

A world map with a light blue background and greenish-yellow landmasses. The map is centered on the Atlantic Ocean, showing the Americas on the right and Europe and Africa on the left. The text is overlaid on the map.

WORD DOMAINS

Balthasar Bickel

Kristine A. Hildebrandt

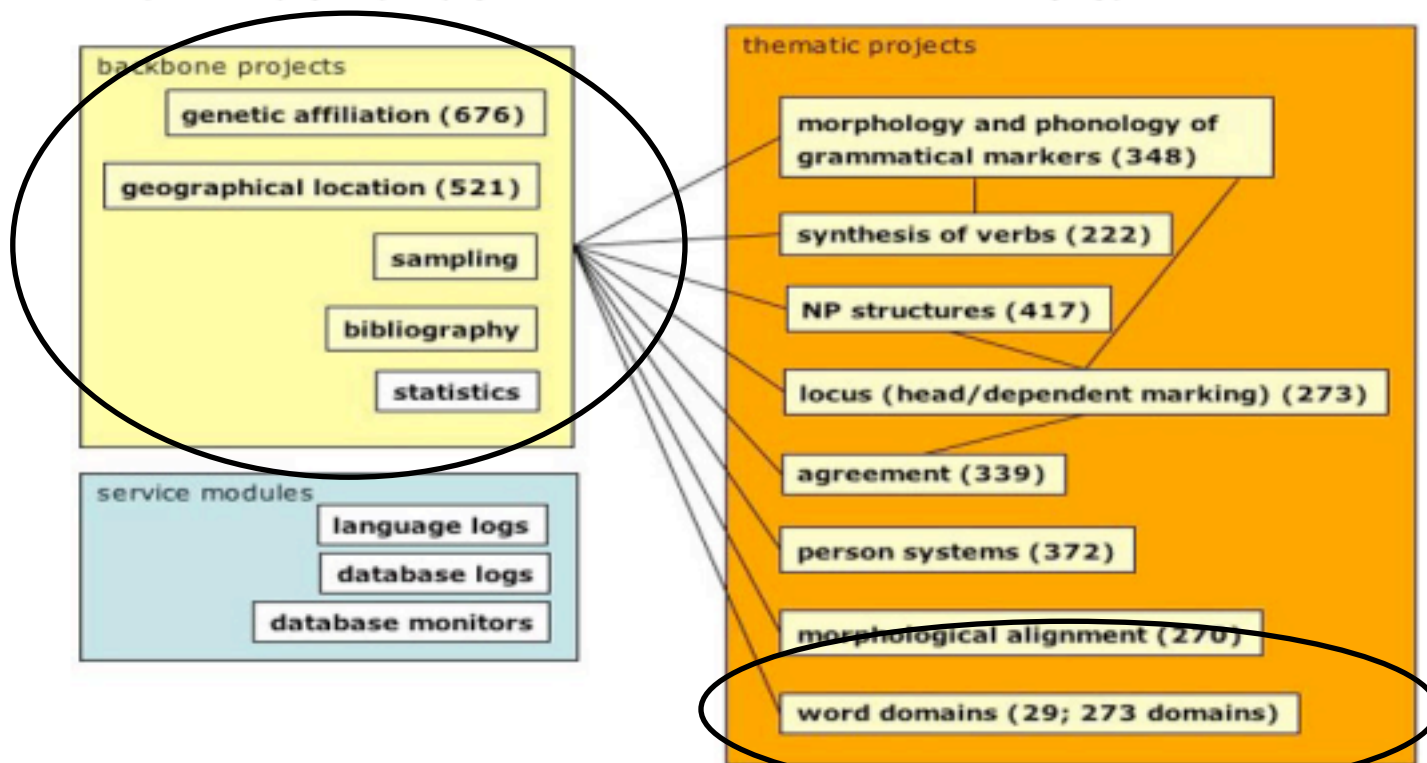
khilde@uni-leipzig.de

I. "Autotyp": <http://www.uni-leipzig.de/~autotyp/>

PROJECTS

AUTOTYP is not an single individual database project. Rather, it is a large-scale research program involving various thematically specific projects. In each project we develop a series of **data and definition files** linked together relationally. Currently, the number of files across all projects is about 70. All database files are currently implemented in FileMaker Pro™ for the Macintosh, but will in due time be transferred into a more general, platform independent format.

The following projects are fully implemented and data collection is currently proceeding full steam (click for further information). Lines indicate the most important relationships between projects. (Most projects are interconnected via shared definition files anyway.)



A faint, light gray world map is visible in the background of the slide, showing the outlines of continents and oceans.

II. Words: Phonological & Grammatical

- Definitions
- Examples from database

The Phonological Word ω

- Selkirk (1980); Booij (1983; 1984; 1985); Nespor & Vogel (1986); Peperkamp (1997); Inkelas (1987); McCarthy & Prince (1993); Hall & Kleinhenz (1999); Wiese (1996); Dixon & Aikhenvald (2002)
- A phonological domain that references morphological information = **Phonological Word**

Phonological processes/ rules/constraints that...

- Reference domains ‘beyond the syllable’
- Also reference some morphological or morpho-syntactic domain (e.g. root/stem; stem + affix; stem + stem, etc.)
- ‘morpho-phonology’

P-word in German (“Onset Maximization” Booij 1985; Wiese 1996)

Onset: *bl* surfaces word-initial:

1. *blitz* -> [blits] ‘lightning’

Coda word-medial: *bl* will often resyllabify:

2. *bibl + isch* -> [bi.blɪʃ] ‘biblical’

= ‘Onset Maximization’

Surfacing coda obstruents devoice

3. *Lob* -> [lo:p] ‘praise’

Rad -> [ʁa:t] ‘wheel’ = ‘Final Devoicing’

German P-word

- *lieb + e* love-1SG.PRES -> [li:..bə] ‘I love...’
- BUT: *lieb + lich* love-ADV -> [li:p.liç] ‘dearly’
 - no onset maximization for a *bl*- cluster
 - rather, /b/ devoices to [p]
- Consonant-initial suffixes parsed as their own domains for this process (separate p-domains)
- (libə)_ω & (biblif)_ω vs. (lib)_ω (liç)_ω
- Domain of ONSET MAX: stem ± v-initial suffix
- Prefixes excluded: *ab-ändern* [(?Λp)_ω.?ɛn.dəʁn]
- Compounds excluded: *bergabfahrt* [(bɛʁk)_ω.?ap.fəʁt]

German Onset Max Word

WTID 309 LID 87 German UNIT p-domain

Inputter 24 Kristine

Date 02-02-2005

Reliability 5 Other published sources

WTYPE 210 Coda C to Onset Resyllab

PHON DOMAIN 3 right

RESOLUTION 28 resyllabification

Morph Domain

Align ID 3

Domain ID 14

Align right edge

Domain stem ± suffix

Gram Domain Size 2

Synthesis

Number Formative Slots 2

Bipartite Stem 0

Max number categories in sequence

Number of Available Morphemes 5

Strata

Stratum ID 52

only v-initial suffixes

Source native

if source is loanword:

LID1

LID2

LID3

LID4

LID5

LID6

KNOWN LENDING LANGUAGES

Notes

data: Jochen Geifuss-Wolfgang (Autotyp Bib); Hall 1999; Booij 1985; Wiese 1996

ONSET MAXIMIZATION applies if 2nd syllable or morph is v-initial. If cv(c), then no ONSET MAX, and rather devoicing of stem/1st syllable coda C:

lieb + e => /li.bə/ 'love.1sg' vs. lieb + lich => /li.ɸ.liç/ 'dearly'

Manange (Tibeto-Burman)

- Single Tone Contour Word = stem ± prefix± suffix ± particle (‘clitic’)
- For example, tone /3/: a high, sharp falling tone
- [sΛ⁵³] ‘good/tasty/wholesome’
- [a⁵⁴-sΛ⁴³] NEG-good ‘not good’
- [a⁵⁴-sΛ⁴³-pΛ^{32/22}] NEG-good-NOM ‘not good one’
- [kjẽ¹¹ a⁵⁴-sΛ⁴³-pΛ=ko³²] rice NEG-good-NOM=DEF ‘the bad-tasting rice’

Manange Tone Contour Word

WTID 1 LID 690 Manange UNIT p-domain

Inputter 24 Kristine

Date 2-10-2003

Reliability 3 Grammar Explicit

WTYPE 1 Single Tone Contour/Melody

PHON DOMAIN 4 span

RESOLUTION 17 preservative spread/assimilation

Morph Domain

Align ID 7

Domain ID 64

Align spans Domain stem ± prefix ± suffix ± particle

Gram Domain Size 4

Synthesis

Number Formative Slots

Bipartite Stem 1

Max number categories in sequence

Number of Available Morphemes 4

Strata

Stratum ID 8 all but approx 15 disyllabic noun roots Source all

if source is loanword:

LID1

LID2

LID3

LID4

LID5

LID6

KNOWN LENDING LANGUAGES

206 Nepali

375 Tibetan (Standard Spoken)

Notes Hildebrandt 2003, 2004, 2005 see WTID 327

G-Words

- In **morphology & syntax**: general assumption of a maximum of 2 or 3 domains:

M (Σ , Root, Affix)

X⁰ (Any syntactic head/terminal node)

XP (Any syntactic phrase)

- **Other accounts of grammatical units also very general:**
Fixed combination or ordering possibilities for elements
A certain “conventionalized meaning”
A degree of coherence
(e.g. Dixon & Aikhenvald 2002)

G-word in Belhare (Tibeto-Burman)

WTID 234 LID 35 Belhare UNIT g-domain

Inputter 24 Kristine

Date 6-11-2004

Reliability 8 Questionnaire

WTYPE 162 Terminal Word (X0)

PHON DOMAIN 5 n/a

RESOLUTION

Morph Domain

Align ID 7

Domain ID 27

Align spans Domain (object + intrans affxes) ± stem + affxes

Gram Domain Size 4

Synthesis

Number Formative Slots	13
Bipartite Stem	1
Max number categories in sequence	7
Number of Available Morphemes	4

Strata

Stratum ID

Source

If source is loanword:

LID1
LID2
LID3
LID4
LID5
LID6

KNOWN LENDING LANGUAGES

206	Nepali	▾
169	Maithili	

Notes e.g. phak se'-yu pig 3sS-kill-NPST; phak 'pig' cannot project a phrase of its own in this type of construction

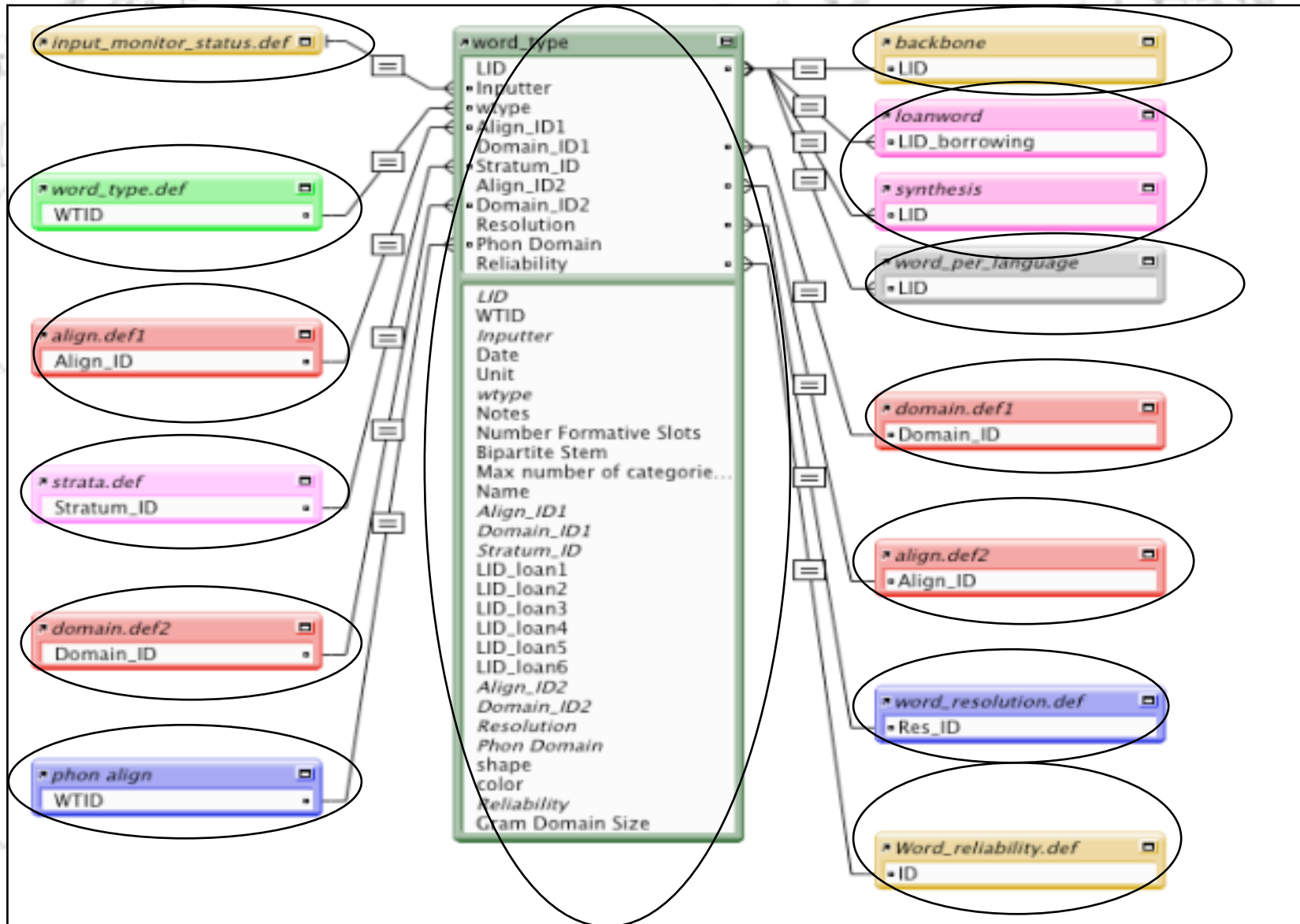
Belhare ‘Terminal Word’

- Domain: (object+intransitive affixes) ± all stem parts +affixes
- p^hak seʔ-yu
pig [3sS-]kill-NPST
'S/he kills pigs'
- p^hak ‘pig’ cannot project a phrase on its own:
 - *eik^ha p^hak seʔyu ‘S/he kills big pigs’
- Not incorporation: p^hak=to seʔyu

III. A comprehensive catalogue of word types

- Description of **process/constraint & resolution** (if applicable)
- Description of **phonological domain & alignment type**
- Description of **grammatical domain & alignment type**
- Exceptions or 'strata'
- Distribution of types, sizes, alignments

What goes into Word Records?



In addition to Manange...

LID	::Language	WTID	l...	::<File...	...	Unit	wty...	word_type.def::word_type
690	Manange	1	24	<File	2	p-domain	1	Single Tone Contour/Melody
691	Tibetan (Lhasa)	84	24	<File	1	p-domain	1	Single Tone Contour/Melody
664	Kham	282	24	<File	8	p-domain	1	Single Tone Contour/Melody

WordType_1: Single Tone Contour

WTID	1	word_type	Single Tone Contour/Melody
phon_def1	10	Tone	
phon_def2			
phon_def3			
Phon_Domain	4	span	
Notes_Explanations	1 contour or melody per unit		
Gword_Definition			

Some other recurring P-Domain Types

- 5 languages: *CC ‘coda’ position restrictions
- 10 languages: *CC ‘onset’ position restrictions
- 5 languages: Single Main Stress
- 6 languages: *V-initial syllable restrictions

A 'Rare' P-domain: Yidiny

WTID 97	LID 580	Yidiny	UNIT p-domain	Inputter 24	<File Missing>
WTYPE 74	Max 7 Syllables			Date	10-31-2003
PHON DOMAIN	<input type="text" value="4"/>	<input type="text" value="span"/>			
RESOLUTION	<input type="text"/>	<input type="text"/>			
Morph Domain: Default					
Align_ID1	<input type="text" value="7"/>	Domain_ID1	<input type="text" value="26"/>		
Align	<input type="text" value="span"/>	Domain	<input type="text" value="stem ± affixes"/>		
		Gram Domain Size	<input type="text" value="2"/>		
Morph Domain: Alternative					
Align_ID2	<input type="text"/>	Domain_ID2	<input type="text"/>		
Align	<input type="text"/>	Domain	<input type="text"/>		
		Gram Domain Size	<input type="text"/>		
Strata					
Stratum_ID	Definition	Source			
if source is loanword:		KNOWN LENDING LANGUAGES			
LID_loan1	<input type="text"/>	<input type="text" value="74 English"/>			
LID_loan2	<input type="text"/>	<input type="text"/>			
LID_loan3	<input type="text"/>	<input type="text"/>			
LID_loan4	<input type="text"/>	<input type="text"/>			
LID_loan5	<input type="text"/>	<input type="text"/>			
LID_loan6	<input type="text"/>	<input type="text"/>			

Synthesis
Number Formative Slots
Bipartite Stem
Max number of categories in sequence <Table

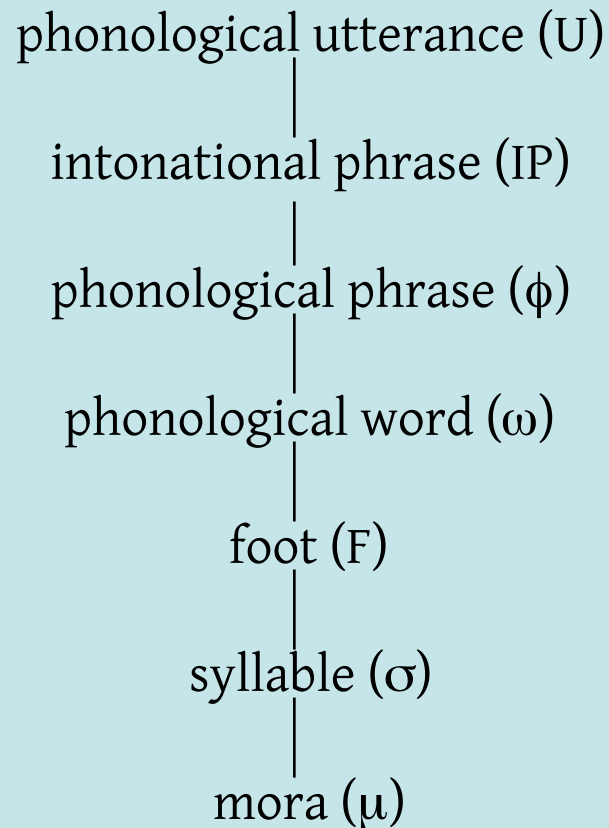
Notes Dixon notes that a phonological unit may not be more than 7 syllables in size (not true for non-cohering affixes); need more morph detail on participatory affixes

IV. Traditional Approaches to P & G Domains

- The Prosodic Hierarchy
- Prosodic, Grammatical Alignments, & Cross-Domain Alignments
- What we are finding so far...

Prosodic Hierarchy

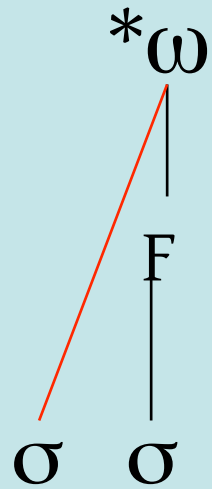
Series of nesting domains for phonological rules/processes



Prosodic Hierarchy: Alignment

- UG specifies these prosodic constituents, arranged in a hierarchy
- Hierarchy characterized by principles & constraints (Selkirk 1980; Nespor & Vogel 1986)
 - Strict Layer Hypothesis: “A prosodic constituent of rank n is immediately dominated by a single constituent of rank $n + 1$ ”

Alignment: Strict Layer Hypothesis



'skipping' of a level



'recursivity' of levels

- Recursivity permittable (Ladd 1986)
- Violable constraints (Optimality Theory) (Ito & Mester 1992; Selkirk 1995; Peperkamp 1997):
 - NONRECURSIVITY: No C_i dominates another C_i
 - EXHAUSTIVITY: No C_i immediately dominates a C_k , $k < i-1$
- Non-Recursivity (Truckenbrodt 1999)=
 - P-phrases in recursive structures must be maximally alike in extension

Grammatical Domains: Alignment

- Correspondence constraints kept very general:
 - ALIGN (PCat, Edge; GCat, Edge), where PCat = $\{\mu, \sigma, F, \omega, \phi\}$ and GCat = $\{Af, Rt, \Sigma, X^0, XP\}$ (McCarthy & Prince 1993)
 - Wrap-XP (Truckenbrodt 1999) Syntactic phrases are contained within only 1 phonological phrase
 - Tautomorphemicity (Crowhurst 1994; Bickel 1998; 2003)

*(G-cat1)(G-cat2)

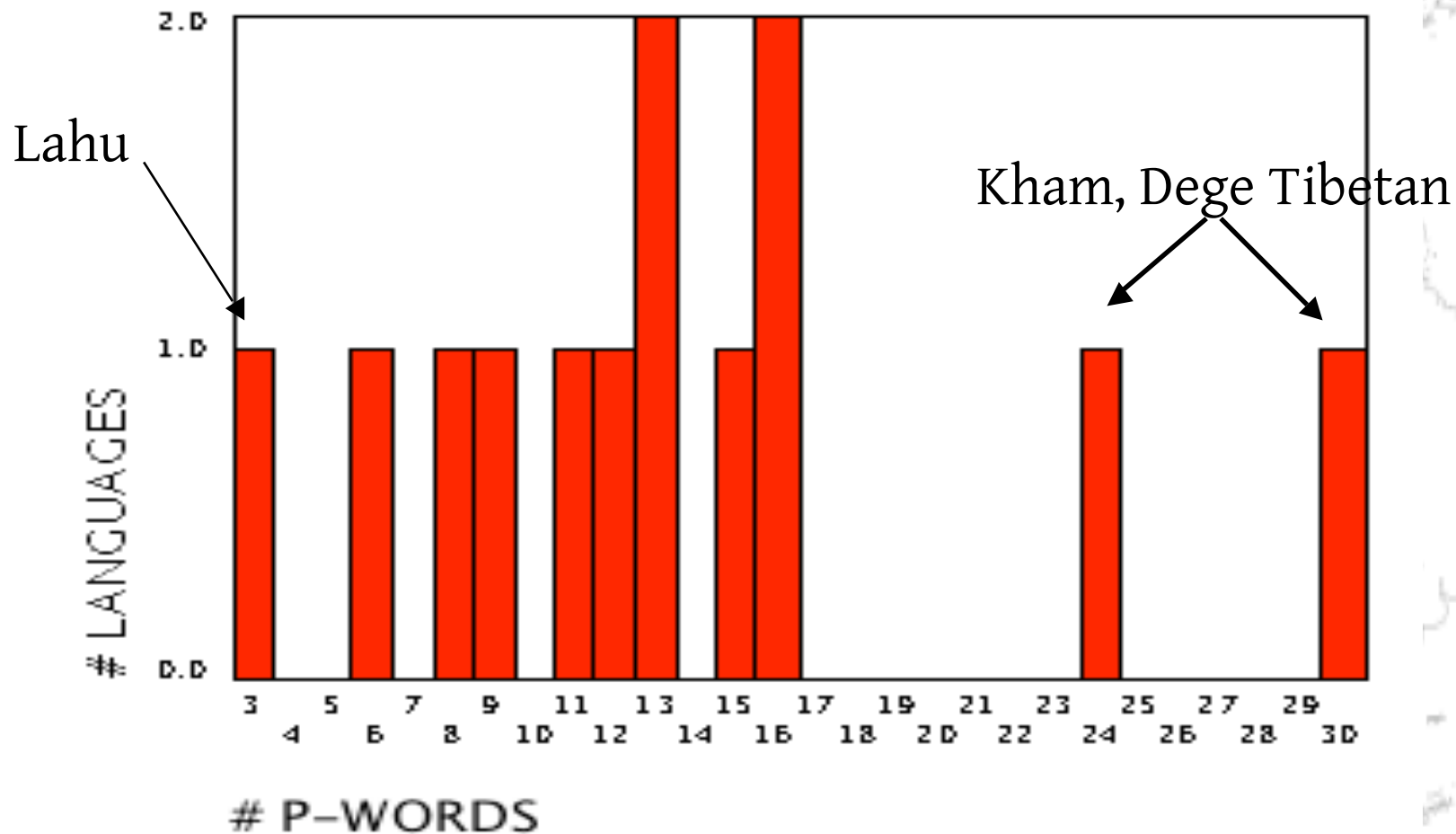
~~[P-CAT]~~

Our Findings: Diversity in Word Domains

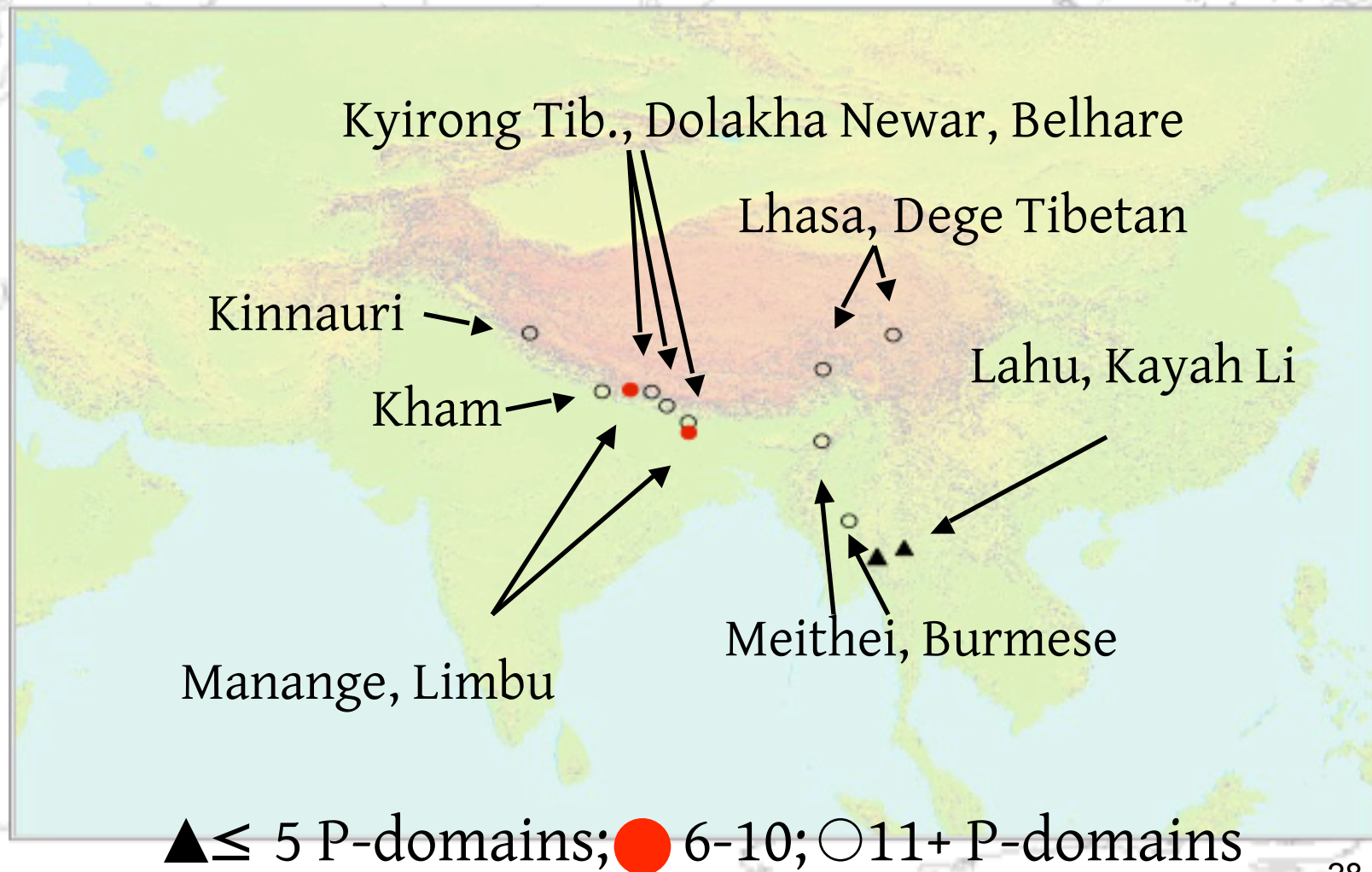
- Languages often differentiate more than one P-domain
- Different P-domains reference different morphological information (non-isomorphism)
- P-domains not necessarily contained within each other à la traditional 'recursion' expectations

Numbers of P-Word Types: 13 Sino-Tibetan Languages

Kinnauri, Meithei, Lhasa Tibetan

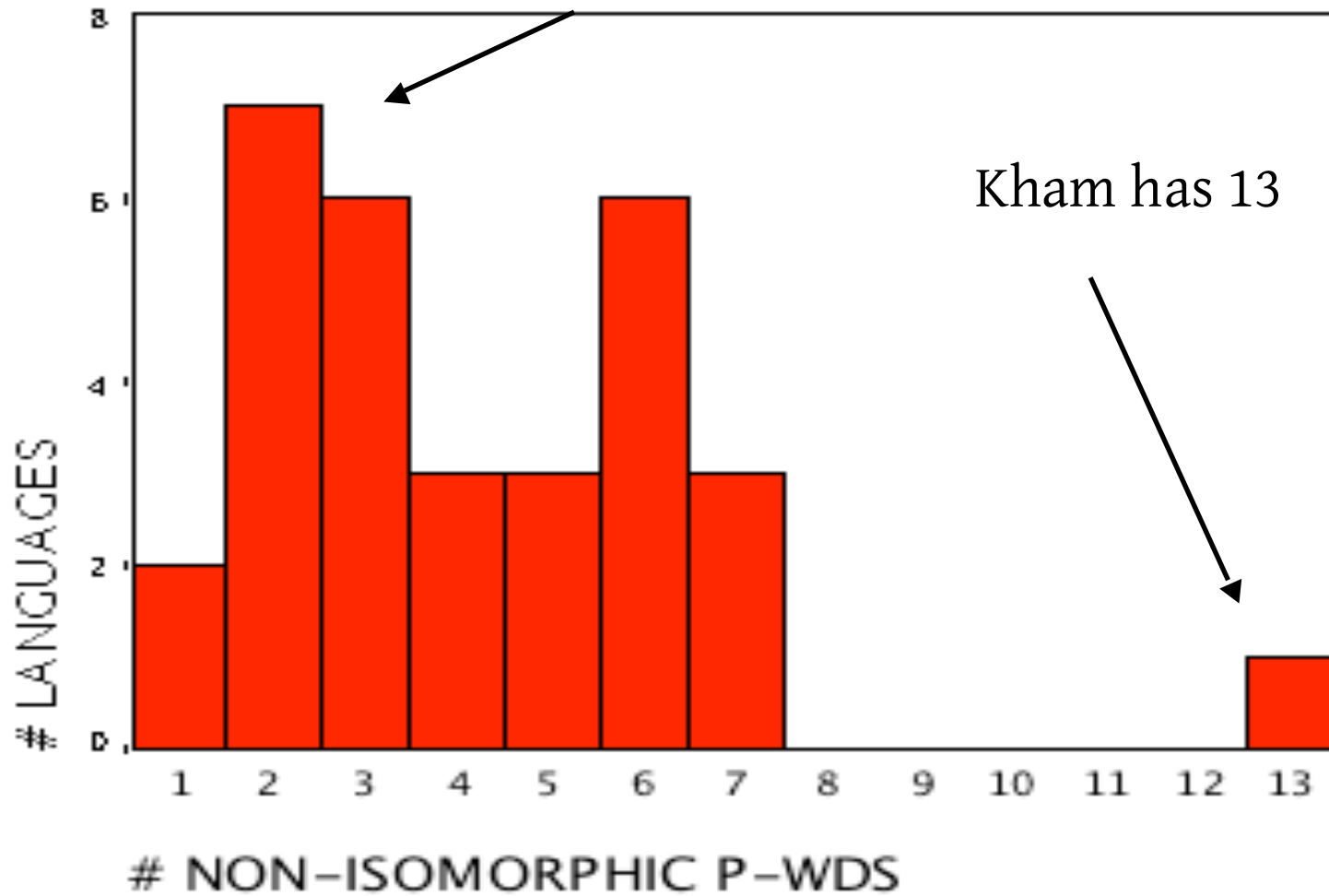


Geographic Distribution: Sino-Tibetan

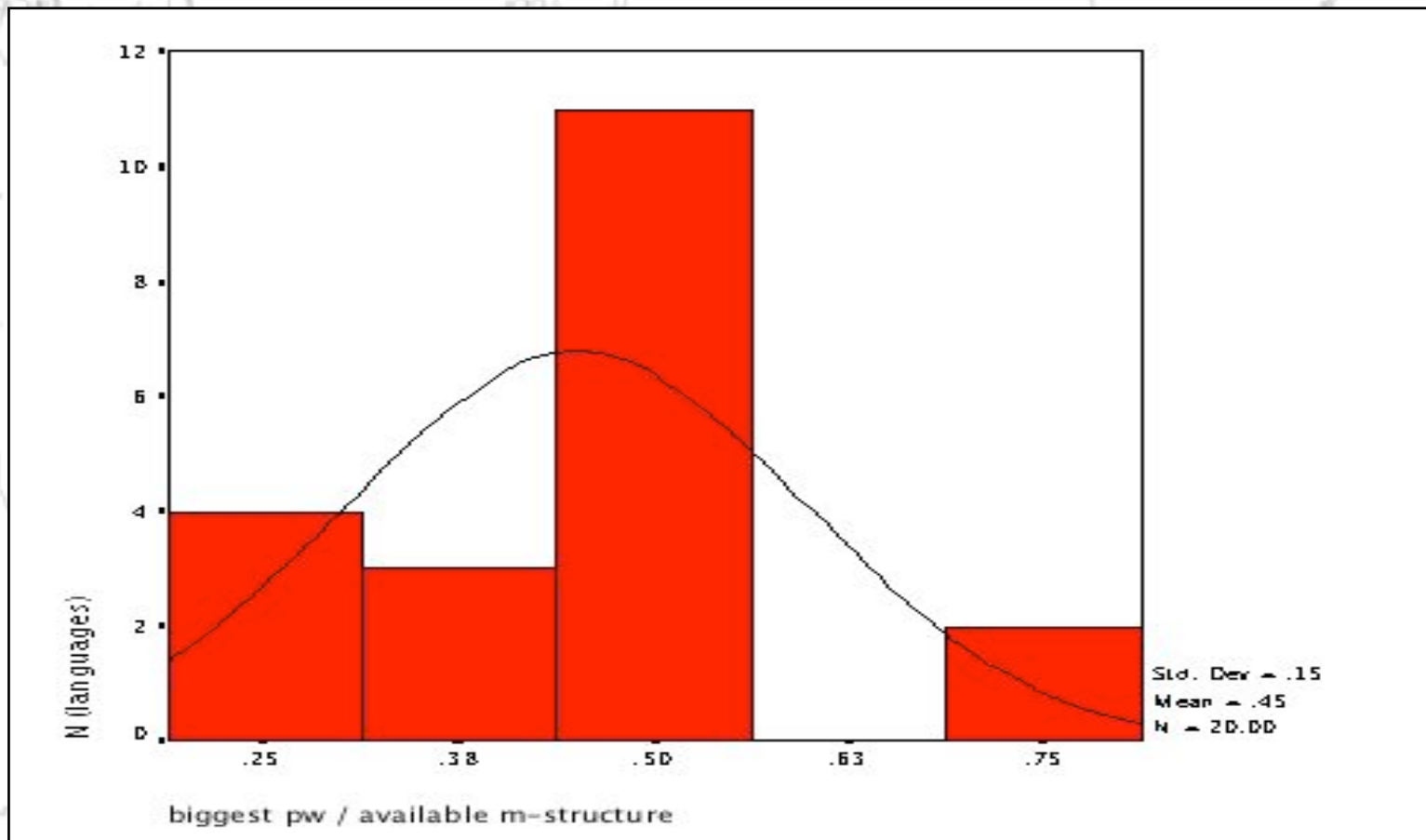


Non-Isomorphism: 27 languages

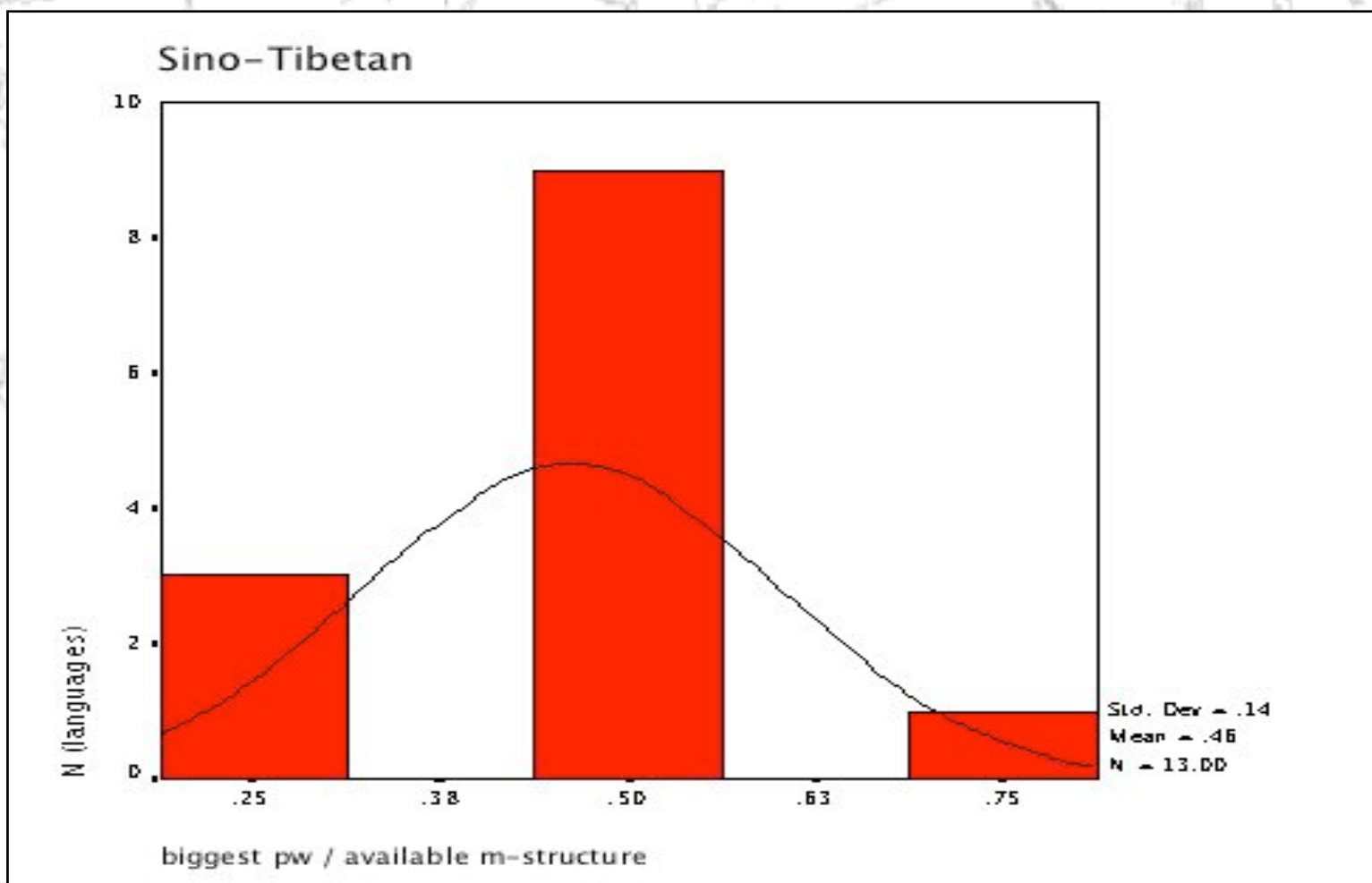
7 languages with 2 non-isomorphic p-words



Based on Available Morpheme Structure



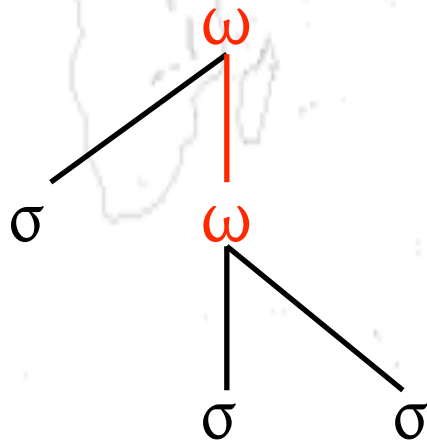
Ratio: Sino-Tibetan only



Recursion: Lahu (Sino-Tibetan)

PW1: Reduced Stress
prefix-stem-suffix

PW2: Stem Tone Change



Recursion: Belhare

- Belhare: 12 P-domains; 6 G-domains
- Bipartite Stems

GD4/PD2, 10: Spans stem, suffixes, particle

GD2: Spans both stem pieces

Prefix- Σ_1 . Σ_2 -Suffix-Particle

GD1: Spans right piece stem

PD5: Contains at least stem + suffix

GD3, 5: Spans stem and all affixes

PD1: Rt. edge stem (+ suffixes + particle)

PD4: Lft. edge stem (+ suffixes + particle)

PD7: Lft. edge prefix + stem (+ suffixes + particle)

PD3, 6, 8, 9: Lft. edge stem only

V. Future Goals

- Catalogue & Typologize on diversity of domain types
- Focus: Sino-Tibetan & Austro-Asiatic vs. 'World' sample
- Enhance database with very detailed information (e.g. Questionnaires)
- Outcome: updating of traditional terminologies: fusion types, affix~clitic~particle

“The Questionnaires”

- <http://www.uni-leipzig.de/~autotyp/download/index.html>
- Designed for use in multiple settings, scenarios:
 - Reading/writing of grammars & descriptive phonology
 - Fieldwork/elicitation
 - Transcription & analysis of text data
- P-domains: a ‘bottom-up’ approach:
 - σ >> polysyllabic, monomorphemic >> polymorphemic >> phrasal/clausal constructions...
 - charting phonology that is sensitive to morphological (& syntactic) structure

AUTOTYP: People and funding

- **The AUTOTYP research team (as of October 2004)**
 - Johanna Nichols (Co-Director, Berkeley)
 - Balthasar Bickel (Co-Director, Leipzig)
 - Tracy Hall (Co-Director, Word Domains, Indiana University)
 - Fernando Zúñiga (Research Associate, Eugene/Zürich/Manchester/Santiago)
 - Kristine Hildebrandt (Post-Doc, Leipzig)
 - *RAs in Berkeley:* Gabriela Caballero, Suzanne Wilhite
 - *RAs in Leipzig:* Michael Riessler, Sven Siegmund, Sindy Poppitz, Franziska Crell, Kathi Stutz, Josh Wilbur, Jenny Seeg, Anja Gampe, Sebastian Hellmann
 - *Past team members:* Sandra Biewald, Aimee Lahaussos-Bartosik, Dave Peterson, Keith Sanders, Alena Witzlack-Makarevich, Rebecca Voll
- **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)