Diversity in phonological domains

Balthasar Bickel & Kristine A. Hildebrandt

University of Leipzig

www.uni-leipzig.de/~autotyp

The prosodic hierarchy

μ	Mora	no direct reference to morphological structure
σ	Syllable	
Φ	Foot	
ω	Word: direct reference to morphological structure, at most one stem	
PhP	Phonological phrase: reference to more than one stem and/or syntactic phrases	
IntP	Intonation Phrase: multiple PhP	
U	Utterance	

Booij 1983, Selkirk 1984; Nespor & Vogel 1986, McCarthy & Prince 1993, Hall 1999, Peperkamp 1997, etc.

Predictions of the Prosodic Hierarchy

1. Phonological processes cluster on exactly one domain between Φ and PhP, i.e. one domain referencing a single stem: the word (ω)
2. More domains between Φ and PhP only by strict recursion (same process, e.g. stress, on recursive levels: Peperkamp 1997)

   Neapolitan Italian: [$_{\omega}$[$_{\omega}$($_{\phi}$cónta)] ($_{\phi}$ténnə)] ‘tell=you=of.it’

   [$_{\omega}$($_{\phi}$tənə) [$_{\omega}$($_{\phi}$cóntə)]] ‘you=of.it=tell’

3. Domains stack only as proper containment (proper bracketing; Itô & Mester 1992): no ***[… (…] …)**

Goals

1. Test theory-based hypotheses against a rich database

   N (languages with *exhaustive* information): 31

   N (domains): 304

   Focus on data from Sino-Tibetan because of its great internal diversity and controversial status of “words”

2. Explore the typological distribution of word domains

Prediction 1: only one ω

Counterexamples: some languages have more than one word domain, e.g. Lahu (Matisoff 1973, 2003):

I. Stress unit: prefix + stem

a. [ɔ̀-u] NMLZ-lay.egg

Not a single-stress unit: stem + suffix

b. [vɨ̀-tā] buy-PFPM

II. Tone change: stem + suffix

c. ši-ɛ̀ > [ší-ɛ̀] yellow-ADVLZ

No tone change: prefix + stem

d. á-qhâ > [á.qhâ] NFP-ragweed

Lahu word domains

	PF	Σ	SF
Stress	———		
Tone Change		———	

Chukchi multiple domains (subset)

Prediction 2: more ω only by recursion

Counterexamples: some languages stack by “pseudo-recursivity” (different domains, different processes), not proper recursivity, e.g. Belhare:

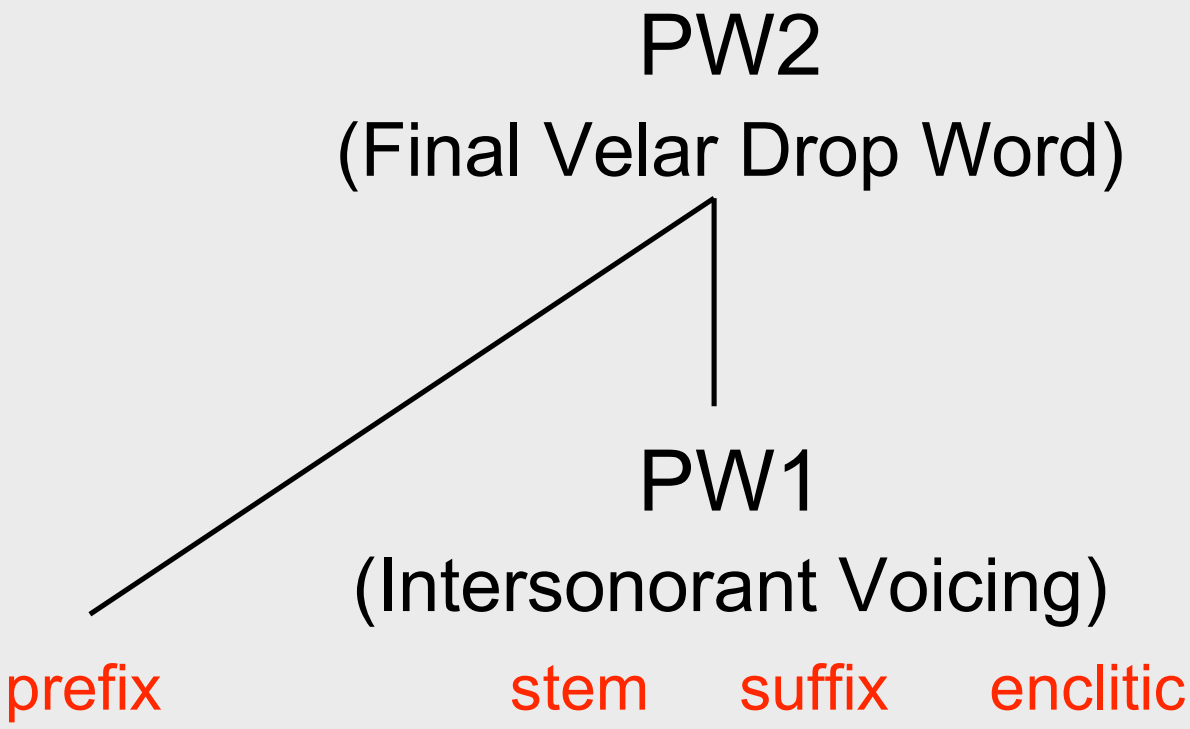
I. Intersonorant voicing: stem + suffix + enclitic

a. ka-teĩ-ʔ-ni-kak > ka($_{\omega}$teĩʔniga), *($_{\omega}$kareĩʔniga)
   1sP-hit-NPST-NEG-2 ‘You won’t hit me.’

II. Final velar drop: prefix + stem + suffix + enclitic

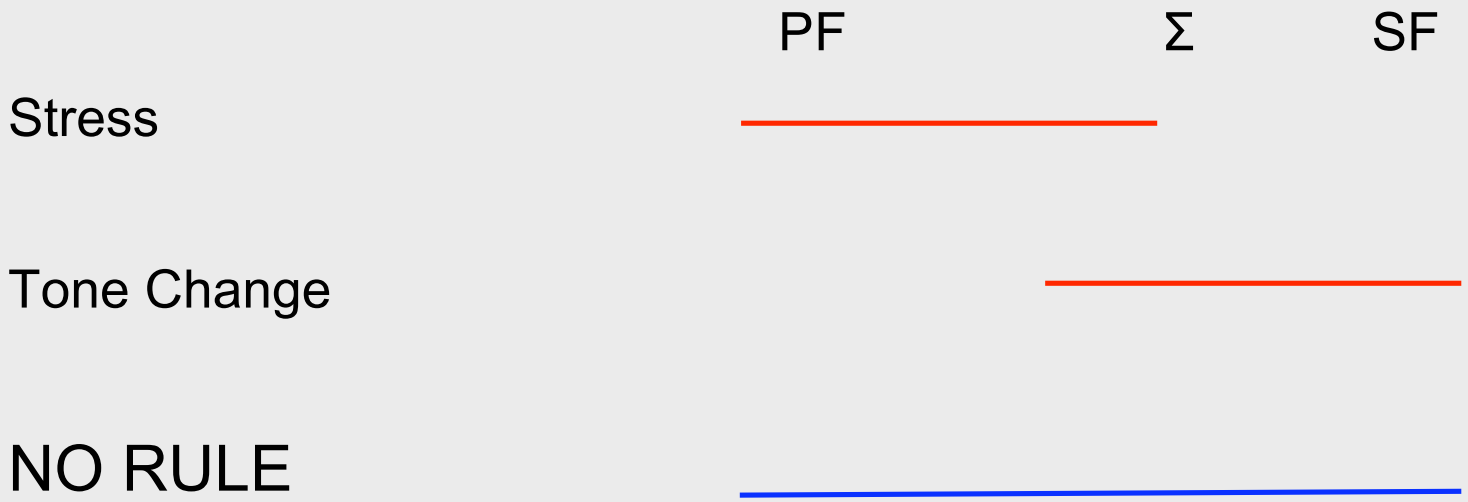
b. ka-ak-lu-kak=phu > ($_{\omega}$kaaklugakphu), *($_{\omega}$kaa)($_{\omega}$lugakphu)
   1sP-OPT-tell-2A=REP ‘You may tell me, they say’

Belhare Pseudorecursivity (Partial)



Prediction 3: Proper containment

Counterexamples: in some languages the biggest ω in a particular language may not include all available affix types at once, cf. Lahu again:



Interim summary

- Our database does not support the predictions entailed by the Prosodic Hierarchy Hypothesis.
- Instead, we find substantial diversity.

How, then, do p-domains distribute typologically? What, if anything, governs their distribution?

Typological distribution

Test genealogical and areal factors

- Area: within Sino-Tibetan
- Stock: Sino-Tibetan compared to others
- against chance by using permutation methods (Janssen, Bickel & Zúñiga 2005)

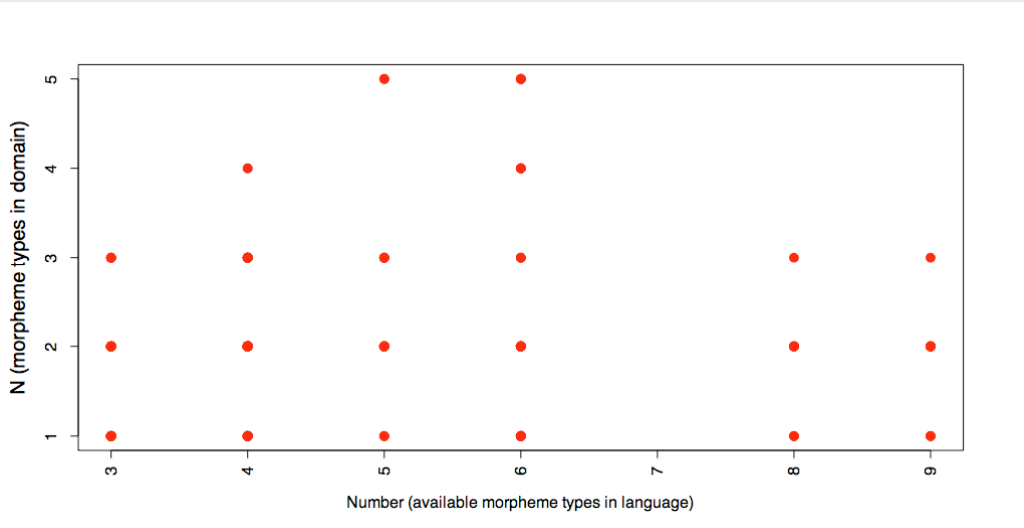
But, how to measure the distribution?

Typological measurements

- 2-13 ω per language
- Need
  - some measure of coherence (‘short’, ‘disruptive, noncohering’ vs. ‘long’, ‘all-encompassing’ pw)
  - some measure of diversity (2 ω vs. 13 ω)

Coherence (*c*)

- Coherence: how many morpheme types are included in the domain? (stem alone? stem plus prefix? plus prefix and suffix? etc.)
- *N* (morpheme types in domain) correlates with *N* (available morpheme types in the language):



Kendall's $\tau = 3.55$, p (rnd) = .001, $N = 303$ from 30 languages

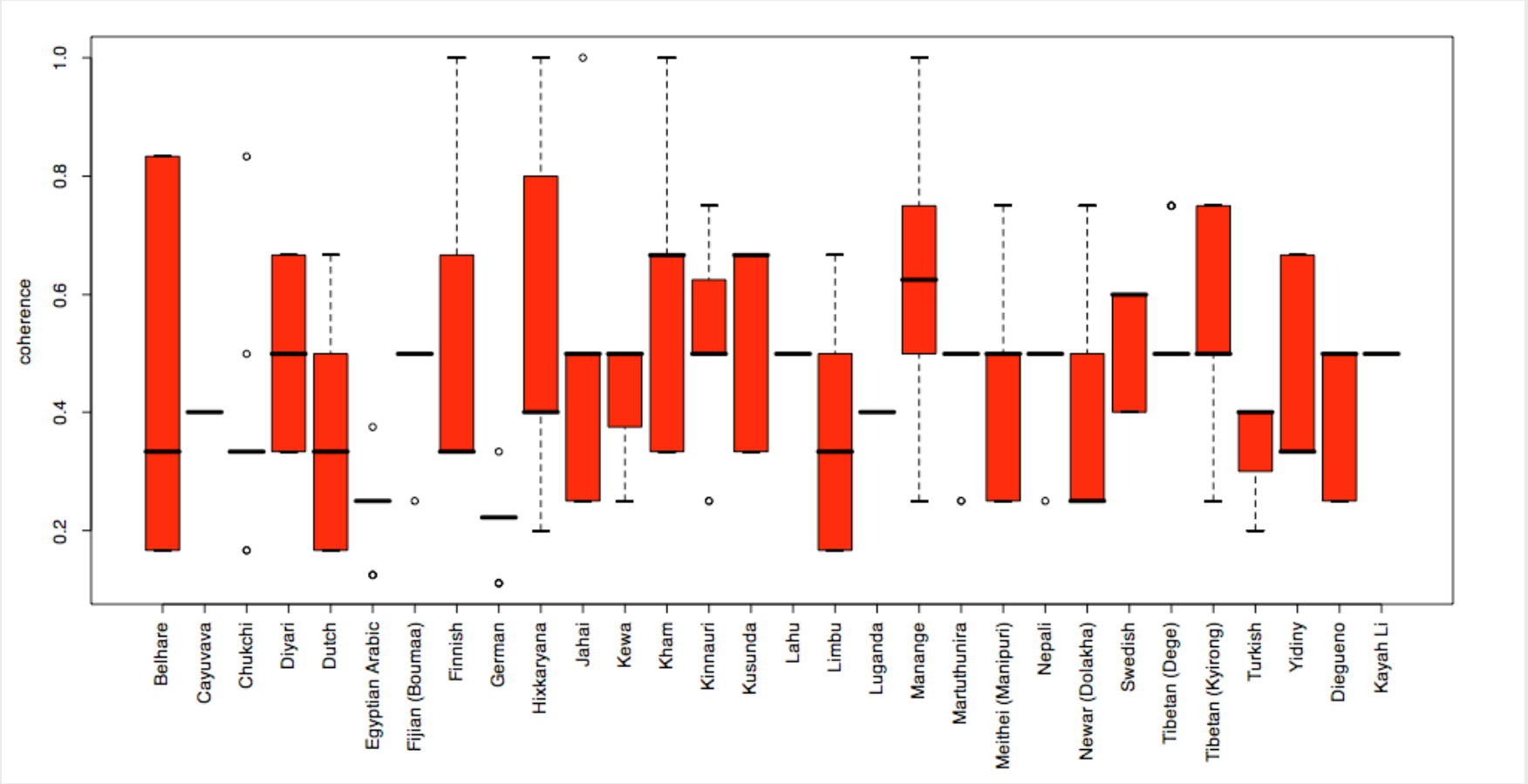
Coherence (*c*)

- Coherence of a domain is relative to the number of available morphemes:

$$c = \frac{N\text{ (morpheme types in domain)}}{N\text{ (available morpheme types)}}$$

Is *c* a typological (cross-linguistic) variable?

Coherence (*c*)



The variance between languages is greater than the variance within languages: F (30, 466) = 3.89, p (rnd) = .0001

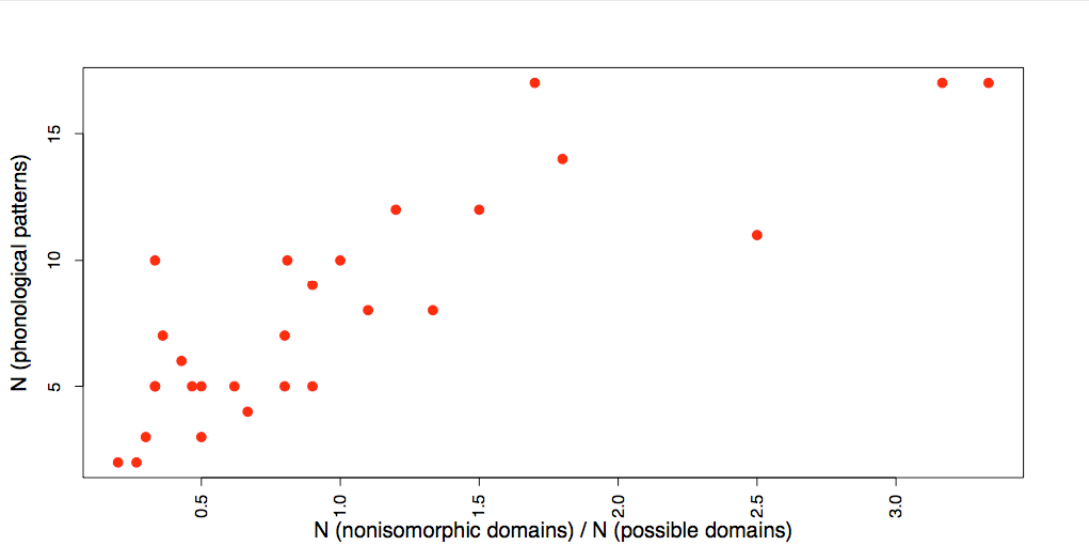
Diversity (*d*)

- Since *c* is a typological variable, it is reasonable to take a *per-language* measurement on language-internal diversity
- $d = \delta$, the number of non-isomorphic domains
- but δ depends on the number of *logically possible* nonisomorphic domains, e.g. if there is only {prefix, stem}, there are only 3 possible domains: (pf-st), (pf), (st)
- possible number of domains with v morpheme types:

$$\sum_{k=1}^{v} k$$

Diversity (*d*)

- Ergo, define *d* relative to number of possible domains
- But *d* also depends (obviously) on the number of phonological processes in the language φ:



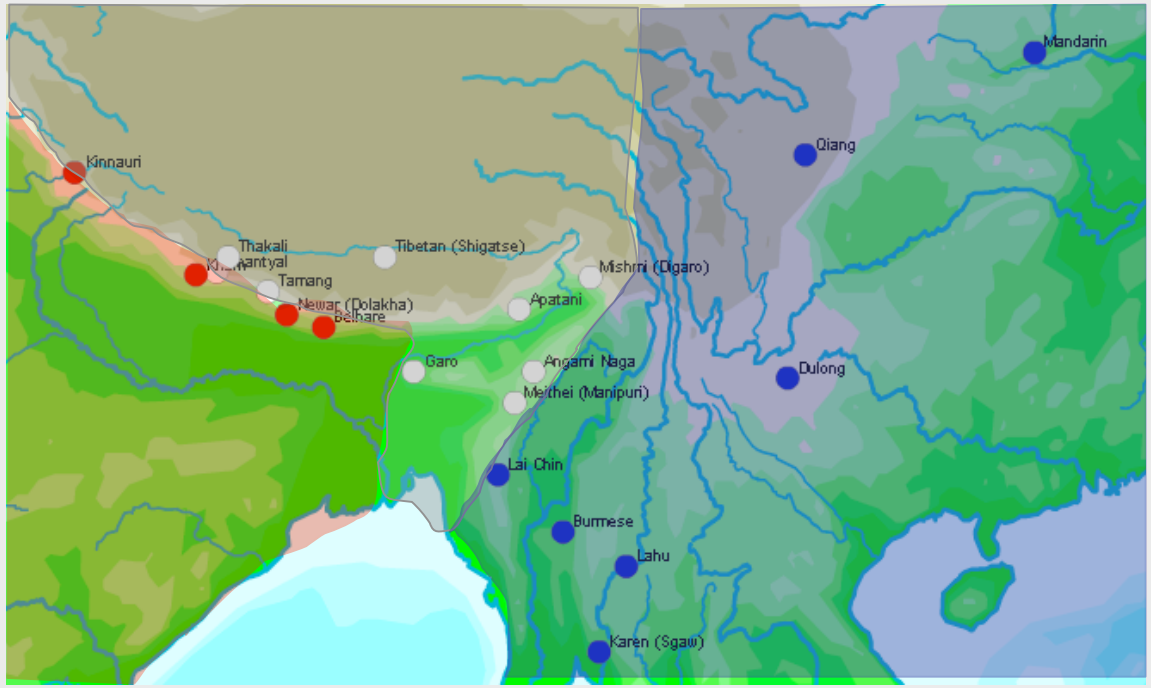
Kendall's τ = 4.25, *p* (rnd) =2.2e-16, *N* (languages) = 31

Diversity (*d*)

- Therefore, $d = \dfrac{\delta}{\varphi \sum_{k=1}^{\nu} k}$

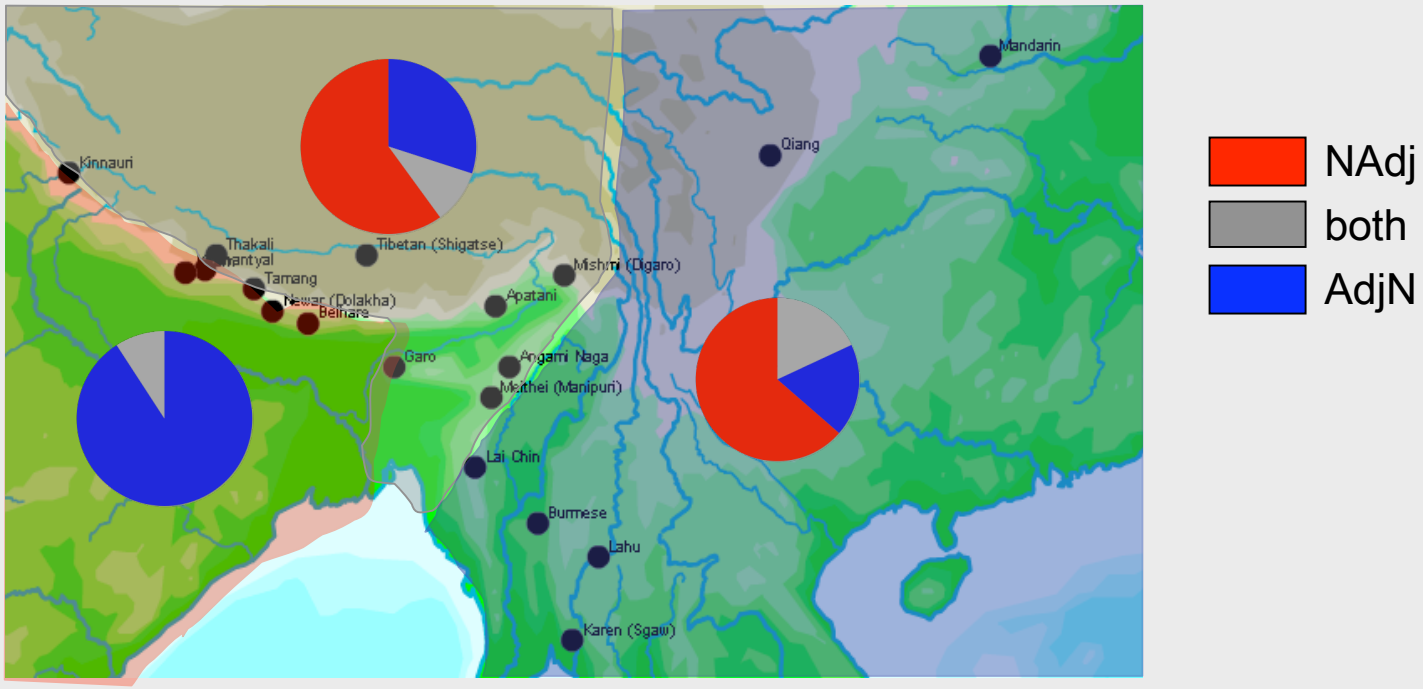
Areal factors

- Our db focuses on Sino-Tibetan, so test within ST
- Prominent areal factors in ST
  - Indosphere vs. Sinosphere (Matisoff 1991, 1999)
  - plus “Buffer Sphere” between the two



Areal factors in ST: previous evidence

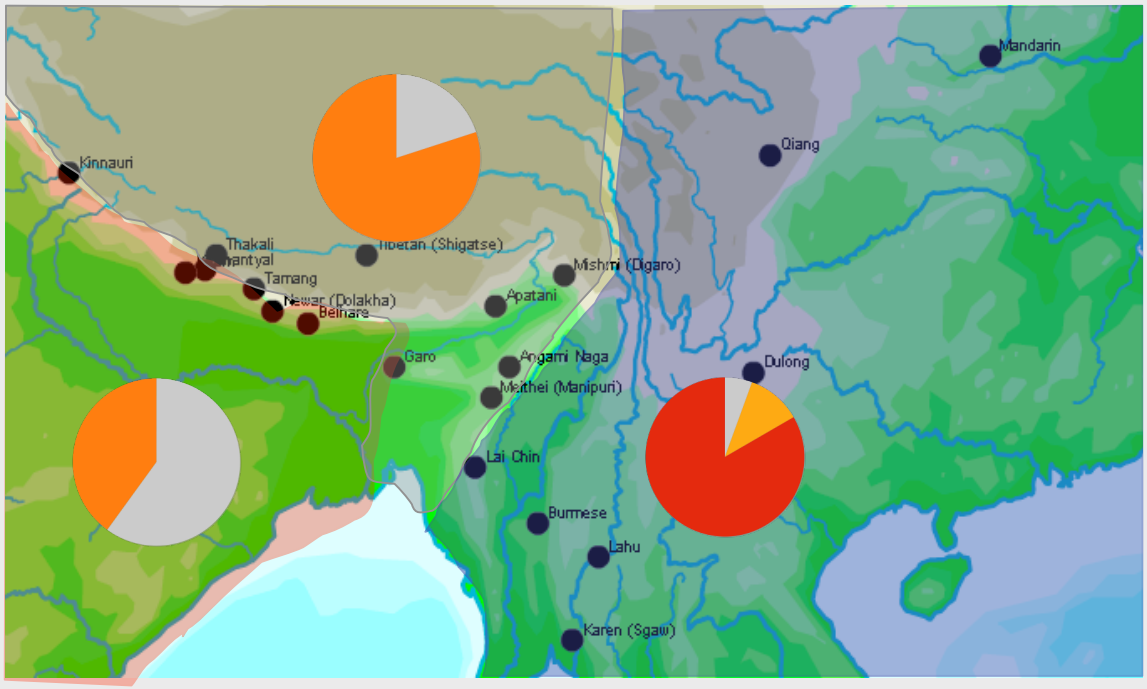
- Order of Adj&N (Dryer 2004, 2005)



χ^2 (4,32) =14.35, p (rnd) = .0001

Areal factors in ST: previous evidence

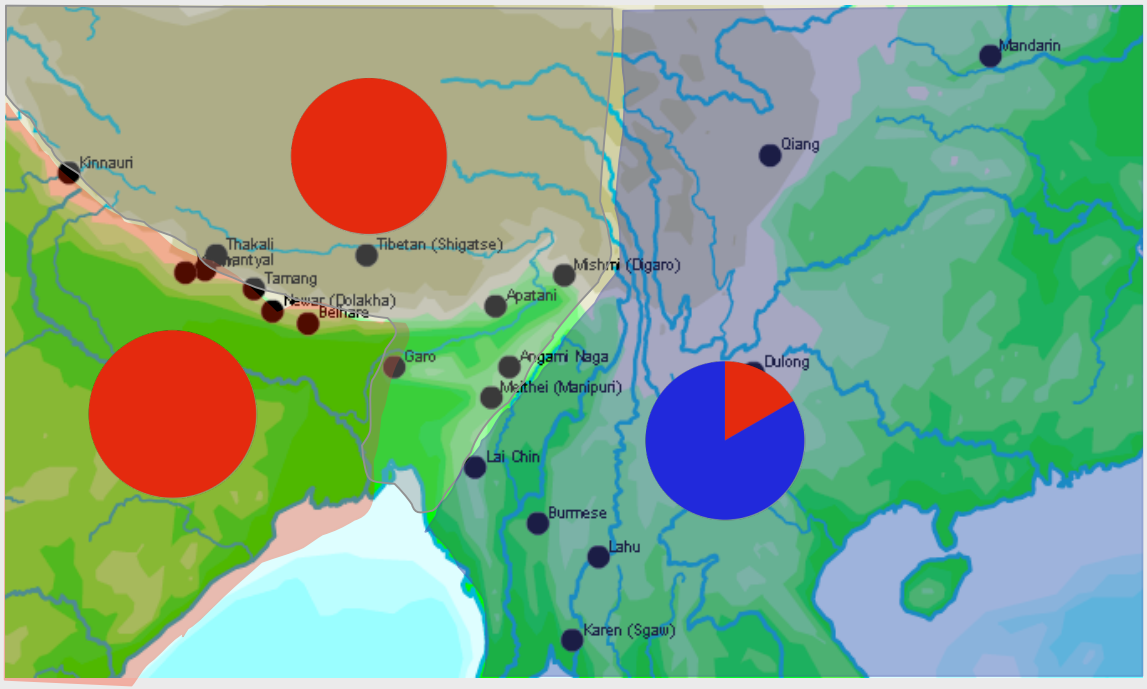
- Tone (Matisoff 1999, Maddieson 2005, own data)



χ^2 (4,20) =14.96, p (rnd) = .0001

Areal factors in ST: previous evidence

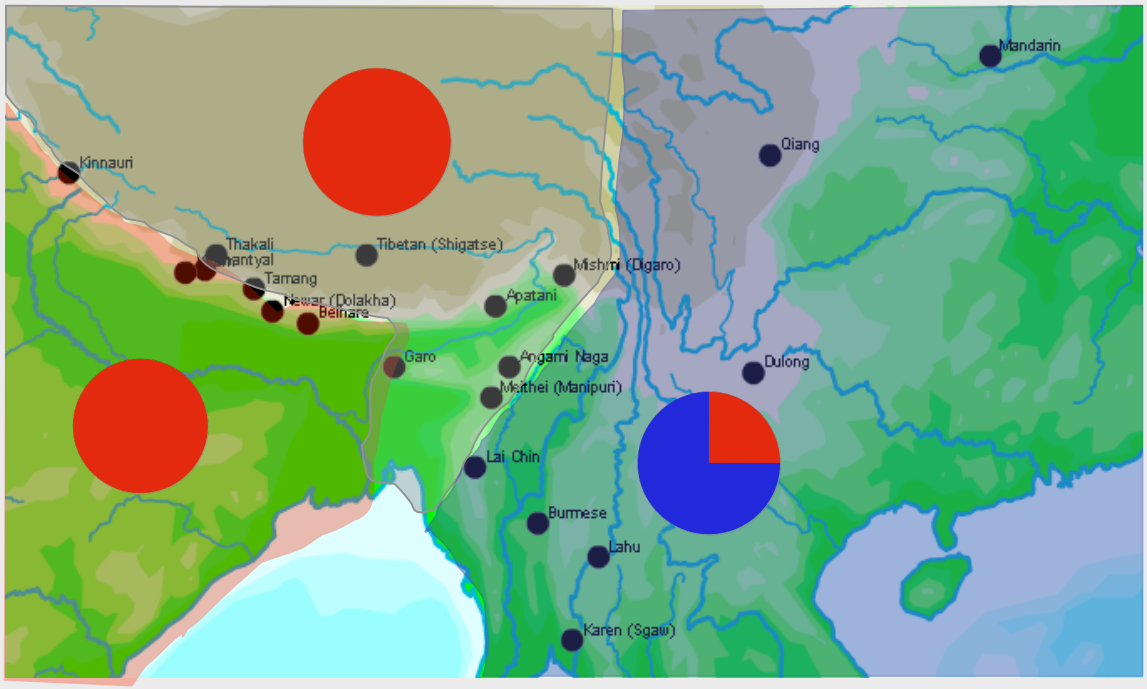
- Fusion of negation markers (Bickel & Nichols 2005)



χ^2 (4,10) = 6.67, p (rnd) = .046

Areal factors in ST: previous evidence

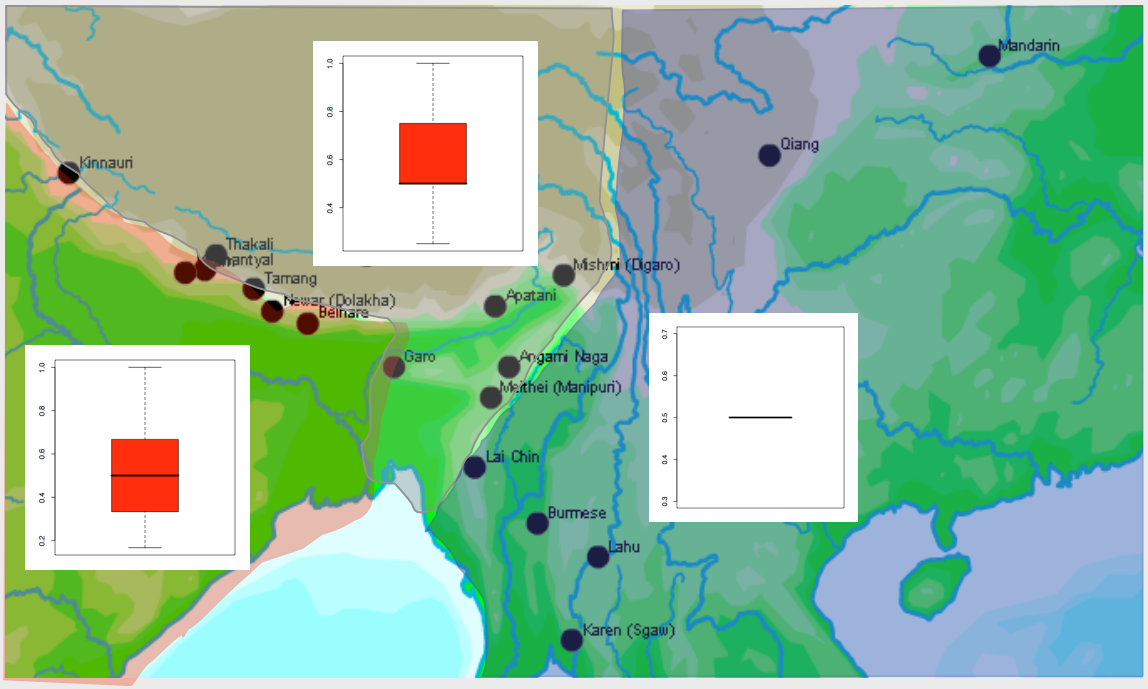
- Fusion of case markers (Bickel & Nichols 2005)



χ^2 (4,12) = 7.22, p (rnd) = .05

Areal factors in ST: testing *c* and *d*

- Coherence (*c*)

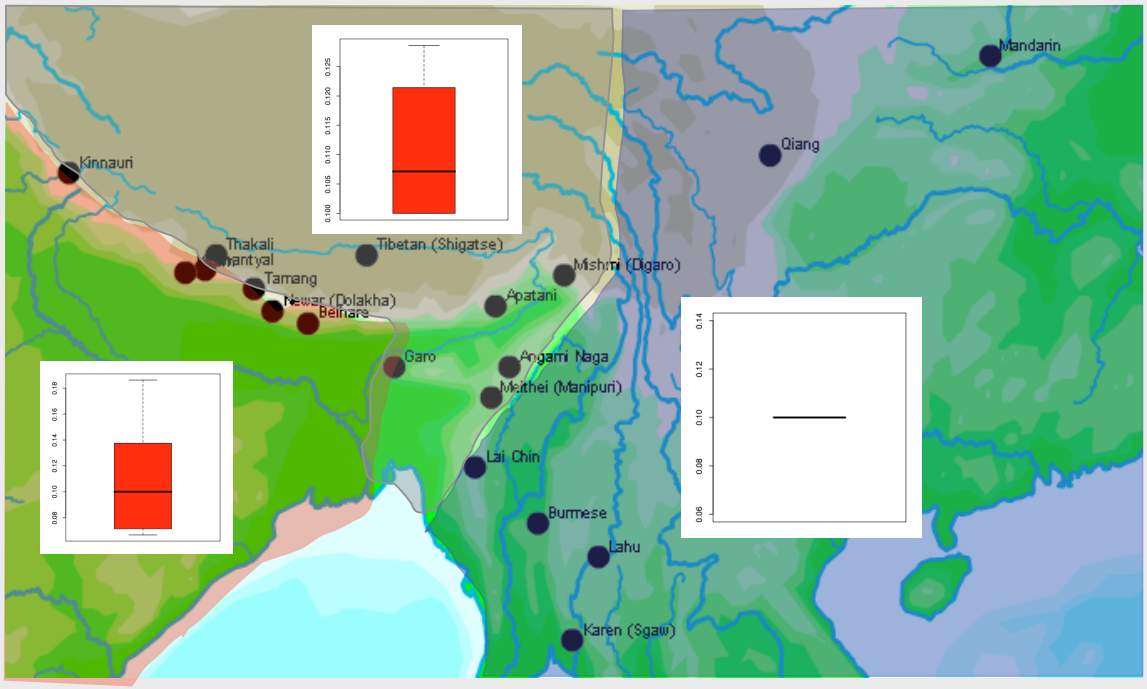


F (2,114) =.88, *p* (rnd) = .41

Combined Indosphere and Buffer Sphere: *F* (1,115) = .04, *p* (rnd) =.84

Areal factors in ST: testing *c* and *d*

- Diversity (*d*)



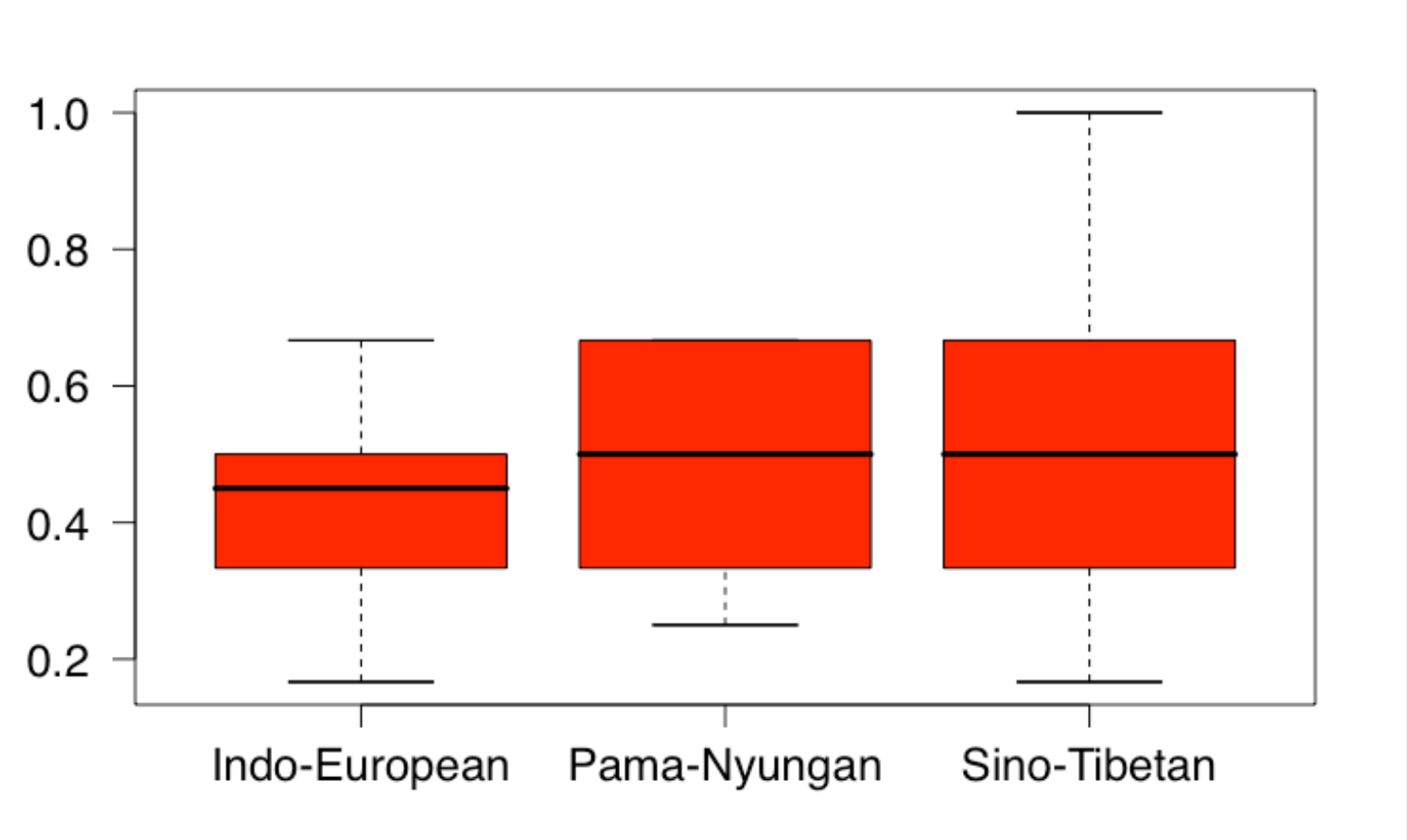
F (2,8) =.08, p (rnd) = .93

Combined Indosphere and Buffer Sphere: F (1,9) = .19, p (rnd) =.64

Areal factors: summary

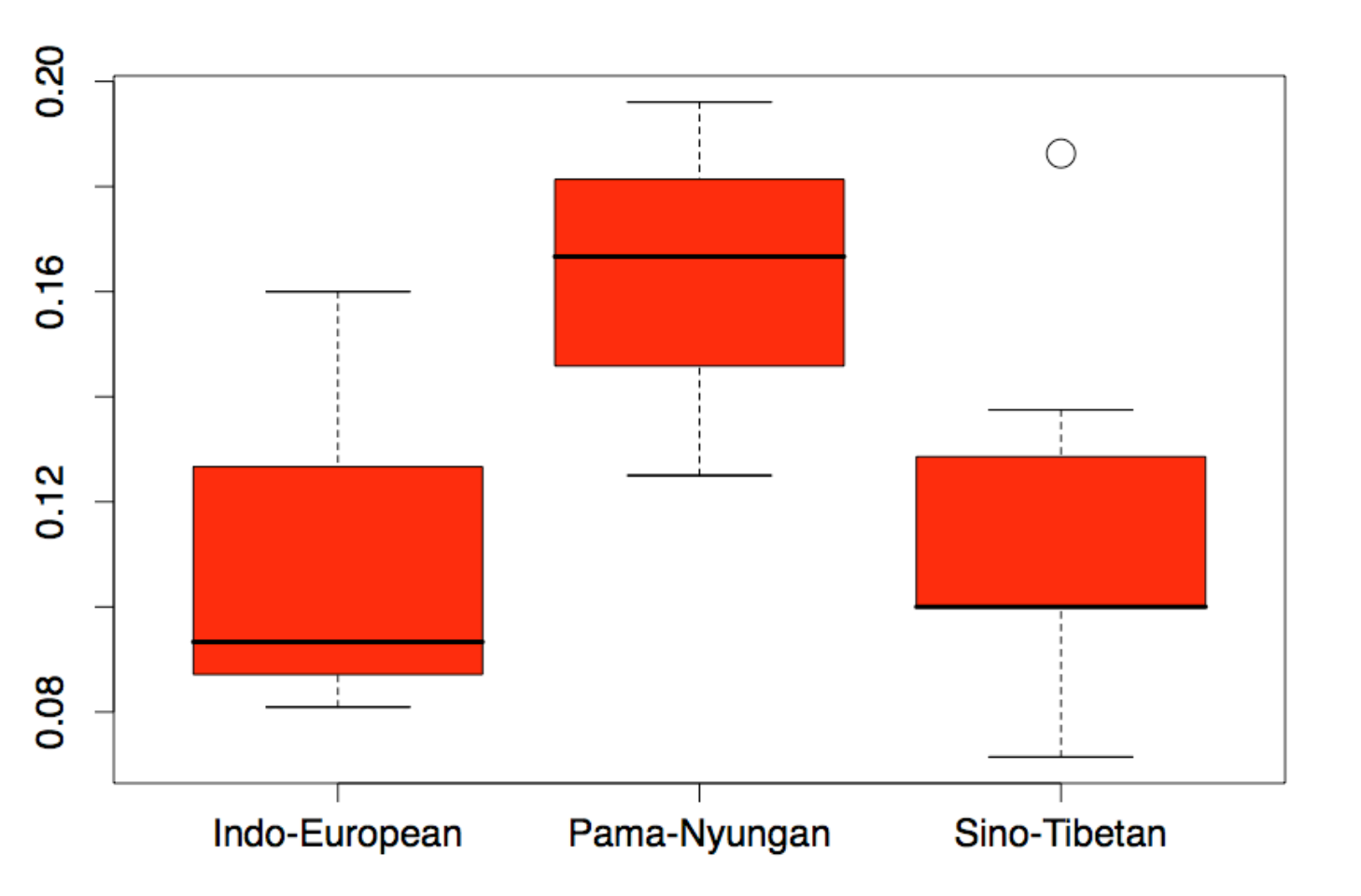
- Despite ample evidence for Matisoff's spheres in ST, no evidence for sphere effects on *c* and *d*
- ST seems surprisingly consistent!
- Are *c* and *d* genealogically stable?
- Database still too poor for extensive testing, but there is *preliminary* evidence that between-stock variance is larger than within-stock variance
  - Sino-Tibetan (10)
  - Pama-Nyungan (3)
  - Indo-European (3)

Genealogical factor: stock x coherence



F (2, 282) = 4.09, *p* (rnd) =.017 — preliminary, small non-ST samples!

Genealogical factor: stock x diversity



F (2, 13) = 2.58, *p* (rnd) =.10

Conclusions

Factors governing the distribution of phonological word domains

- no support for universal constraints
- no support for areal patterns (spheres) within Sino-Tibetan
- limited support for genealogical stability, perhaps on the stock level

⇨ Overall distribution result of individual historical developments

Contrast with coherence of individual formatives (case, negation) that do evidence areal patterns in Sino-Tibetan

- individual formatives can escape the overall coherence profile of a language
- further support for individual historical sources of the observed distribution

Acknowledgments

- **The AUTOTYP research team**
  - Johanna Nichols (Co-Director, Berkeley)
  - Balthasar Bickel (Co-Director, Leipzig)
  - Kristine Hildebrandt (Post-Doc, Leipzig)
  - Tracy Alan Hall (Research Associate, Bloomington)
  - Fernando Zúñiga (Research Associate, Santiago)
  - *RAs in Berkeley:* Gabriela Caballero, Suzanne Wilhite
  - *RAs in Leipzig:* Michael Riessler, Sven Siegmund, Sindy Poppitz, Franziska Crell, Jenny Seeg, Thomas Goldammer, Anja Gampe, Sebastian Hellmann
  - *Past team members:* Kathi Stutz, Josh Wilbur, Rebecca Voll, Alena Witzlack-Makarevich, Sandra Biewald, Aimee Lahaussois-Bartosik, Dave Peterson, Keith Sanders
- **Funding:** Swiss NSF Grant Nos. 08210-053455 (1998-2001, Bickel) and 610-0627 (2001-2002, Bickel), German DFG Grant No. BI 799/2-1 (2003-2005, Bickel & Hall), US NSF Grant No. 96-16448 (1998-2001, Nichols)

AUTOTYP