

Tone in Historical Linguistics – A view from Mixtec

Typology Seminar - Faculty of Languages
Uppsala University
16.09.2025

Sandra Auderset
University of Bern

sandra.auderset@unibe.ch

“no other phonological feature is treated with such indifference as tone”
(Hyman 2011)

“The study of sound change has focused overwhelmingly on segmental change, while general and typological knowledge about tone change (excluding tonogenesis) lags far behind. (...) Furthermore, tone is often reported to change faster or more sporadically than segments, leaving more changes to unravel and/or less-regular sound correspondences to handle.” (Campbell 2021)

“This [lack of studies on tone change] can be at least partially attributed to impressionistic statements on the diachronic volatility of tones (...), and a prevailing assumption that tones play at best a minor role in unraveling the history of a language family.” (Auderset, Dockum, Gehrman 2025)



faster change

irregularity &
drastic variation

large number of
correspondence sets

limited availability
of data

loss of conditioning
environments

**LOOK, SIMBA. EVERYTHING THE
LIGHT TOUCHES IS HISTORICAL LINGUISTICS**



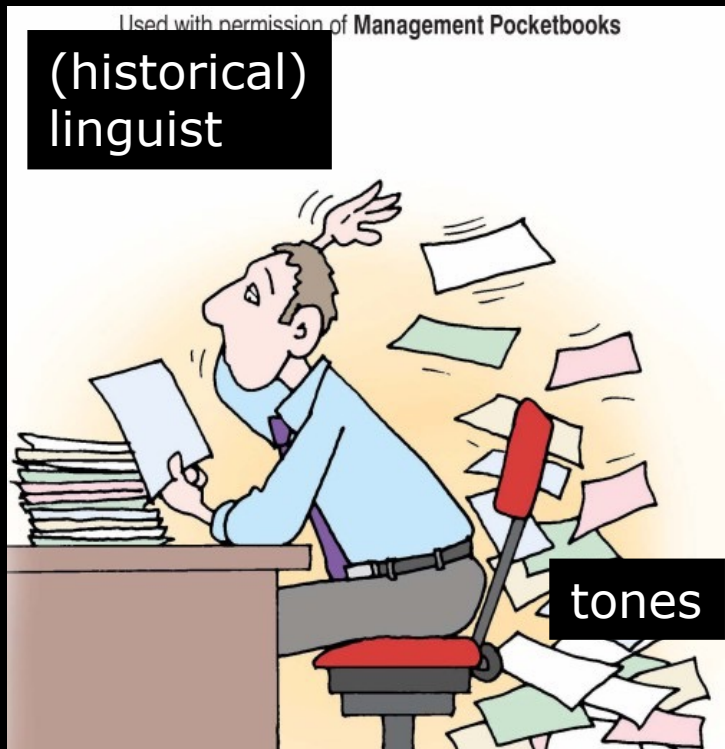
**But what's that
shadowy place over there?**



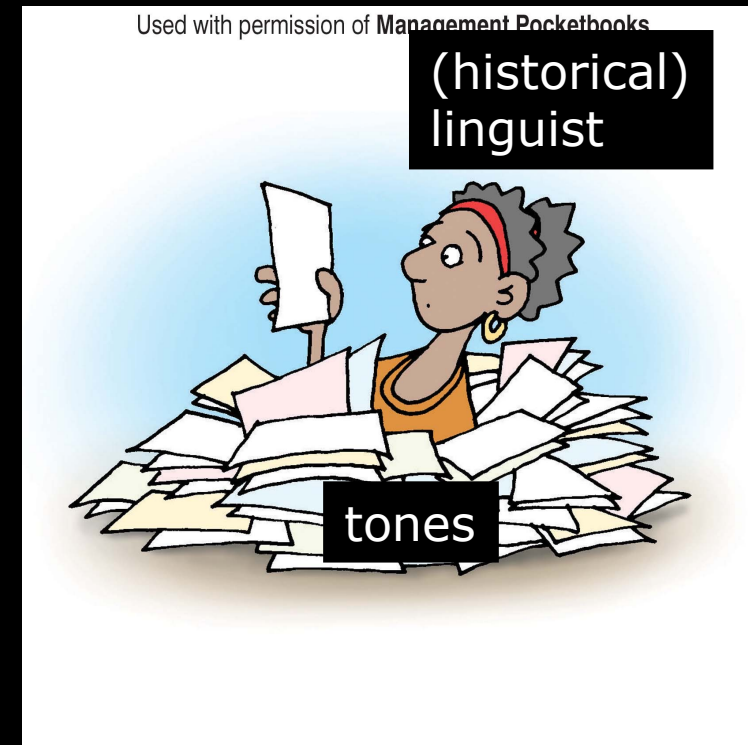
**THAT IS TONE
CHANGE. YOU
MUST NEVER GO THERE**

This talk in a nutshell

NO



YES



Roadmap



Brief intro to Mixtec languages and their tone systems



Case study: Rates of change and phylogenetic signal in Mixtec tone



Pilot study: Distinguishing grammatical and lexical tone through diachrony



Wrap-up and outlook on tone in historical linguistics

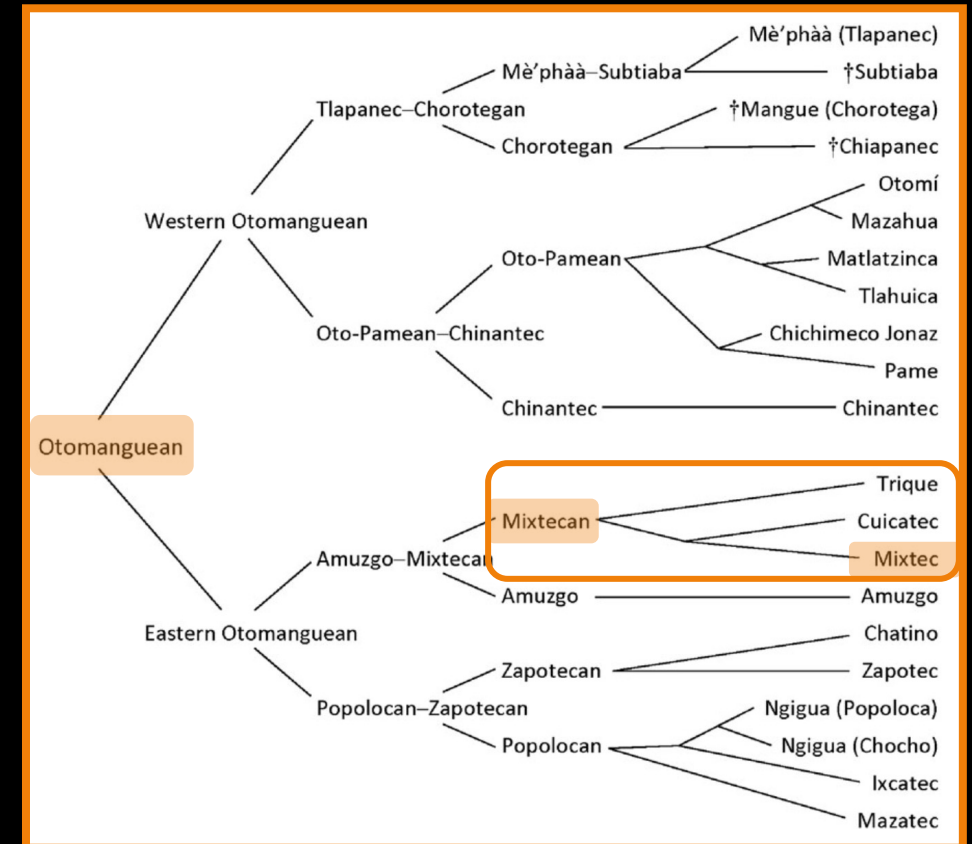
The Mixtec (Tu'un Savi) languages of southern Mexico

Location



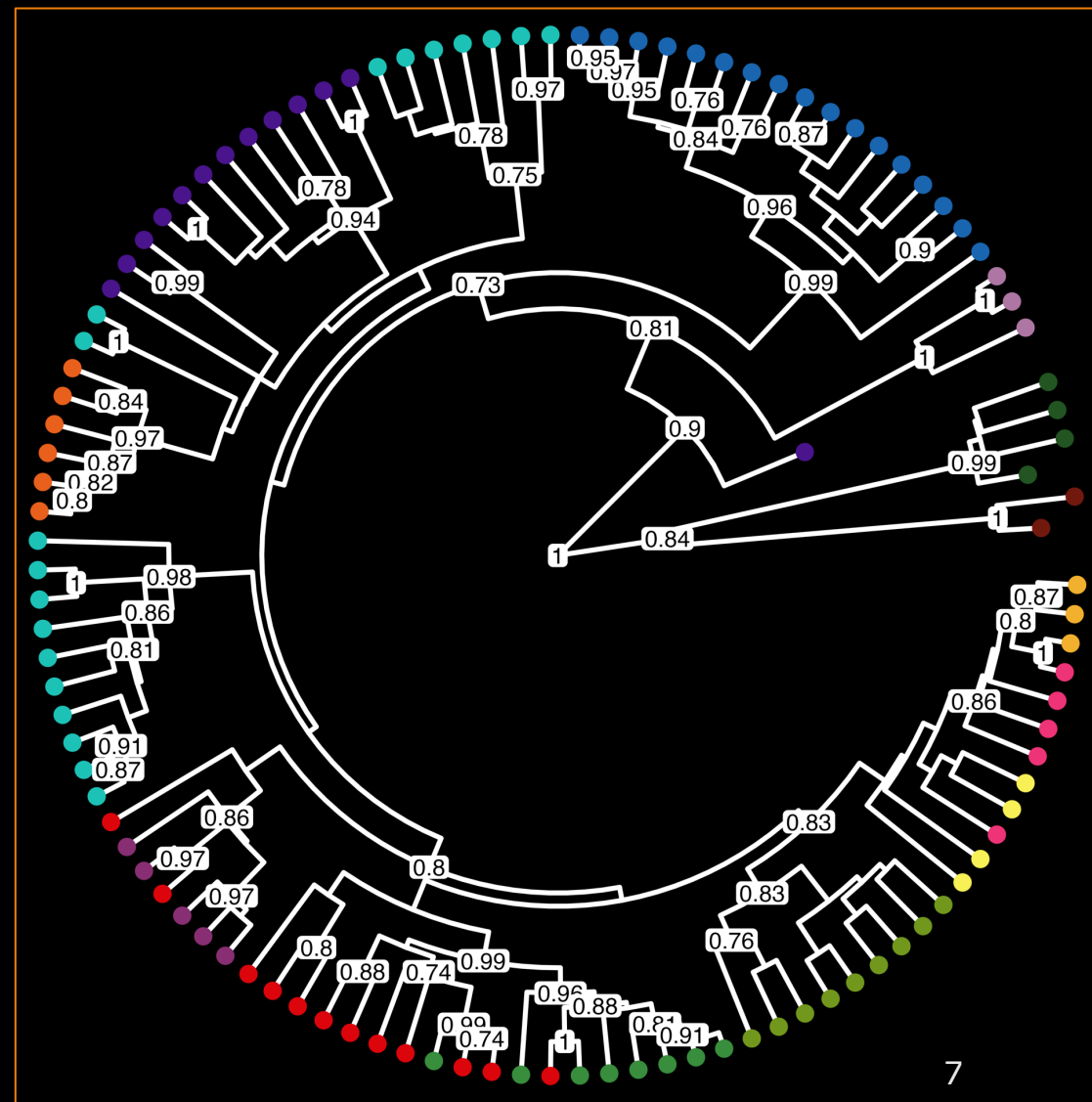
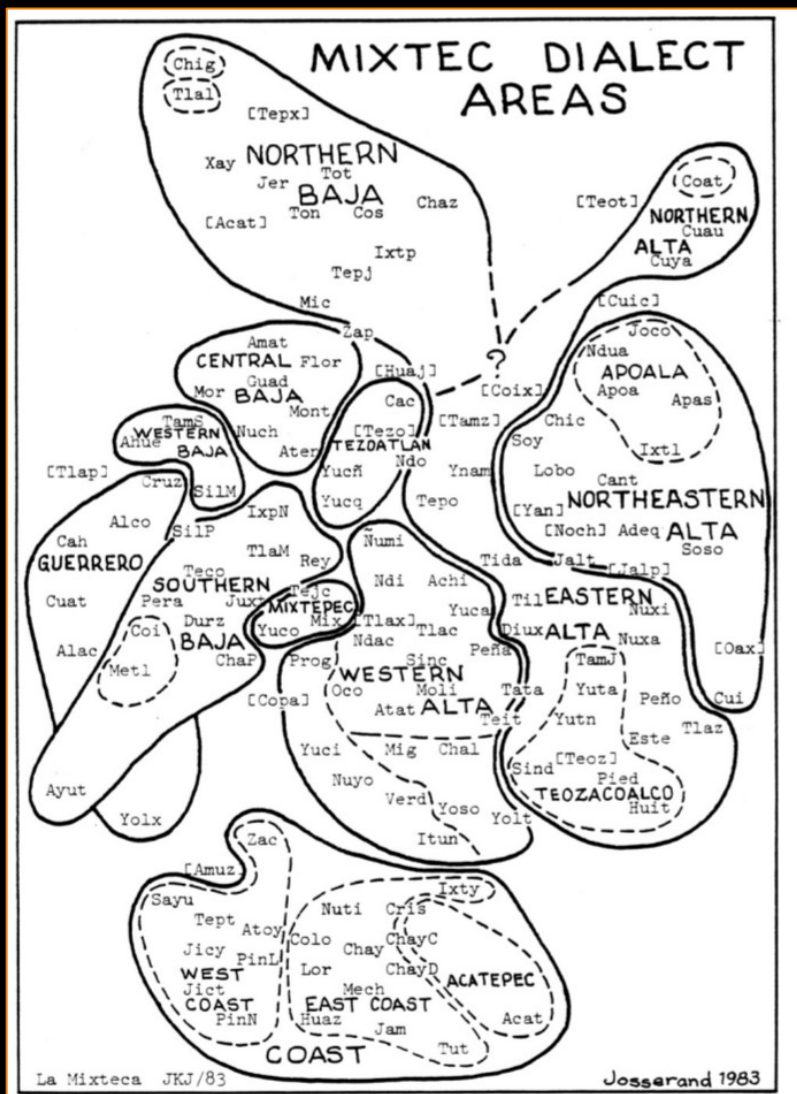
Aguilar Sánchez. 2020:24. *Nuu Savi: Pasado, presente y futuro*. Diss U Leiden

Classification



Campbell, E.W. 2017. *Lang. Ling. Compass* 11(7) (based on Kaufman)

Diversification of Mixtec



Example from San Martín Duraznos Mixtec

soo yoó tiki inkà kì'v-ì ñà íin Tijuána
but exist REP other sister.of.F-1SG 3SG.F INC.live Tijuana
'but there is my other sister who lives in Tijuana,

taa xínì ñákán súvà'-án ñà'a chávà leka
and INC.know 3SG.F.DEM POT.prepare-3SG.F thing other bag
'and she knows how to make other types of (woven) bags'

núchí kitxì-va ntá-an
pretty much=INTENS COP=3GEN
'they are very pretty'



Some phonological basics

- lexical items in Mixtec are always at least bimoraic (mora = short vowel)

(C)V^T(C)V^T

- grammatical markers can also be monomoraic (or have no mora)

(C)V^T | C(V^T)

- there are limited consonant clusters and no final consonants
- final vowels show contrastive vowel length and nasalization

Tlahuapa Mixtec

Orth.	IPA	Gloss
iín	i ³ ĩ ⁵	peel off
xaa	ʃa ³ a ³	start
kàxi	ka ¹ ʃĩ ³	choose
ka'nta	ka ³ ʔ ⁿ da ³	cut sth.
sntaxi	s- ⁿ da ³ ʃĩ ³	make wet [CAUS-be.wet]
sntákòo	s- ⁿ da ⁵ -ko ¹ o ³	make sb. stand up again [CAUS-ITER-stand.up]

Tones in Mixtec

- TBU (tone bearing unit) is the mora
- tone systems analyzed with 2, 3, or 4 level tones; some with 1-3 contour tones (usually rising and falling)

standardized notation based on
Chao 1930

Chao's number	Musical comparison	Label
5	G#	high
4	F	half-high
3	E	mid
2	D	half-low
1	C	low

contrastive tone melodies in
Duraznos Mixtec (own data)

level tones only		
1.1	3.1	5.1
1.3	3.3	5.3
1.5	3.5	5.5

contrastive tone melodies in Yoloxóchitl Mixtec
(Castillo García 2007)

level tones			contour tones	
1.1		5.1	1.32	
	3.2	5.2		13.2
1.3	3.3	5.3	3.52	15.3
1.5	3.5	5.5	5.25	15.5

The functional load of tone

Lexicon

Form	Meaning
ni ¹ i ¹	salt
ni ¹ i ³	skin
ni ⁵ i ⁵	hail
n [̃] da ¹ ki ³	broom
n [̃] da ¹ ki ⁵	pozole
n [̃] da ³ ki ³	strong
i ³ ʃi ¹	hair
i ¹ ʃi ⁵	hairy, bearded

Grammar

Irrealis	Imperfective	Meaning
kã ³ ã ³	kã ⁵ ã ⁵	make holes
kã ¹ ã ³	kã ⁵ ã ³	get used to
ki ³ ʔβi ¹	ki ⁵ ʔβi ¹	be sick
ki ¹ ʔβi ³	ki ⁵ ʔβi ³	enter
n [̃] da ³ n [̃] do ⁵ so ⁵	n [̃] da ⁵ n [̃] do ⁵ so ⁵	forget
tʃi ¹ n [̃] dʒe ³ e ⁵	tʃi ⁵ n [̃] dʒe ³ e ⁵	help

Examples from Duraznos Mixtec (own data)

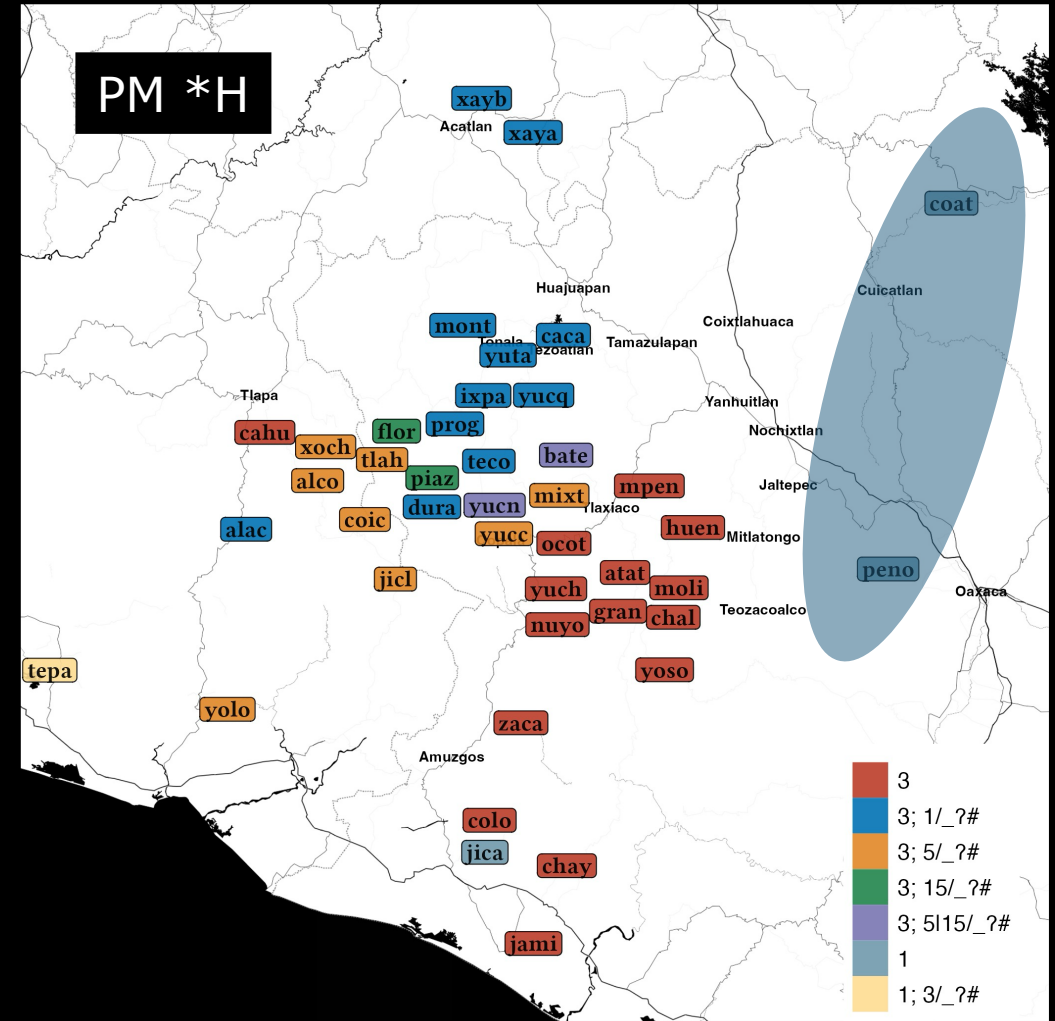
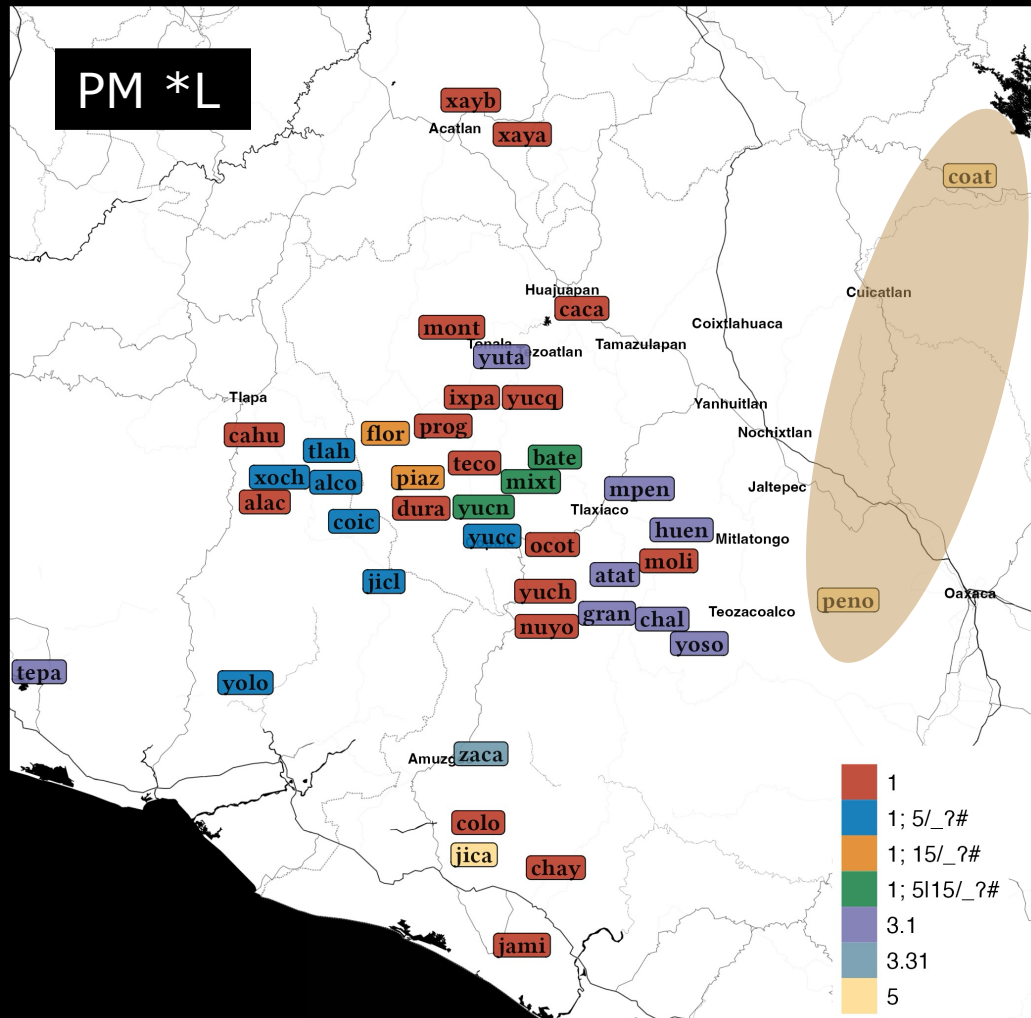
Reconstructing tone

- tone can be reconstructed to Proto-Mixtec (and Proto-Mixtecan and beyond)
- correspondences are regular and relatively well understood for 'basic' vocabulary

Variety	ANIMAL	PALM MAT	Tones
Proto-Mixtec	*ki ^H ti ^H ʔ	*ju ^H wi ^H ʔ	*HHʔ
Peñoles	ki ¹ ti ¹	ʒu ¹ u ¹	1.1
Zacatepec	ki ³ ti ³ ʔ	ju ³ βi ³ ʔ	3.3
Cacaloxtepec	ki ³ ti ¹	ʒu ³ u ¹	3.1
Xochapa	ki ³ ti ⁵	i ³ βi ⁵	3.5
Piedra Azul	ki ³ tsi ¹⁵	ji ³ βi ¹⁵	3.15

Area A		Proto-Mixtec		Area B
MID		*high		LOW
LOW	←	*low	→	HIGH
HIGH		*ʔ# (modified)		lowering

Reconstructing tone



Reconstructing tone

Label	Reconstruction	Possible environment	Swanton and Mendoza Ruiz (2021)
A ₁	*HH		*HH
A ₂	*HH?		*HH?
(A ₃)	*?-HH	TBD	—
A ₄	*H ^A -HH(?)	adjectival high tone	*?²-HH
A ₅	*H ^A -HH(?)	adjectival high tone	*?¹-HH
A ₆	*L-HH(?)	low tone from classifier	—
A ₇	*?-HH(?)	TBD	*?-HH?
A ₈	*?-HH?	TBD	*HH?
B ₁	*LL		*LL
B ₂	*LL?		*LL?
(B ₃)	*?-LL	TBD	*?-LL
(B ₄)	*?-LL	TBD	—
B ₅	*H ^A -LL?	adjectival high tone	—
B ₆	*?-LL?	TBD	*?-LL?
C ₁	*HL		*HL
C ₂	*?-HL	TBD	*HL
C ₃	*?-HL	TBD	*HL
C ₄	*H ^A -HL	adjectival high tone	*?-HL
C ₅	*L-HL	low tone	*L-HL, *L?-HL
C ₆	*?-HL		—
D ₁	*LH		*LH

Case study (Auderset 2024)

How can we empirically assess whether tone change is more rapid and irregular than segmental change or not?

Data:

- tone is old in Mixtec
- previous reconstructions, including tone
- recent documentation efforts = more & better data
- lots of variation between varieties
- Bayesian phylogenetic study (Auderset et al. 2023)
- database of sound changes (Auderset & Campbell 2024)

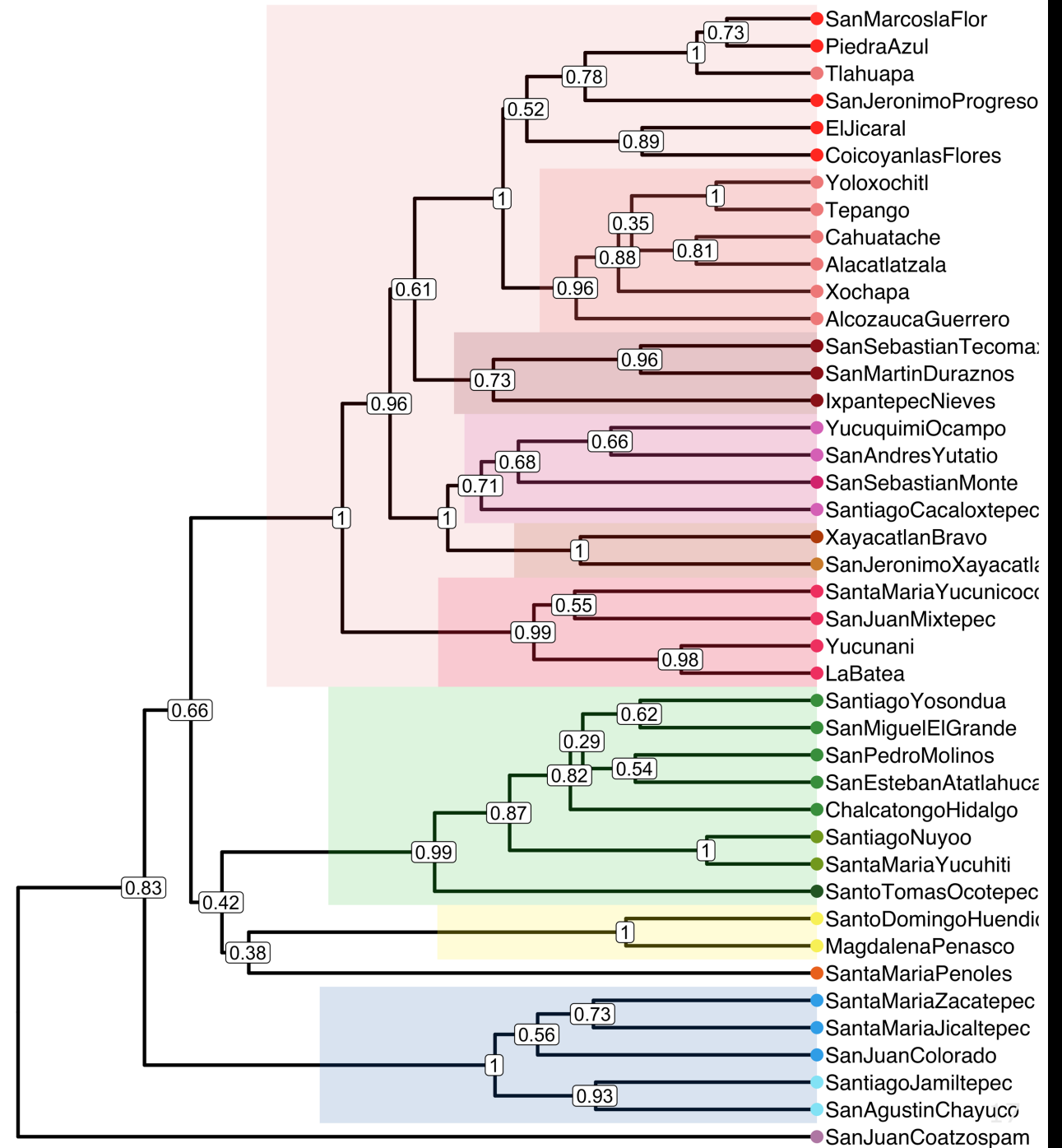
Methods:

- measuring how 'fast' tone changes vs. segments: evolutionary rates (hidden Markov model)
- measuring how well tone changes align with subgrouping (tree topology) vs. segmental changes: phylogenetic signal (D)



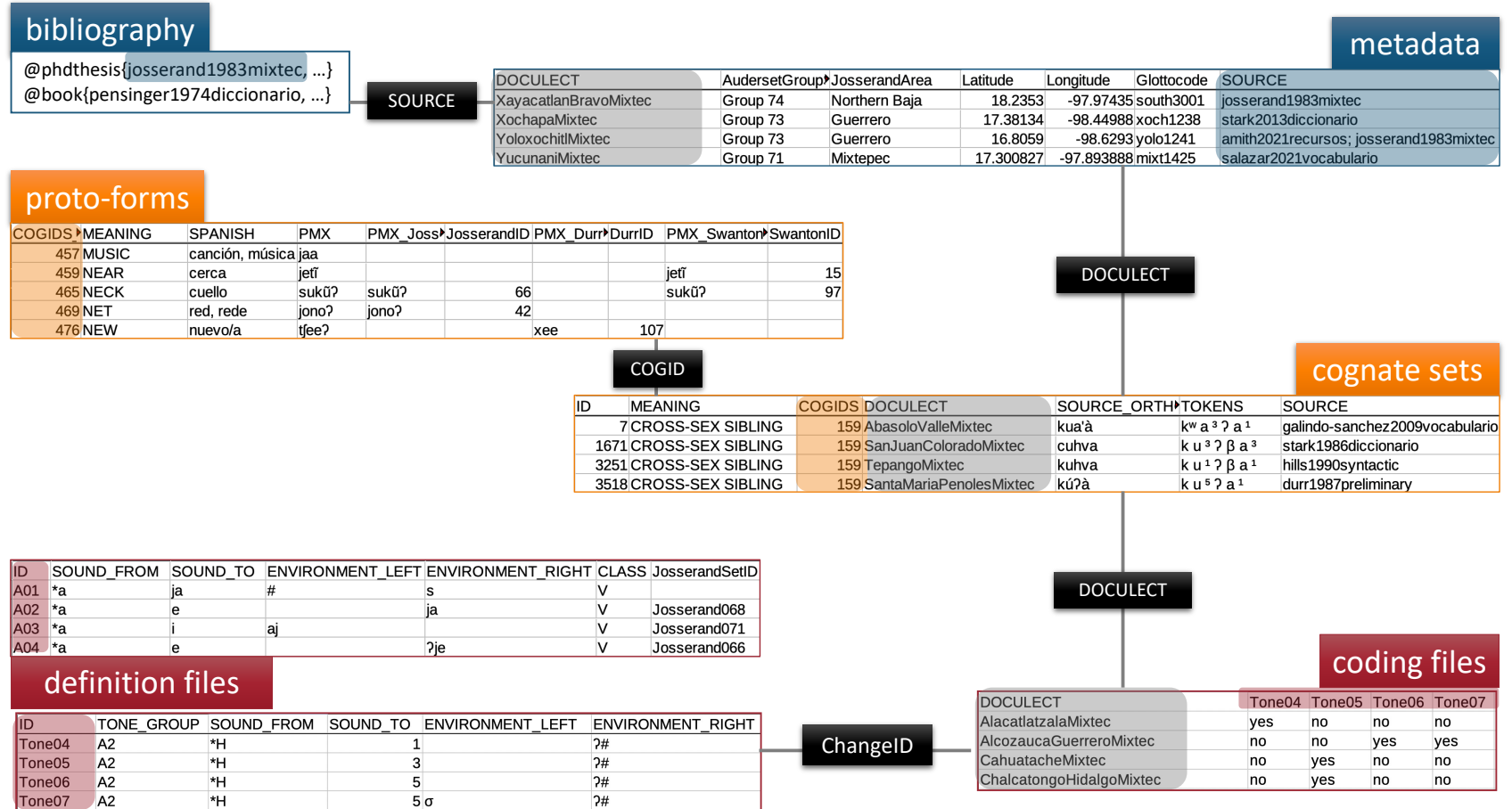
Phylogenetic Trees

M(aximum) C(lade) C(redibility)
tree



Annotation of sound changes

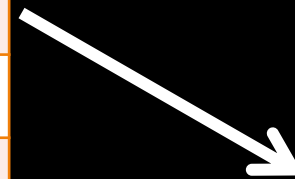
- AUTOTYP design principles:
- modularity & connectivity
 - autotypology
 - separation of definition and coding files
 - late aggregation



Tone correspondences & variables

Variety	ANIMAL	PALM	MAT Tones
Proto-Mixtec	*ki ^H ti ^H ʔ	*ju ^H wi ^H ʔ	*HHʔ
Peñoles	ki ¹ ti ¹	ʒu ¹ u ¹	1.1
Zacatepec	ki ³ ti ³ ʔ	ju ³ βi ³ ʔ	3.3
Cacaloxtepec	ki ³ ti ¹	ʒu ³ u ¹	3.1
Xochapa	ki ³ ti ⁵	i ³ βi ⁵	3.5
Piedra Azul	ki ³ tsi ¹⁵	ji ³ βi ¹⁵	3.15

correspondence sets



change variables

Variable	PM	Reflex	Environment
Tone01	*H	1	#_
Tone02	*H	3	#_
Tone04	*H	1	_?#
Tone05	*H	3	_?#
Tone08	*H	15	_?#

Methods & Analysis

phylogenetic signal:

D-metric

- sum of sister-clade differences across a tree
- for binary traits

input:

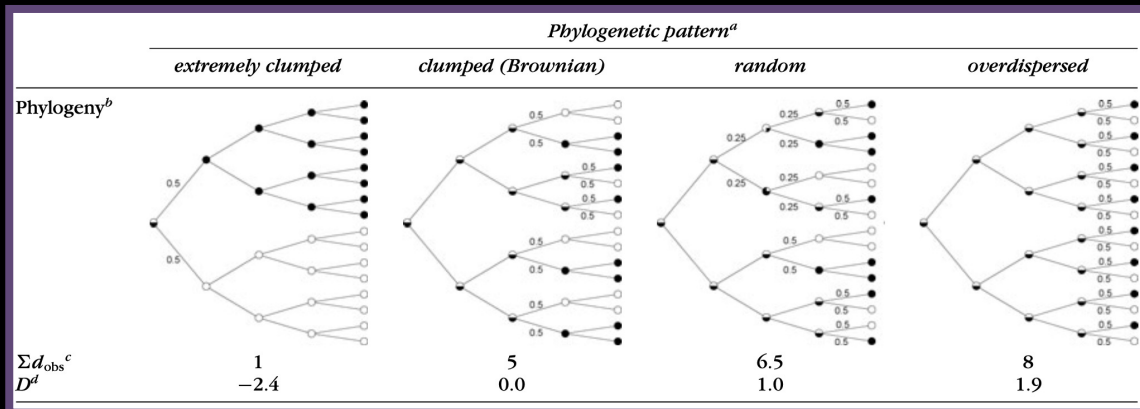
coded sound changes

sample of 1000 trees from posterior distribution

evolutionary rates:

hidden Markov model

- transition rates between presence and absence of a trait across a tree
- all rates differ (ARD)

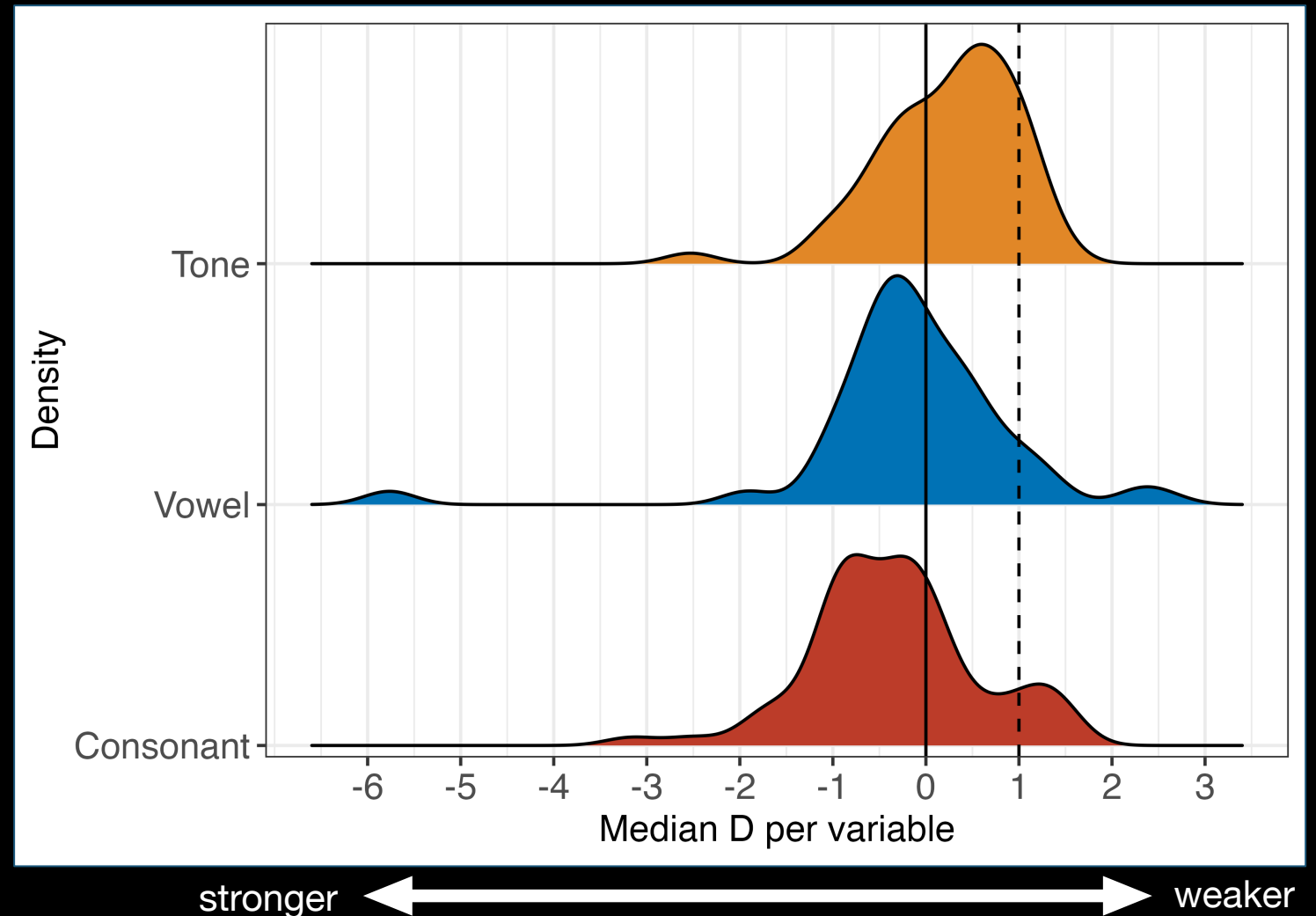


Phylogenetic signal in sound changes

Are tone changes distributed over subgroups significantly differently from segmental changes?

Interpretation of D

$D < 0$	overclumped
$D \approx 0$	clumped
$D \approx 1$	random
$D < 1$	overdispersed

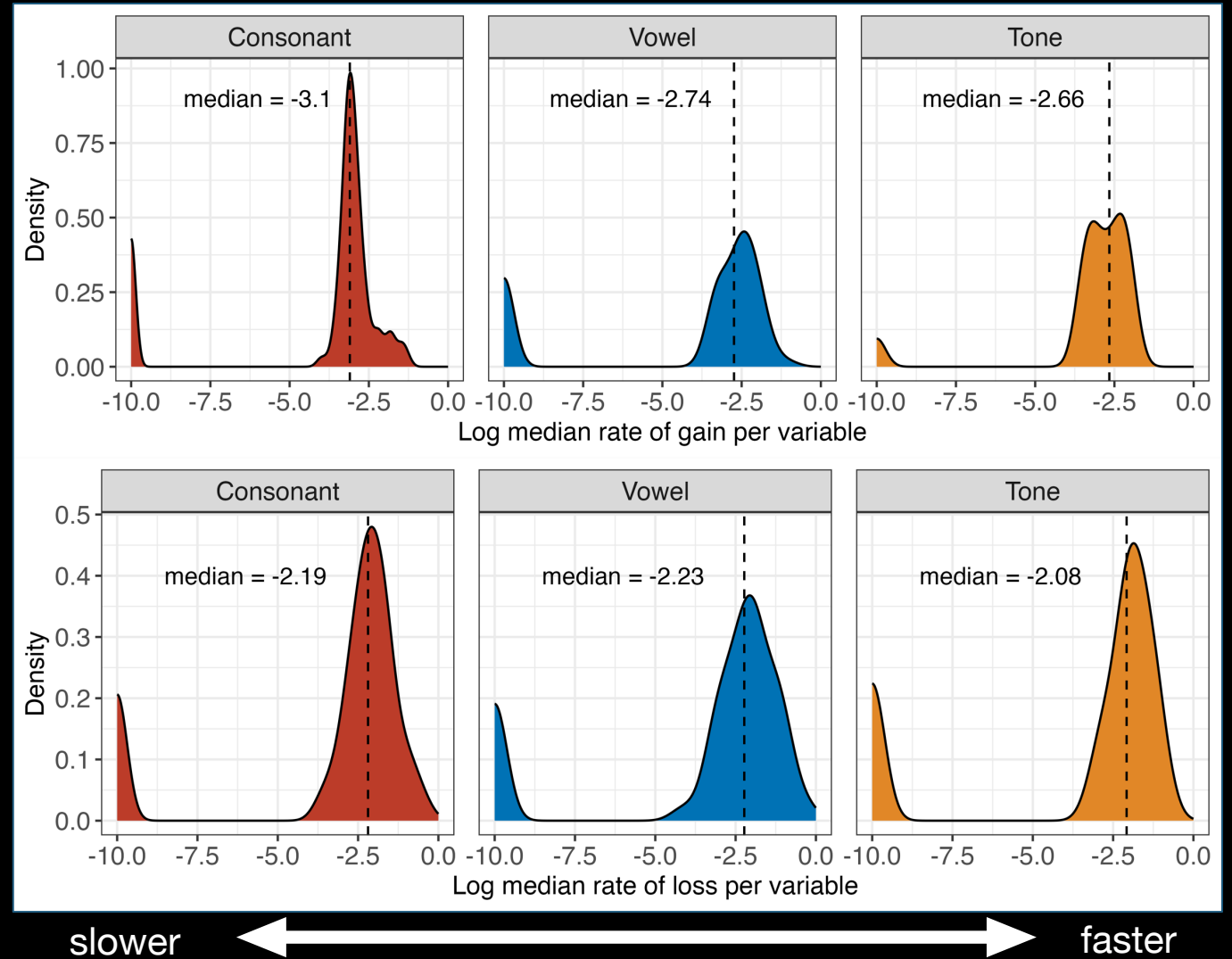


Rates of sound changes

Are tone changes generally gained and lost faster than segmental ones?

Interpretation of log rates:

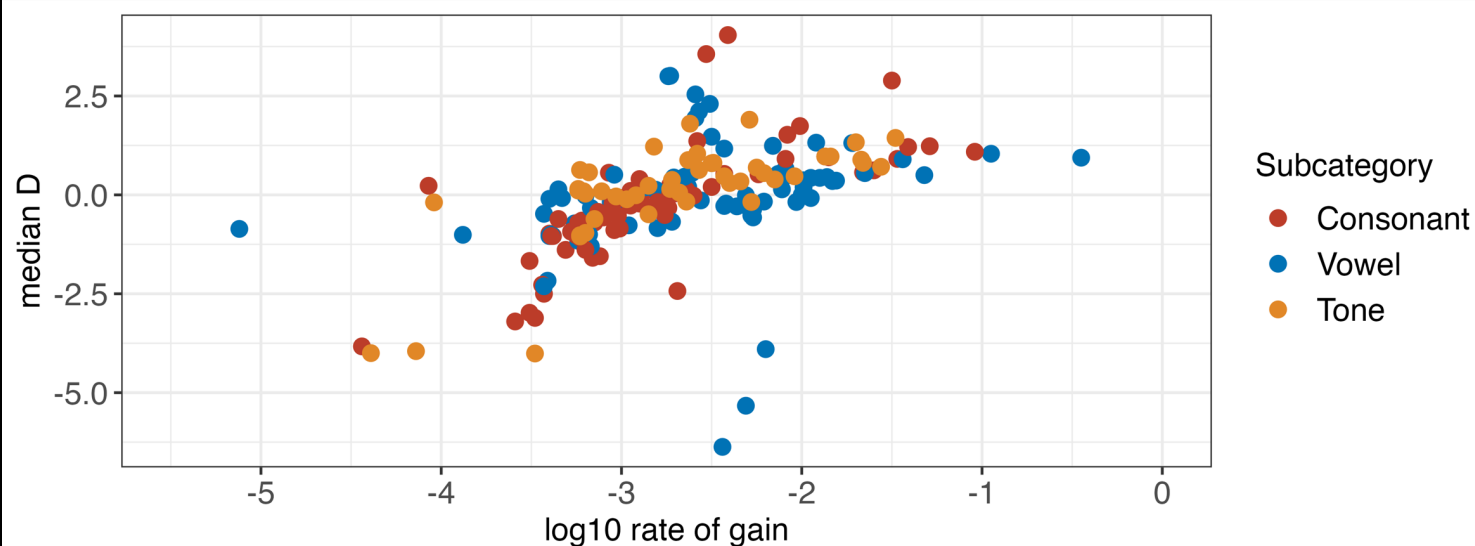
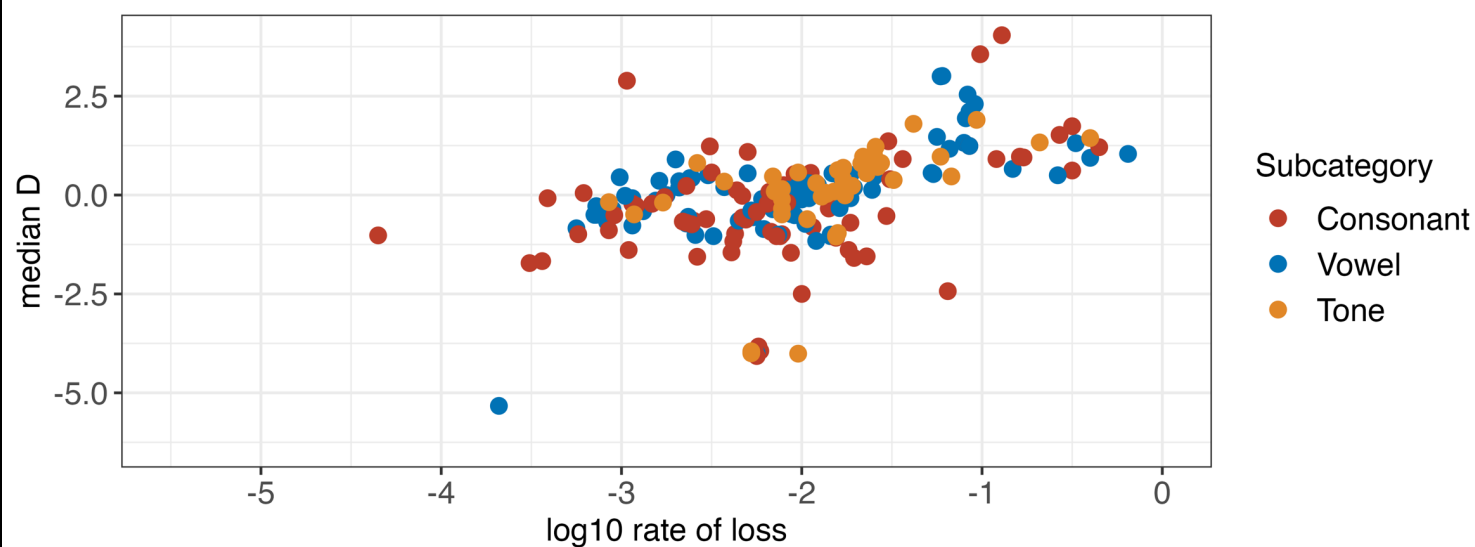
0	fast
-5	slower
-10	no change



Summary

no signal
↑
↓
strong signal

Rate of loss and gain (log10) against D colored by type of sound change variable



slower ← → faster

Pilot study

Does grammatical tone change differently from lexical tone? Can grammatical tone be delineated from lexical tone based on diachrony?

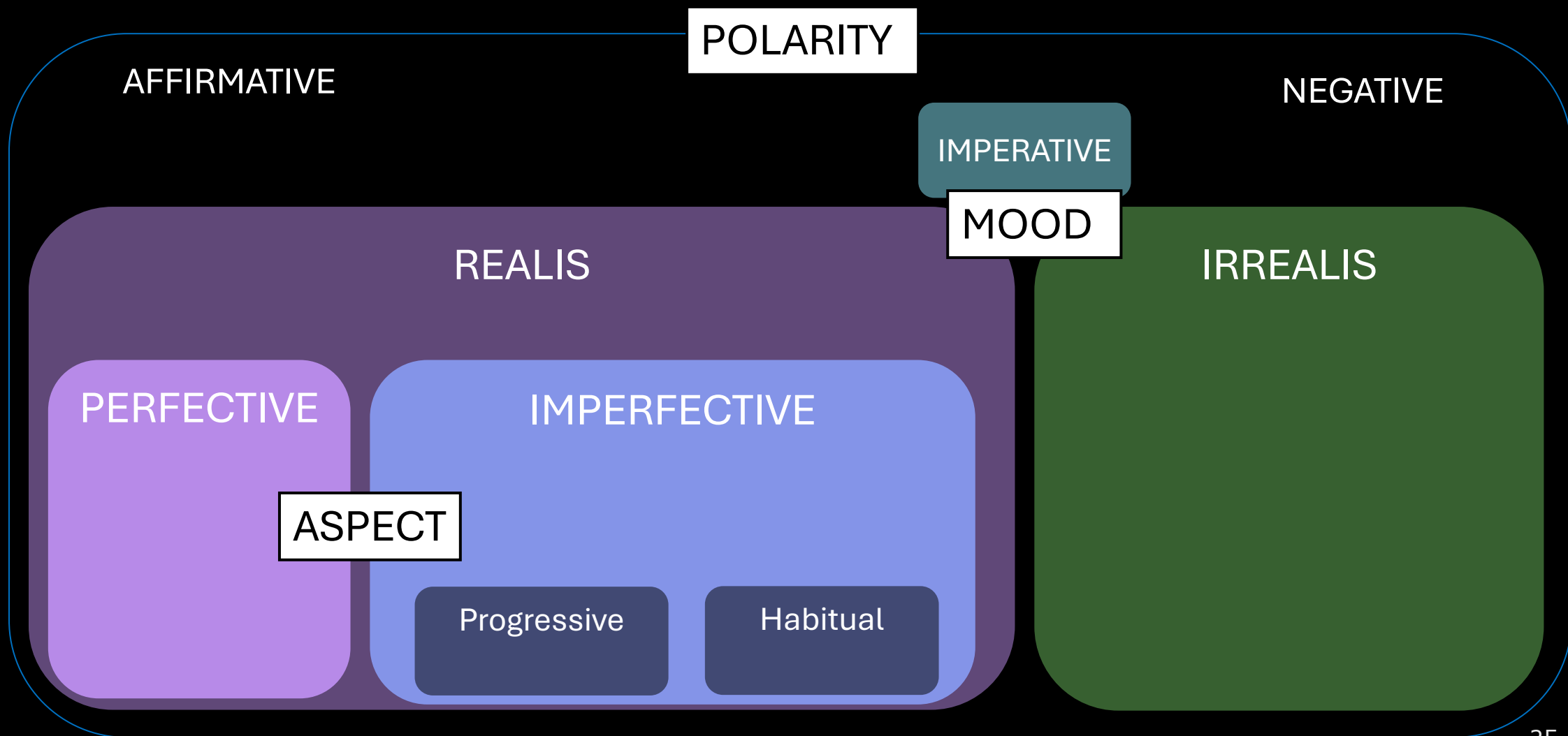
Data:

- cognate-coded lexical word lists (Auderset et al. 2023)
- database of verb paradigms, coded for segmental & tonal marking and cognacy (~12 varieties, in progress)

Methods:

- compare correspondence sets in inflected verb forms to lexical correspondences
- reconstructing marker(s) of categories
- grammaticalization?

Aspect-Mood-Polarity categories



Aspect-Mood-Polarity categories

- all verbs (with a few exceptions) have an affirmative and negative form of the
 - imperfective
 - perfective
 - irrealis
- so the 'paradigm' is small BUT
 - there are inflectional classes because these categories are marked by various tonal and segmental markers that cross-cut each other
 - rough estimate: ~30 classes per variety

Encoding of AMP categories

- exact configuration varies:
- between each languages
 - for each inflectional class

Category	Prefixes	Stem Alt.	Suppletion	Tone
Irrealis	✓	✓	✓	✓
Imperfective	✓	✓	✓	✓
Perfective	✓	×	×	✓
Imperative	×	×	×	✓
NEG.IRR	✓	×	×	✓
NEG.REAL	✓	×	×	×

Encoding of AMP categories

exact configuration varies:

- between each languages
- for each inflectional class

most frequent/'default' marking circled in red

Category	Prefixes	Stem Alt.	Suppletion	Tone
Irrealis	✓	✓	✓	✓
Imperfective	✓	✓	✓	✓
Perfective	✓	×	×	✓
Imperative	×	×	×	✓
NEG.IRR	✓	×	×	✓
NEG.REAL	✓	×	×	×

Grammatical tone?

Variety/Form	IRR	PFV	IMPF	IRR	PFV	IMPF
S. J. Colorado	kwa ¹ ndo ¹ o ¹	ndo ¹ o ¹	ndo ⁵ o ⁵	kwa ¹ ku ³ nu ³	tsi ³ nu ³	tsi ⁵ nu ⁵
M. Peñasco	ndo ³ o ³	ni ¹ ndo ³ o ³	ndo ⁵ o ³	ku ³ nu ³	ni ¹ hi ³ nu ³	hi ⁵ nu ³
S. M. Duraznos	ndo ³ o ³	i ¹ -ndo ³ o ³	ndo ⁵ o ³	ku ³ nu ³	i ¹ -ɕi ³ nu ³	ɕi ⁵ nu ³
Piedra Azul	ndo ³ o ³	ndo ¹ o ³	ndo ⁵ o ³	ko ³ nu ³	ʃi ¹ nu ³	ʃi ⁵ nu ³
Yoloxóchitl	ndo ³ o ³	ndo ¹³ o ³	ndo ⁵ o ⁵	ku ³ nu ³	ʃi ¹³ nu ³	ʃi ⁵ nu ³
S. A. Yutatío	ndo ³ o ³	ni ¹ ndo ³ o ³	ndo ⁵ o ³	kwi ³ no ³	ni ¹ ʃi ³ no ³	ʃi ⁵ no ³
gloss	‘clean, purify, fade’			‘run, flee’		

Grammatical tone?

Variety	IRR	IMPF	PFV	IRR	IMPF	PFV
S. M. Duraznos	$n\text{t}\phi i^3\gamma i^3$	$n\text{t}\phi i^5\gamma i^3$	$i^1 + n\text{t}\phi i^3\gamma i^3$	$sa^5 + n\text{d}\phi i^3\gamma i^3$	$sa^5 + n\text{d}\phi i^3\gamma i^3$	$i^1 + sa^5 + n\text{t}\phi i^3\gamma i^3$
Piedra Azul	$n\text{tsi}^3\gamma i^3$	$n\text{tsi}^5\gamma i^3$	$n\text{tsi}^1\gamma i^3$	$si^3 + n\text{dzi}^3\gamma i^3$	$si^5 + n\text{dsi}^3\gamma i^3$	$si^1 + n\text{tsi}^1\gamma i^3$
gloss	'end, be finished'			'finish sth.'		
M. Peñasco	$ka^3\gamma\tilde{a}^1$	$ka^5\gamma\tilde{a}^1$	$ni^1 + ka^3\gamma\tilde{a}^1$	$n\text{da}^3 - ka^3\gamma\tilde{a}^1$	$n\text{da}^5 - ka^3\gamma\tilde{a}^1$	NA
S. A. Yutatío	$ka^3\gamma\tilde{a}^1$	$ka^5\gamma\tilde{a}^1$	$ni^1 + ka^3\gamma\tilde{a}^1$	$na^5 - ka^3\gamma\tilde{a}^1$	$na^5 - ka^3\gamma\tilde{a}^1$	$ni^1 na^1 - ka^3\gamma\tilde{a}^1$
gloss	'say, speak'			'repeat, say again'		

Verbal correspondences

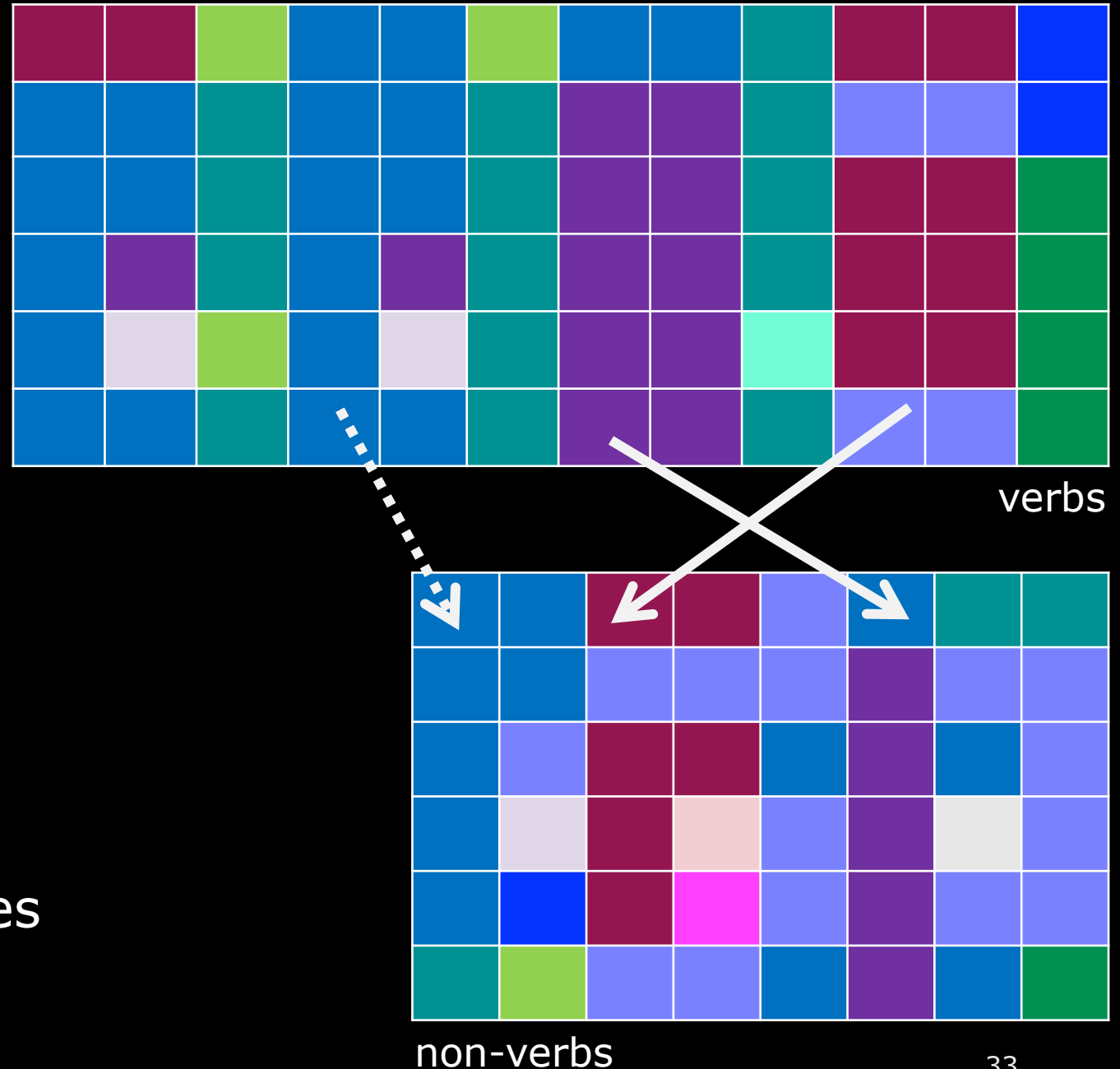
Form/ Variety	IRR	PFV	IMPF	IRR	PFV	IMPF	IRR	PFV	IMPF	IRR	PFV	IMPF
S.J. Colorado	1+1.1	1.1	5.5	1+3.3	3.3	5.5	1+3.3	3.3	5.3	1+1.1	1.1	3.5
M. Peñasco	3.3	1+3.3	5.3	3.3	1+3.3	5.3	1.3	1+1.3	5.3	3.1	1+3.1	3.5
S.M. Duraznos	3.3	1+3.3	5.3	3.3	1+3.3	5.3	1.3	1+1.3	5.3	1.1	1+1.1	5.1
Piedra Azul	3.3	1.3	5.3	3.3	1.3	5.3	1.3	1+1.3	5.3	1.1	1+1.1	5.1
Yoloxóchitl	3.3	13.3	5.5	3.3	13.3	5.3	1.3	1.3	5.13	1.1	1.1	5.1
S.A. Yutatío	3.3	1+3.3	5.3	3.3	1+3.3	5.3	1.3	1+1.3	5.3	3.1	1+3.1	5.1
PM	*n ^{do} ho ^H ? 'clean'			*ku ^H no ^H 'run'			*ki ^L ?βi ^H 'enter'			*ku ^L su ^L 'sleep'		

Lexical correspondences

Corr. Set/ Variety	A1 *HH	A2 *HH?	B1 *LL	B2 *LL?	C1 *HL	D1 *LH	C2 (*X-HL)	C3 (*Y-HL)
S.J. Colorado	3.3	3.3	1.1	1.1	3.1	3.3	5.3	5.3
M. Peñasco	3.3	3.3	3.1	3.1	3.1	1.3	3.1	3.1
S.M. Duraznos	3.3	3.1	1.1	1.1	3.3	1.3	3.3	3.1
Piedra Azul	3.3	3.15	1.1	1.15	3.1	1.3	?	3.1
Yoloxóchitl	3.3	3.5	1.1	1.5	3.1	1.3	3.1	3.1
S.A. Yutatío	5.3	5.5	3.1	3.1	3.3	1.3	3.3	5.1

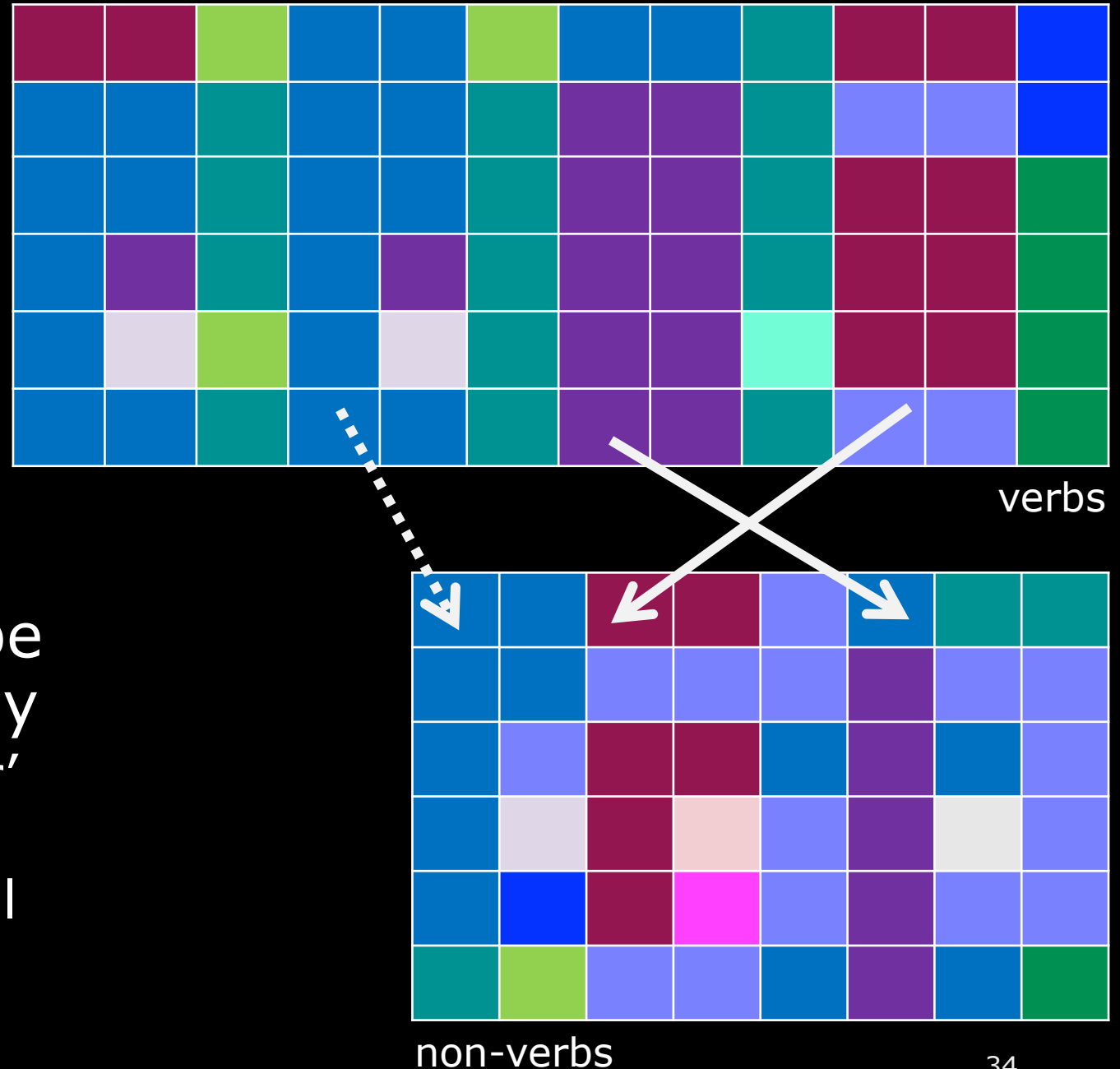
Comparison

- the IRR sets are (largely) the same as found on other word classes -> 'lexical' tone
- the IMPF sets are different and not found elsewhere -> 'grammatical' tone
 - sets all involve high tones
 - many homologous sets



Comparison

- the IRR sets go back to the same PM tone melodies already established
- the IMPF sets need to be reconstructed separately
 - do not show the 'regular' reflexes
 - most likely no segmental origin



Recap and outlook

The idea that tone has little to contribute to historical linguistics comes from various biases and historical contingencies

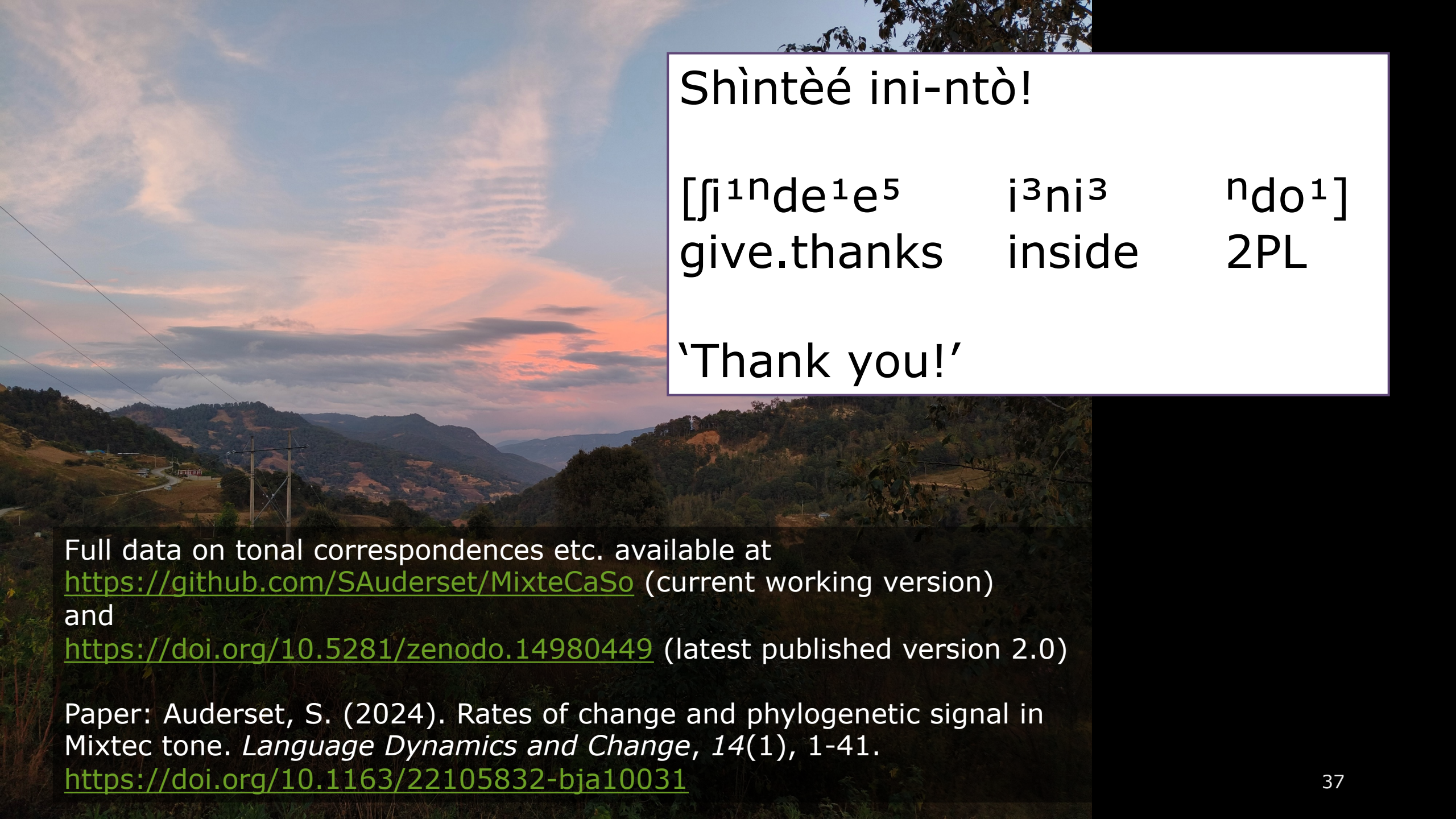
- Eurocentrism
- focus on tonogenesis (emergence vs. change)
- focus on SEA, where tone generally has a low functional load
- conflation of phonetic variation with instability/irregularity
- general data scarcity because tones are often omitted in descriptions (still!)
- etc.

Recap and outlook

In Mixtec:

- tone is reconstructable and the correspondences are regular
- tones do not change faster than segments overall
- tone changes do not show less phylogenetic signal than segmental changes overall
- tones that mark grammatical categories can be distinguished from 'lexical' tones by correspondence sets (i.e. different origin and evolution)

tone is an integral part of the language history
and we should study it as such



Shìntèé ini-ntò!

[ʃi¹n[̃]de¹e⁵ i³ni³ n[̃]do¹]
give.thanks inside 2PL

‘Thank you!’

Full data on tonal correspondences etc. available at
<https://github.com/SAuderset/MixteCaSo> (current working version)
and
<https://doi.org/10.5281/zenodo.14980449> (latest published version 2.0)

Paper: Auderset, S. (2024). Rates of change and phylogenetic signal in
Mixtec tone. *Language Dynamics and Change*, 14(1), 1-41.
<https://doi.org/10.1163/22105832-bja10031>