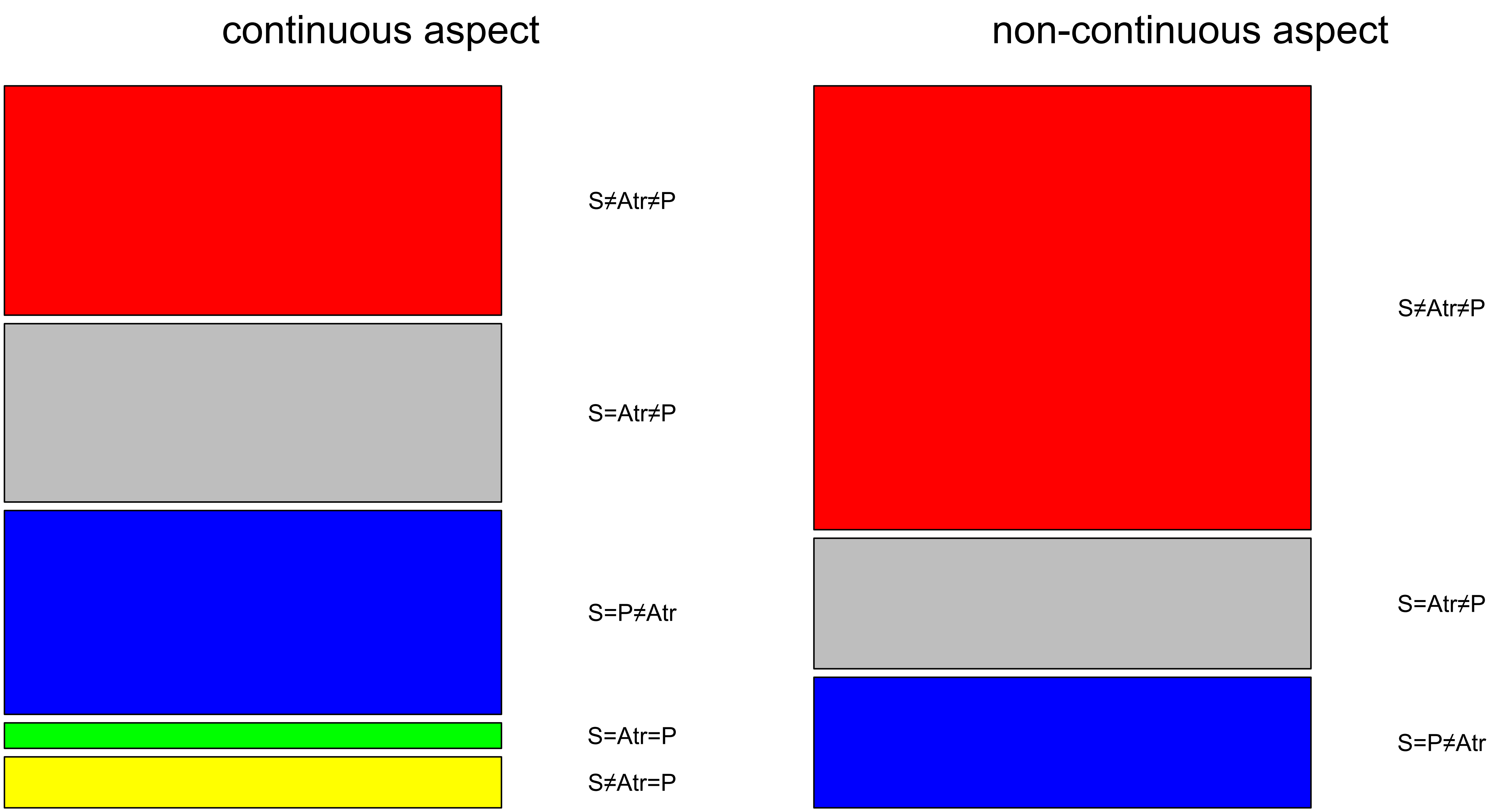
- For Chechen predicates:
  - 4 one-argument predicate classes
  - 7 two-argument predicate classes
  - 4 three-argument predicate classes
- Exhaustive alignment:

➡ $4 \times 7 =$ **28** alignment descriptions for one- and two-arg. predicates (in each aspect subsystem: CONT & NON-CONT)

➡ $4 \times 7 \times 4 =$ **112** alignment descriptions for all predicates (in each aspect subsystem: CONT & NON-CONT)

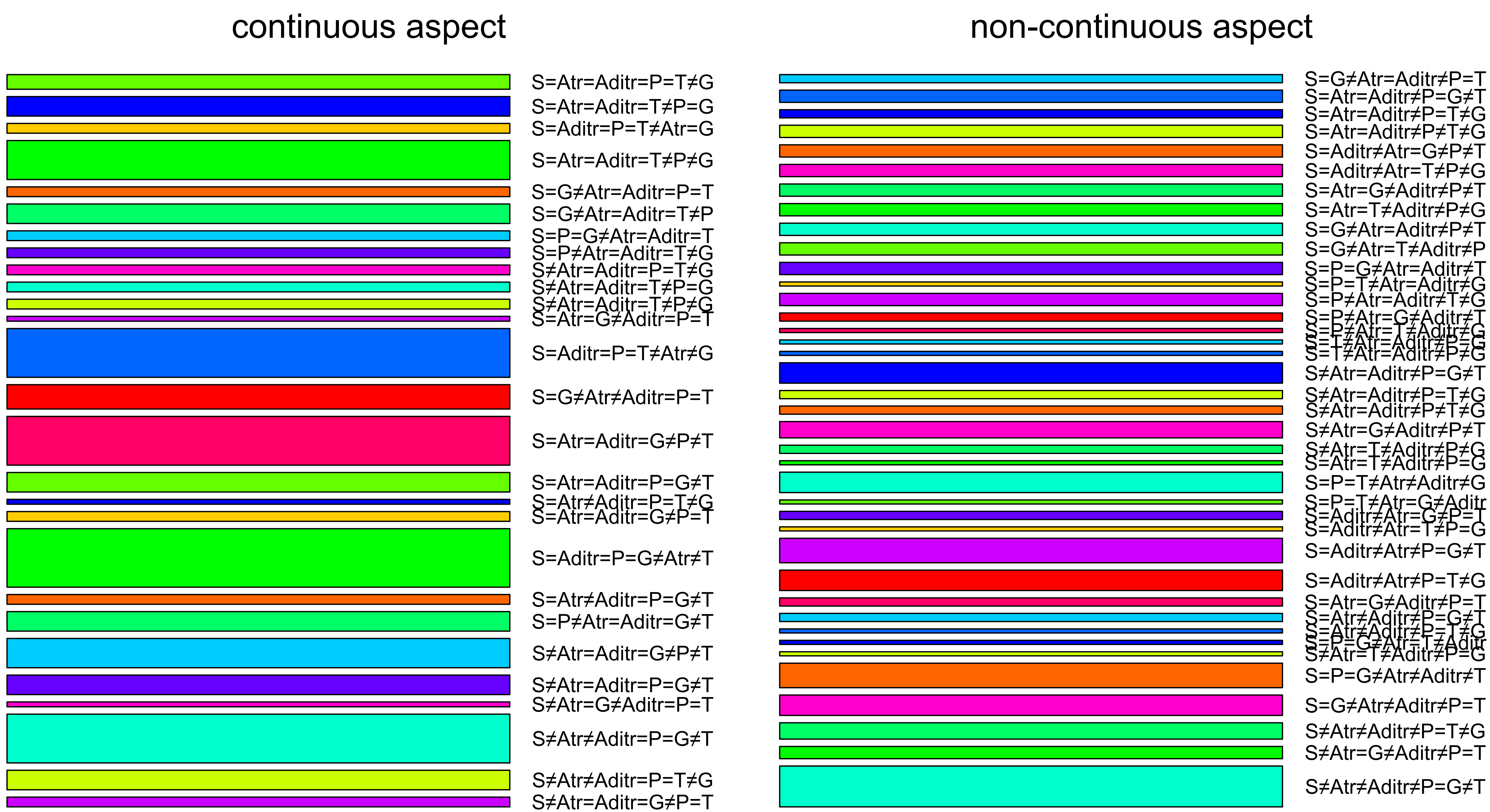
Exhaustive alignment in Chechen

- The true diversity of exhaustive alignment in Chechen (S, A, and P only):



Exhaustive alignment in Chechen

- The true diversity of exhaustive alignment in Chechen (S, A_{tr}, P, A_{ditr}, T, G):



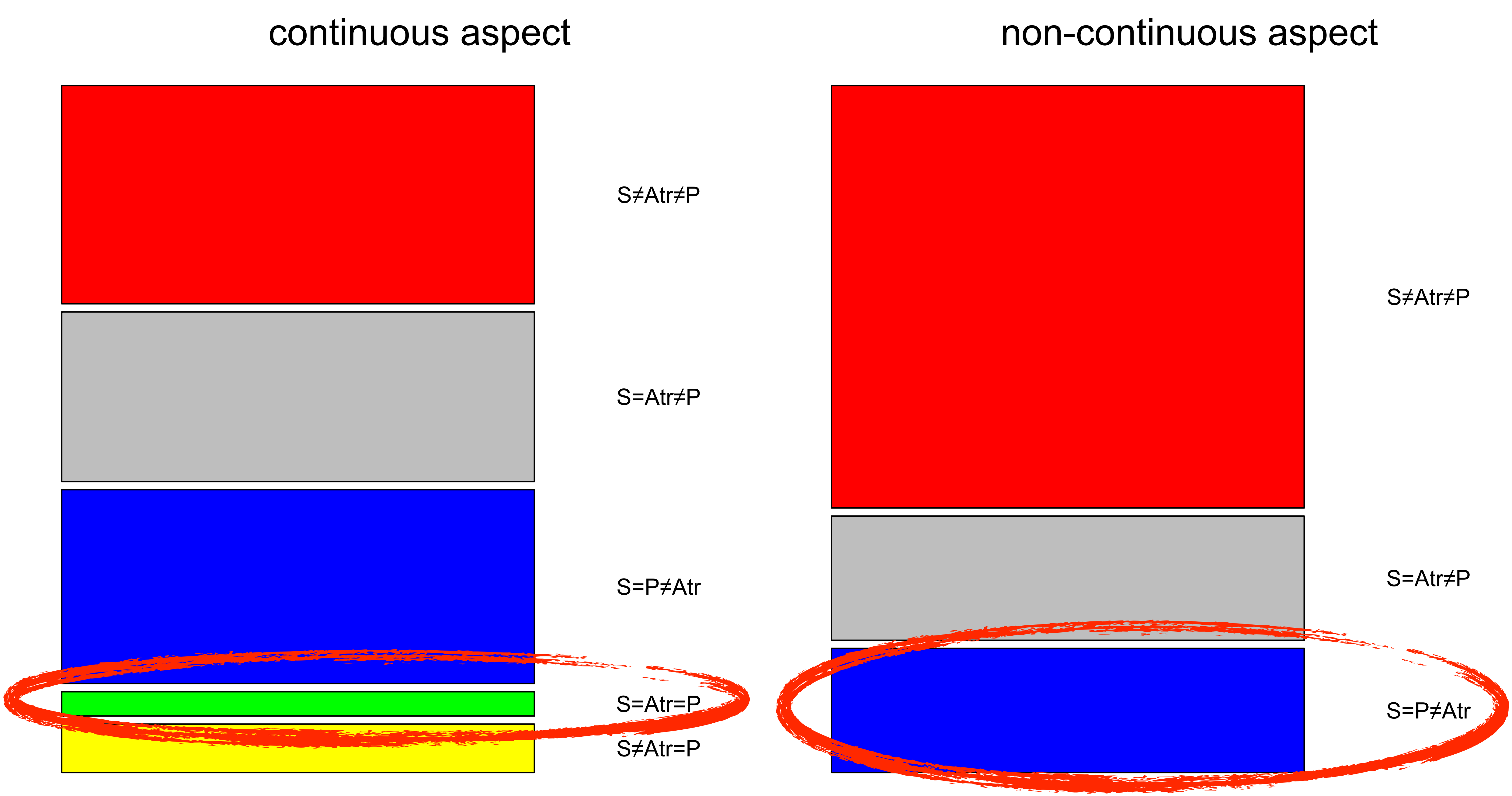
Default vs. non-default predicate classes

- Descriptive grammars and typologies often take the alignment of the default classes as representative of the whole language system (e.g. case marking).

- The most practicable — and, as argued by Bickel et al. (2010), the psycho-linguistically most realistic — definition of the default predicate classes is in terms of **type frequency** (i.e. largest class in the lexicon)
- **Is it really representative of the whole language system?**

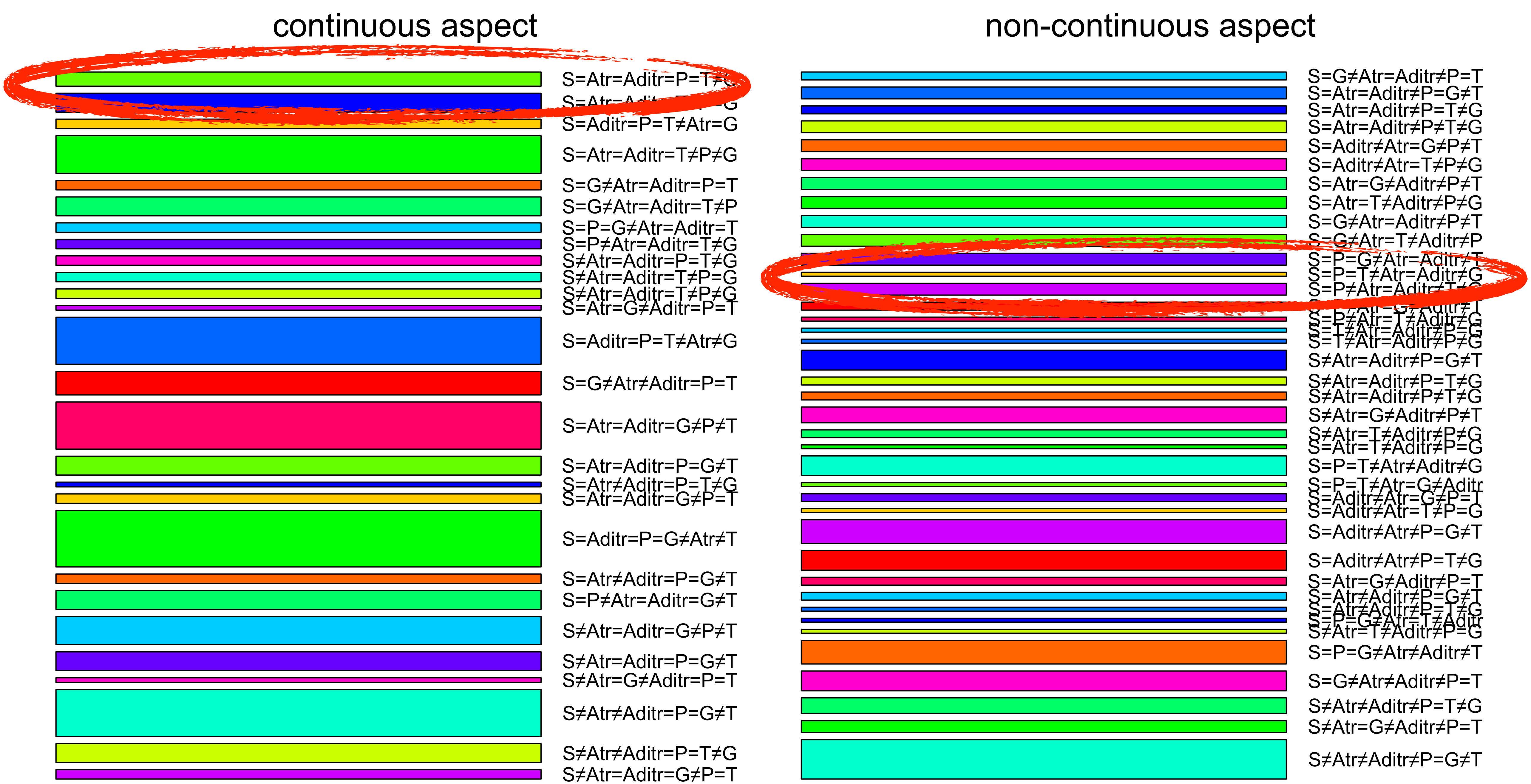
Default vs. non-default in Chechen

- The true diversity of exhaustive alignment in Chechen (S, A, and P only):



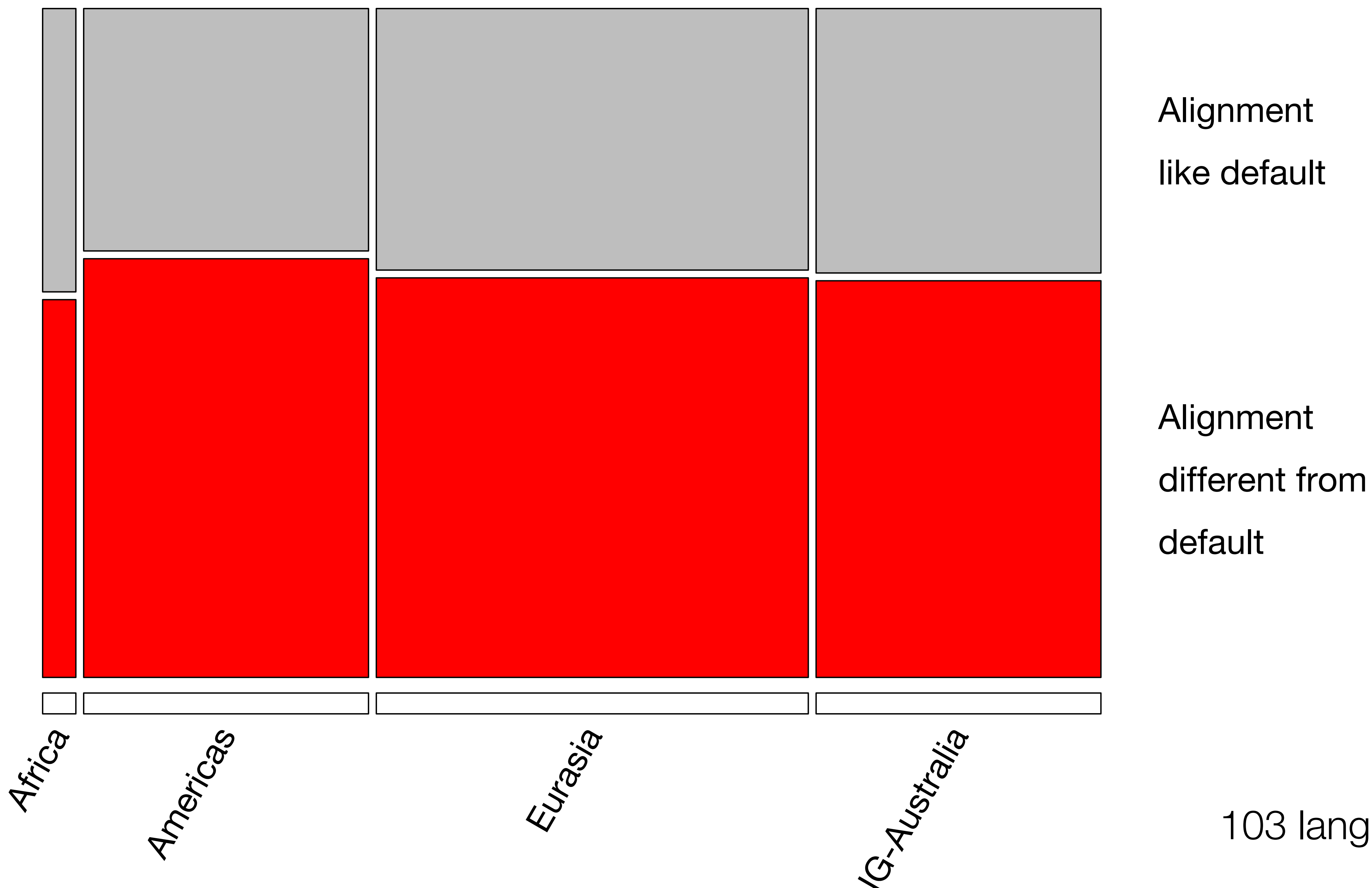
Default vs. non-default in Chechen

- The true diversity of exhaustive alignment in Chechen (S, A_{tr}, P, A_{ditr}, T, G):



Default vs. non-default: S, A, and P

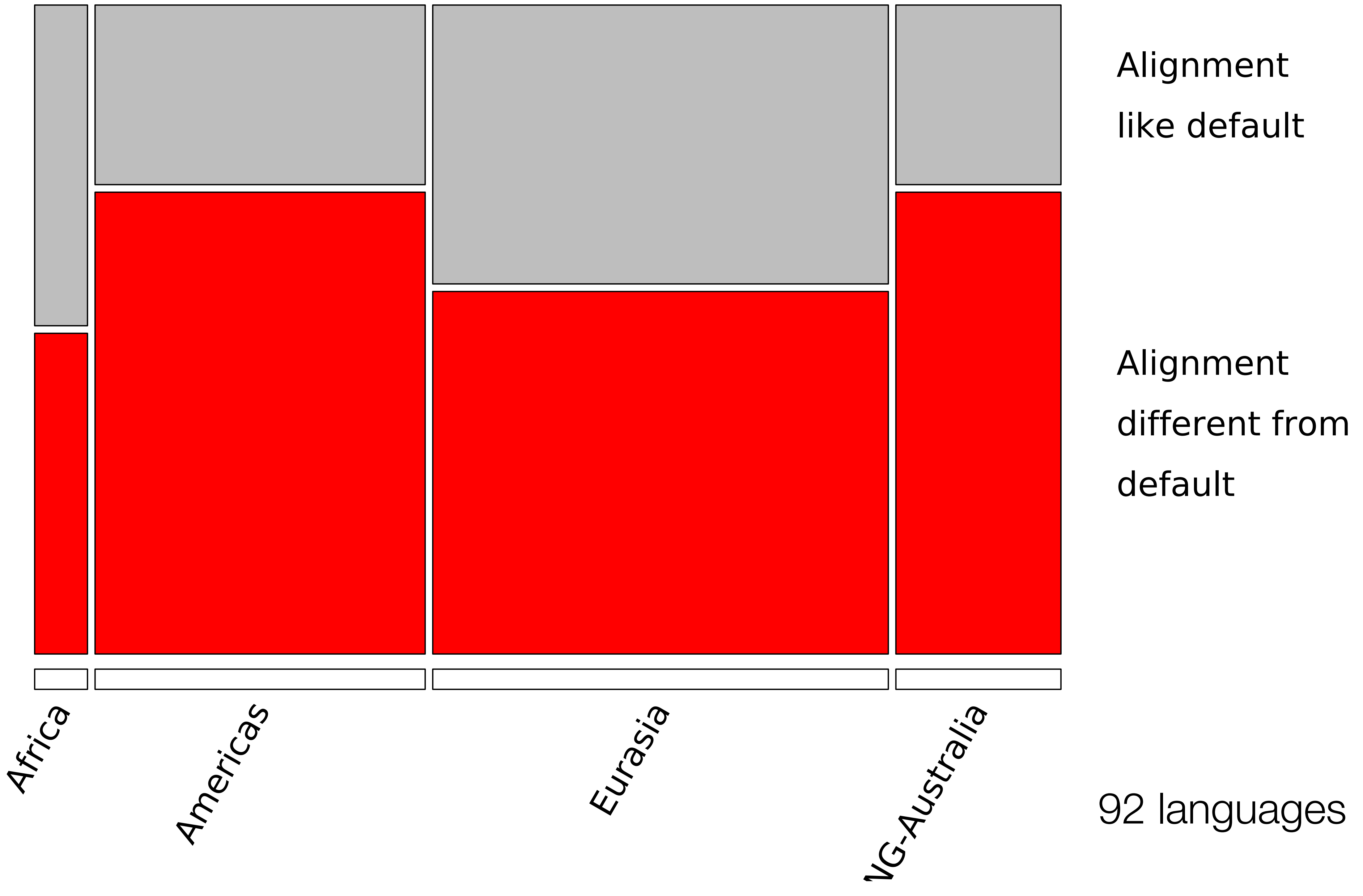
- Is the default predicate class really representative?



103 languages

Default vs. non-default: P, T, and G

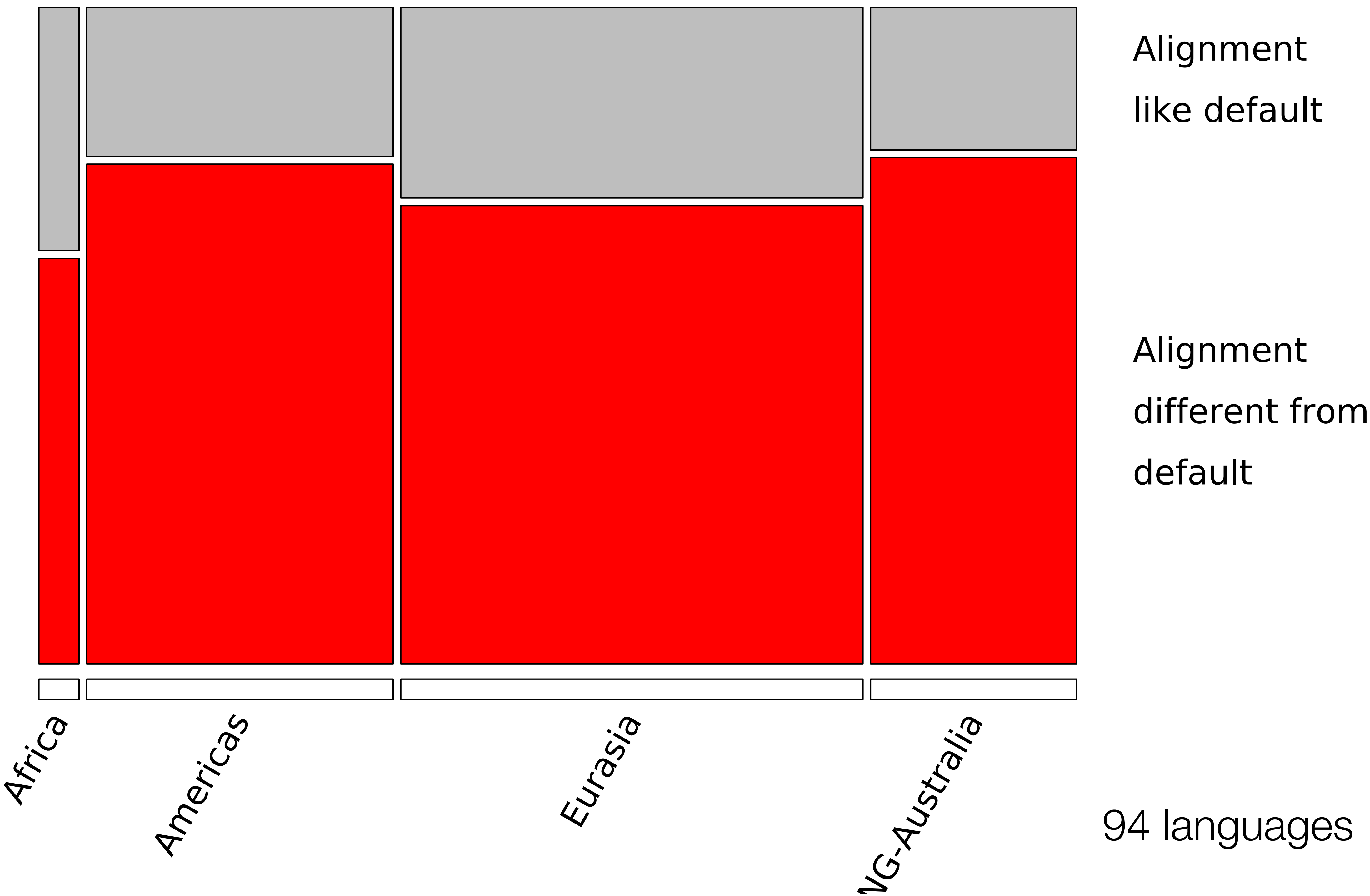
- Is the default predicate class really representative?



92 languages

Default vs. non-default: S, A, P, T, and G

- Is the default predicate class really representative?



94 languages

Default vs. non-default predicate classes

Conclusion:

- The alignment of default classes is **never** a good approximation of the exhaustive alignment (i.e. alignment on the basis of all predicate classes)!

A more substantial issue

- Looking at non-default classes is not only necessary for accurate description
- but also allows exploring the distribution and historical development of alignment types in their own right
- For this, we measure distances between alignment types ...

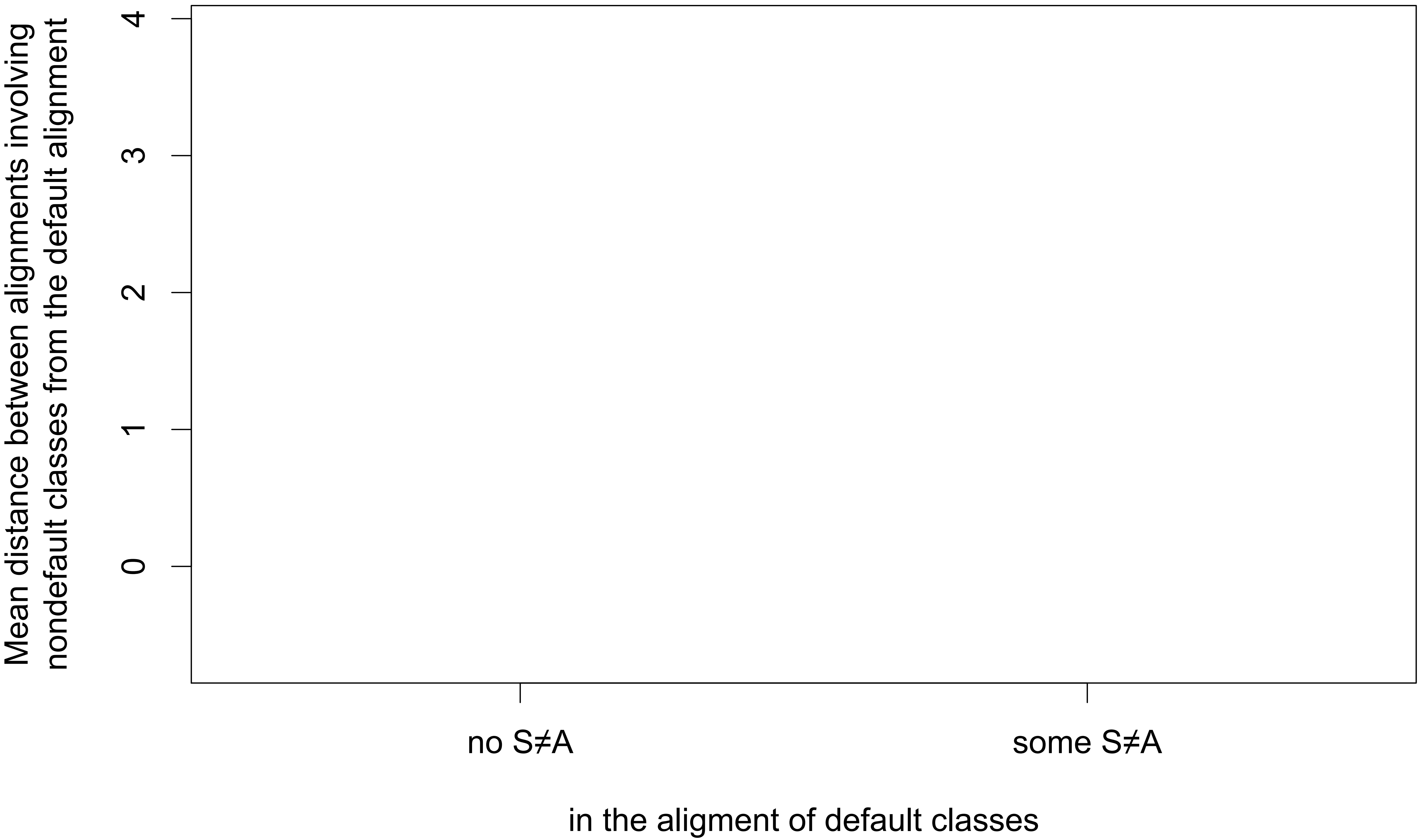
Distances between alignments

- Idea: some alignment types are more similar (historically closer than) to some other alignment types,
    - e.g.S=A=P is closer to S=A≠P than to S≠A≠P
- Levenshtein distance (for S, A, and P):

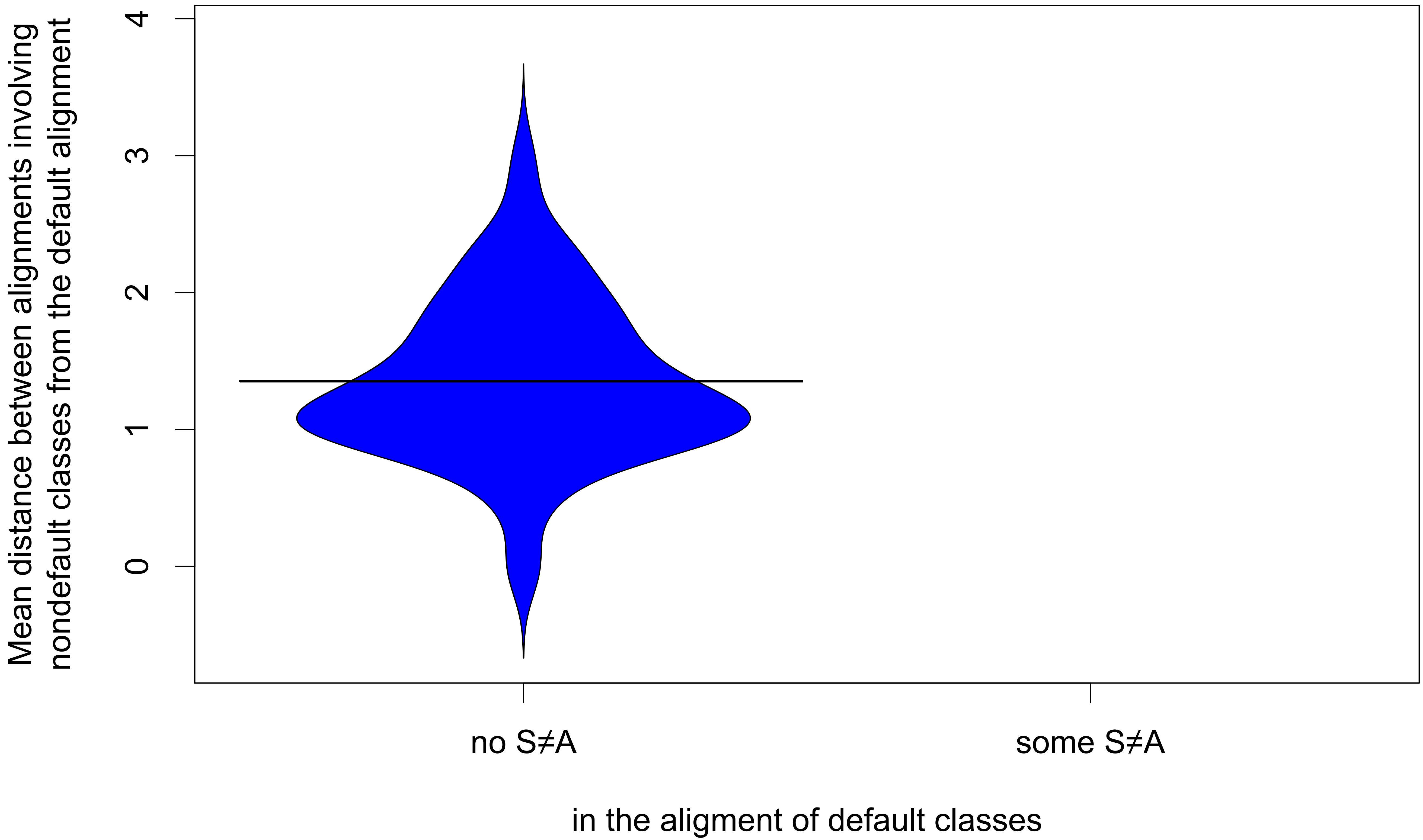
	S=A=P	S=A≠P	S=P≠A	S≠A=P	S≠A≠P
S=A=P	0	1	1	1	2
S=A≠P	1	0	2	2	1
S=P≠A	1	2	0	2	1
S≠A=P	1	2	2	0	1
S≠A≠P	2	1	1	1	0

- For S, $A_{tr/ditr}$, P, T, and G the Levenshtein distance can be up to 5.

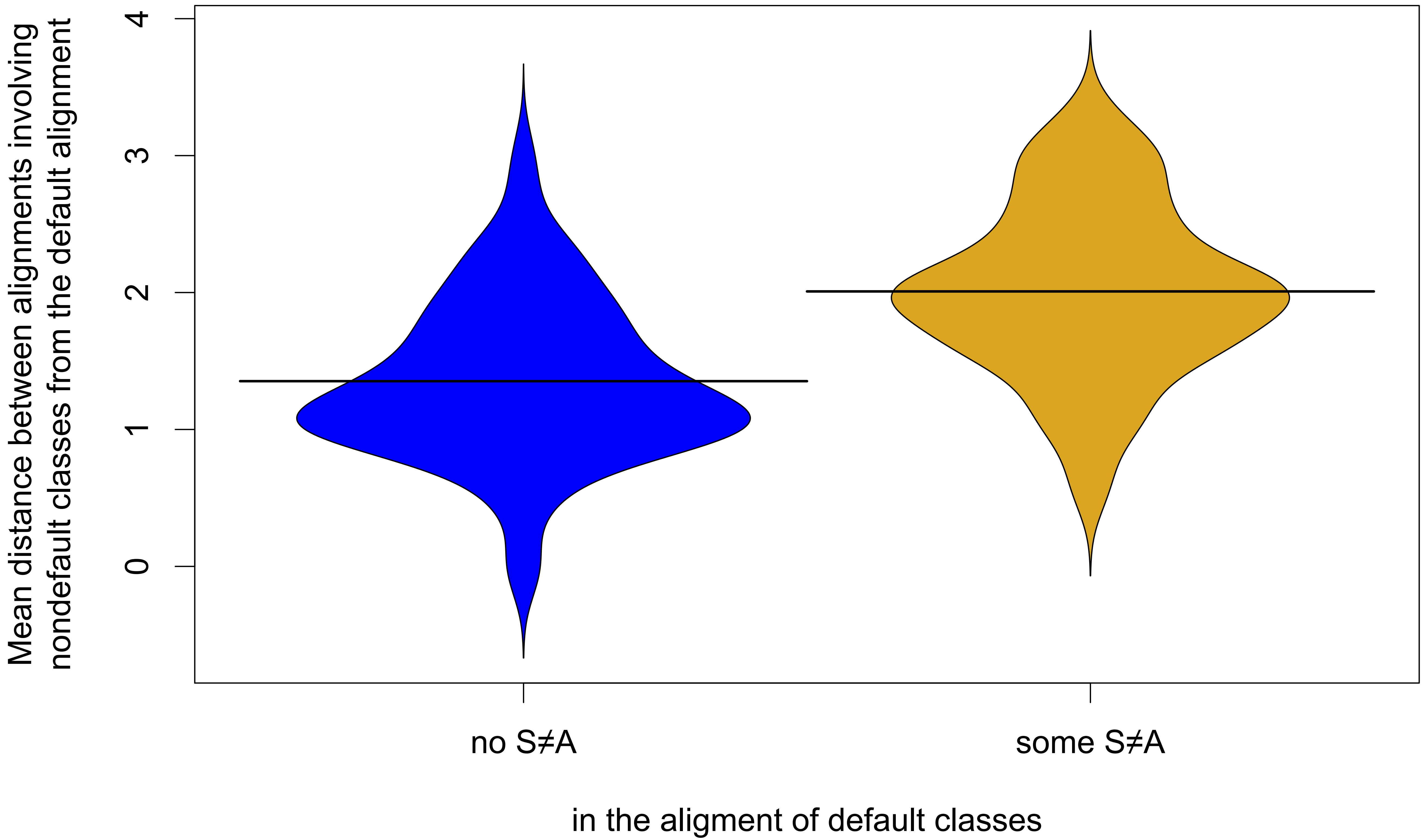
Correlation of distances with presence of ergativity



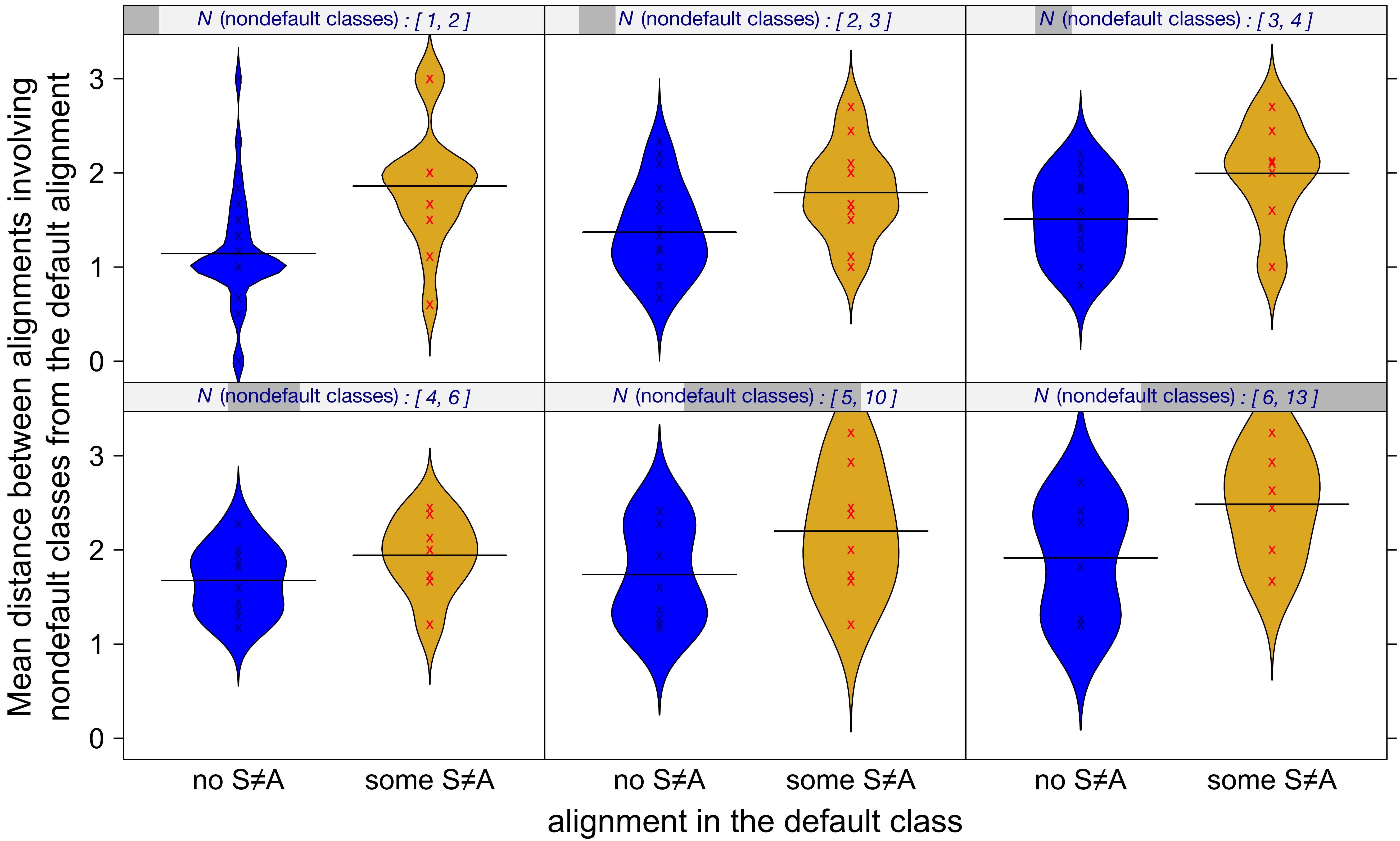
Correlation of distances with presence of ergativity



Correlation of distances with presence of ergativity



Correlation of distances with presence of ergativity



Correlation of distances with presence of ergativity

- The mean distance of non-default alignments to default alignments correlates
  - with the number of non-default classes, $F = 21.85$, $p < .001$ (not on the slides)
  - with the **presence of an ergative alignment** in the default predicate classes, $F = 25.87$, $p < .001$
  - but not with an interaction between these factors, $F = .005$, $p = .99$
- Suggests possible hypotheses to be further tested, with control of geography and genealogy and larger sample (current sample: 94 languages)

Discussion

- S≠A in the default favors more varied exhaustive alignment, and more varied exhaustive alignment favors S≠A in the default.
- This fits
  - with Nichols (1993) proposal that ergativity is “recessive” and therefore more prone to variation
  - and with Butt’s (2001) finding that S≠A traces in Indo-Aryan are a reflex of an extreme and continued variation in exhaustive alignment.
- Hypothesis for future research: a **Diachronic Universal:**
  - S≠A traces in a language increase alignment variation over time, and
  - alignment variation leads to the development and maintenance of S≠A traces over time

Conclusion

- Alignment patterns involving non-default predicate classes are essential
  - from a descriptive perspective because:
    - default classes (or any other “basic” predicate type) are **bad estimators** of the overall alignment in a language
  - from a theoretical perspective because:
    - they help understand how S≠A develops over time
    - they give insights on the recessive nature of S≠A in languages

Thank you!