

(De-)constructing classes: a typological-historical approach to verbal inflection in Mixtec languages

Sandra Auderset
MPI-EVA Leipzig*
sandra_auderset@eva.mpg.de

*from 1.8.: Universität Bern

Kolloquium 11.05.2023
Institut für Linguistik
Universität zu Köln

Roadmap

1. Background on inflectional classes and Mixtec languages
2. Mixtec aspect-mood marking
3. Database design
4. Applications (pilot)
5. Wrap up



Collaborators

database/paper co-authors:



Simon L. Peters (PhD Candidate, UC Santa Barbara)



Inî G. Mendoza (PhD Student, U Chicago)



Zoe Poirier Maruenda (student assistant, MPI-EVA)

data contributors:

Carmen Hernández Martínez (S. Martín Duraznos Mixtec)

Inî G. Mendoza (Piedra Azul Mixtec)

Griselda Reyes Basurto (Tlahuapa Mixtec)

Jeremías Salazar (Yucunani Mixtec)

Juvenal Solano (S. Sebastián del Monte Mixtec)

but
first
~~coffee~~

some background on inflectional classes and Mixtec

What are inflectional classes?

“Inflectional class is the best known type of morphological feature. It partitions the vocabulary items in a given language according to the way in which they realise morphosemantic or morphosyntactic feature specifications.” (Kibort 2008)

“An inflectional class is a set of lexemes whose members each select the same set of inflectional realizations.” (Aronoff 1994:64)

“The constitution of inflectional classes is based on the uniformity and distinctiveness of paradigms, just as every classification is based on the common and distinct properties of the elements to be classified.” (Wurzel 1989:63)

What are inflectional classes?

“Inflectional class is the best known type of **morphological feature**. It partitions the vocabulary items in a given language according to the way in which they realise **morphosemantic or morphosyntactic** feature specifications.” (Kibort 2008)

“An inflectional class is a set of **lexemes** whose members each select the same set of **inflectional realizations**.” (Aronoff 1994:64)

“The constitution of inflectional classes is based on the uniformity and distinctiveness of **paradigms**, just as every classification is based on the common and distinct properties of the elements to be classified.” (Wurzel 1989:63)

Typology and diachrony

- focus on Indo-European, especially in diachrony
- Canonical typology
- whole-language typology
- quantification such as 'complexity measures' and simulations

The canonical view

1	Forms differ as consistently as possible across inflectional classes, cell by cell.	Principle I: Canonical inflectional classes are fully comparable and are distinguished as clearly as is possible.
2	Canonical inflectional classes realize the same morphosyntactic or morphosemantic distinctions (they are of the same structure).	
3	Within a canonical inflectional class each member behaves identically	
4	Within a canonical inflectional class each paradigm cell is of equal status.	
5	The larger the number of members of an inflectional class (up to an equal 'share' of the available items) the more canonical that class.	Principle II: The distribution of lexical items over canonical inflectional classes is synchronically unmotivated
6	The distribution of lexical items over inflectional classes is not phonologically motivated.	
7	The distribution of lexical items over inflectional classes is not syntactically motivated.	
8	The distribution of lexical items over inflectional classes is not motivated by Part of Speech.	
9	The distribution of lexical items over inflectional classes is not motivated by pragmatics (including information structure).	

The Mixtec languages

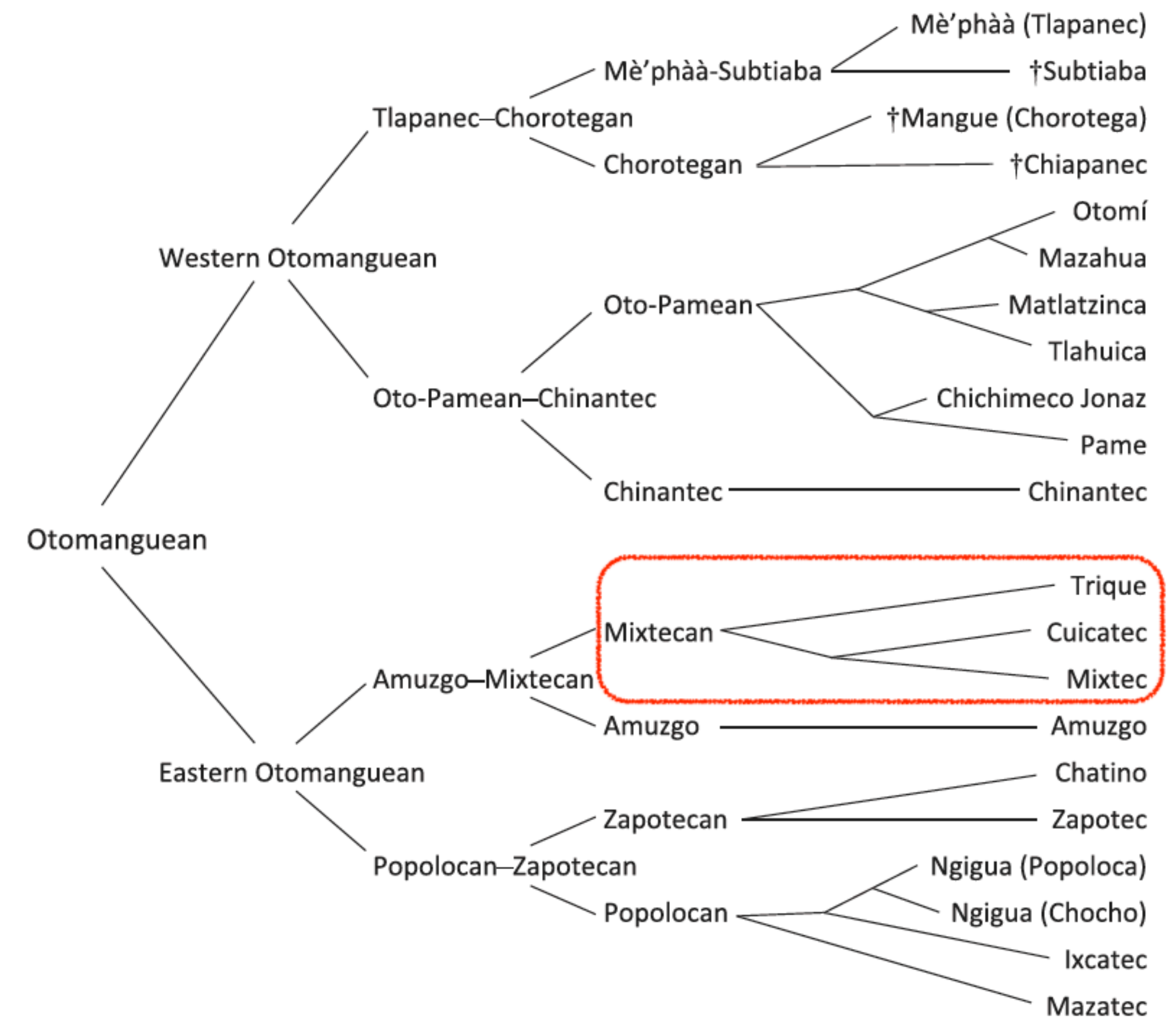
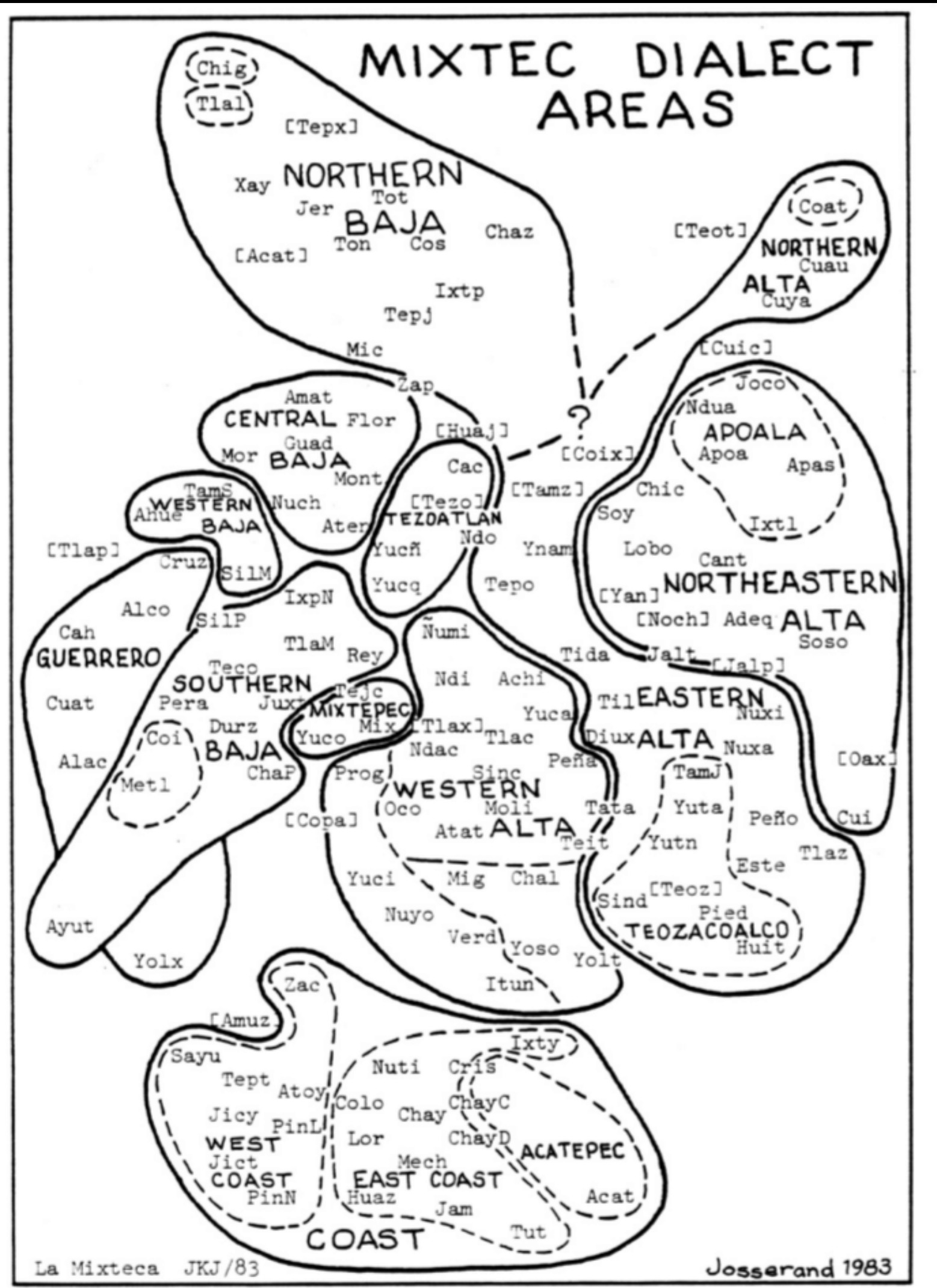
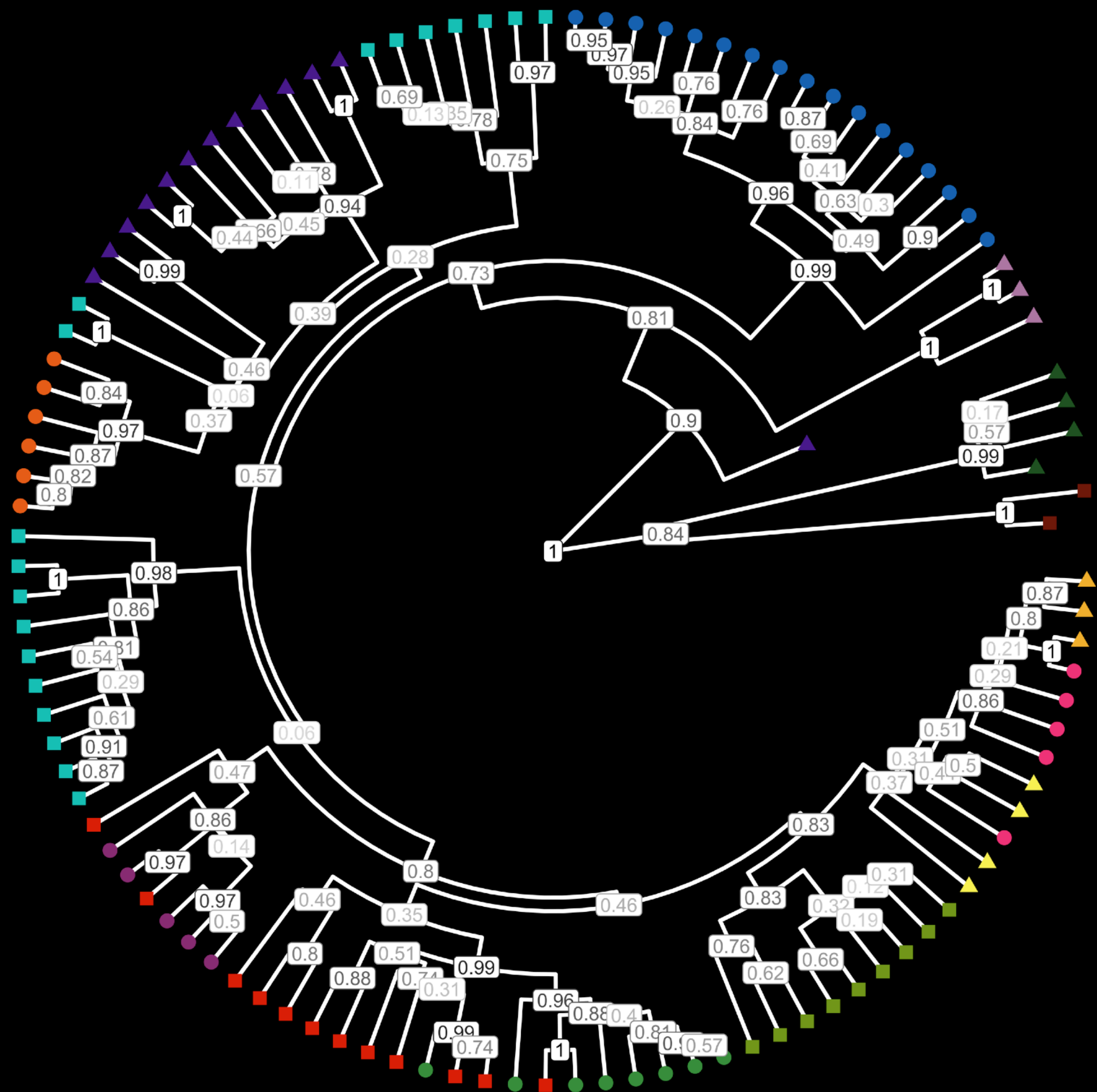


FIGURE 1 Otomanguean classification (based on Kaufman, 1988)

Classification



Jossierand. 1983. *Mixtec Dialect History*. Dissertation



Auderset et al. to appear. *Journal of Language Evolution*

Some facts about tone and syllable structure

- all Mixtec languages are tonal
- the tone-bearing unit is the mora
- final vowels show contrastive vowel length and nasalization
- lexical items in Mixtec are always at least bimoraic
- there are limited consonant clusters and no final consonants



POT Form	Meaning
i ³ ĩ ⁵	peel off
ʃa ³ a ³	start
ka ¹ ʃĩ ²	choose
ka ³ ? ⁿ ta ³	cut sth.
s- ⁿ ta ³ ʃĩ ³	make wet
s- ⁿ ta ⁵ -ko ¹ o ³	make sb. stand up again

Mixtec aspect-mood marking

Irrealis	Realis	
Potential	Incompletive	Completive

- forms are expressed by (a combination of):
 - prefixes
 - stem changes in the initial consonant (and optionally the following vowel)
 - tone changes
 - suppletion

Mixtec aspect-mood marking

`give as a gift, donate'

Language	POT_IPA	INCPL_IPA	COMPL_IPA
MagdalenaPenascoMixtec	sa ³ ha ¹	sa ⁵ ha ¹	
PiedraAzulMixtec	ʃa ³ ʃa ¹⁵	ʃa ⁵ ʃa ¹⁵	ʃa ¹³ ʃa ¹⁵
SanMartinDuraznosMixtec	ɬa ³ ɬã ³	ɬa ⁵ ɬã ⁵	i ¹ +ɬa ³ ɬã ³
TlahuapaMixtec	ʃa ³ ʃa ³	ʃa ⁵ ʃa ⁵	ʃa ¹ ʃa ⁵
YoloxochitlMixtec	ʃa ³ ʃa ⁵	ʃa ⁵ ʃa ⁵	ʃa ¹³ ʃa ⁵
SanAndresYutatioMixtec	ð̥a ³ sa ⁵	ð̥a ³ sa ⁵	ni ¹ +ð̥a ³ sa ⁵

Mixtec aspect-mood marking

`break something'

Language	POT_IPA	INCPL_IPA	COMPL_IPA
MagdalenaPenascoMixtec	ka ³ ?nu ¹	ha ⁵ ?nu ¹	
PiedraAzulMixtec	ka ¹ ?nu ¹	ʃa ⁵ ?nu ¹	ni ¹ +ʃa ¹ ?nu ¹
SanAndresYutatioMixtec	ka ³ ?a ¹ no ¹	ka ⁵ ?a ¹ no ¹	ni ¹ +ka ³ ?a ¹ no ¹
SanMartinDuraznosMixtec	ka ¹ ?nu ¹	ɬa ⁵ ?nu ¹	i ¹ +ɬa ¹ ?nu ¹
TlahuapaMixtec	ka ³ ?nu ¹	ʃa ⁵ ?nu ¹	ʃa ¹ ?nu ²
XochapaMixtec	ka ¹ ?nu ¹	ʃa ⁵ ?nu ¹	ⁿ ti ¹ +ʃa ¹ ?nu ¹
YoloxochitlMixtec	ka? ¹ nu ¹	ʃa? ⁵ nu ¹	ʃa? ¹ nu ¹

Mixtec aspect-mood marking

‘play’

Language	POT_IPA	INCPL_IPA	COMPL_IPA
SanAndresYutatioMixtec	ka ³ +ǝi ³ ki ⁵	sa ⁵ +ǝi ³ ki ⁵	ni ¹ +sa ¹ +ǝi ³ ki ⁵
SanJuanColoradoMixtec	ha ¹ +si ³ ki ³	ha ⁵ +si ⁵ ki ⁵	ha ¹ +si ³ ki ³
SanMartinDuraznosMixtec	ku ³ +si ⁵ ki ⁵	si ⁵ +si ⁵ ki ⁵	i ¹ +si ³ +si ⁵ ki ⁵
TlahuapaMixtec	ku ³ +si ⁵ ki ⁵	si ⁵ +si ⁵ ki ⁵	si ¹ +si ¹⁵ ki ⁵
YoloxochitlMixtec	ku ³ +si ⁵ ki ²⁵	xa ⁵ +si ⁵ ki ²⁵	xa ¹³ +si ⁵ ki ²⁵

Inflectional classes in Otomanguean

- popular topic in the past 10 years
- more documentation => more descriptions available
- core interests revolve around
 - complexity
 - tone marking and the interaction of tones and segments
 - canonicity (Canonical Typology)

Inflectional classes in Otomanguean

- popular topic in the past 10 years
- more documentation => more descriptions available
- core interests revolve around
 - complexity
 - tone marking and the interaction of tones and segments
 - canonicity (Canonical Typology)



- 20 Otomanguean languages
 - 3 Mixtec languages

available online at: <https://oto-manguean.surrey.ac.uk/>

The Otomanguean Inflectional Class Database

Citation form	Spanish / English	Verbal paradigm exhibits variation in:	Inflectional class assignment based on:	
cha ⁴ ?vi ⁴	pagar pay	✗ Affixes ✓ Tone ✗ Stem	✓ Affixes ? Tone ✓ Stem	Less ^
Paradigm		Inflectional class		
	IRR	CPL	INCPL	
Inflected form	cha ³ ?vi ³	cha ² ?vi ²	cha ⁴ ?vi ⁴	
Prefix	—	—	—	
Tone	33	22	44	
		Additional information		
		Stem phonology: Bimoraic Transitivity: tr		

Xochapa Mixtec

Yoloxochitl Mixtec

Citation form	Spanish / English	Verbal paradigm exhibits variation in:	Inflectional class assignment based on:	
su ¹ kun ¹	caerse fall	✓ Affixes ✓ Tone ✗ Stem	✗ Affixes ✗ Tone ✓ Stem	Less ^

Paradigm						
	IRR	NEG.IRR	CPL-1	CPL-2	INCPL	STATIVE
Inflected form	su ¹ kun ¹	su ¹⁴ kun ¹	ni ¹ -su ¹ kun ¹	su ¹ kun ¹	su ⁴ kun ¹	—
Tone	1-1	14-1	1-1	1-1	4-1	

Inflectional class

Stem alternation: Invariable

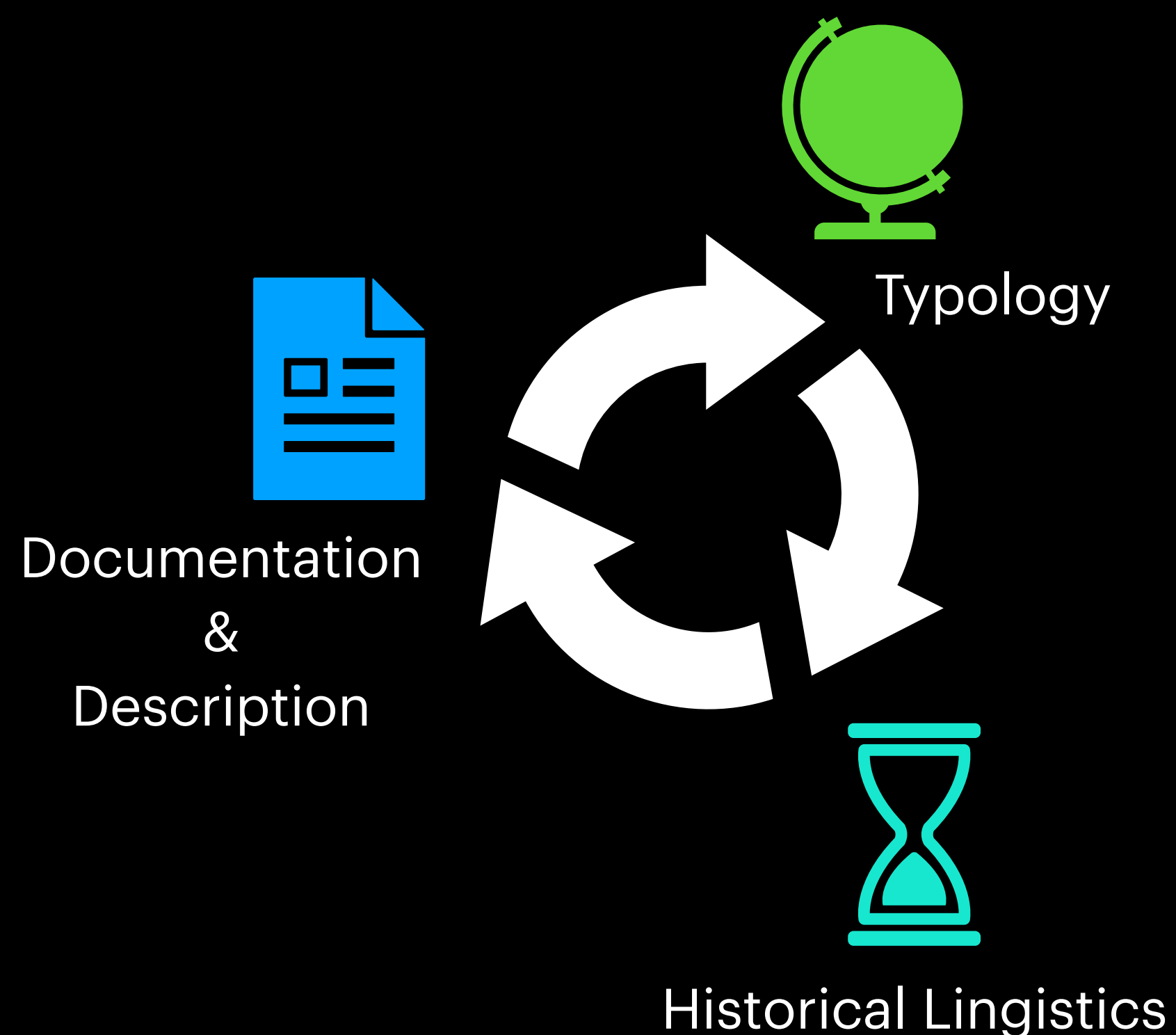
Additional information

Stem phonology: Bimoraic
Transitivity: intr

What is part of the paradigm?

- person
 - pronouns (not indexes)
 - BUT: they interact segmentally and tonally with the verb base
- negation
 - seems to be relatively regular
 - BUT: comparative study is lacking

Research questions and aims



Research questions:

- How did Mixtec aspect-mood marking evolve over time?
- Can we trace the origins of segmental and tonal markers to proto-Mixtec?
- Are `inflectional classes` stable over time?
- if so, can they tell us anything about subgrouping?

Data base design

Design principles

adapted from and inspired by AUTOTYP and FAIR design principles
combined with requisites for historical linguistics/diachronic typology

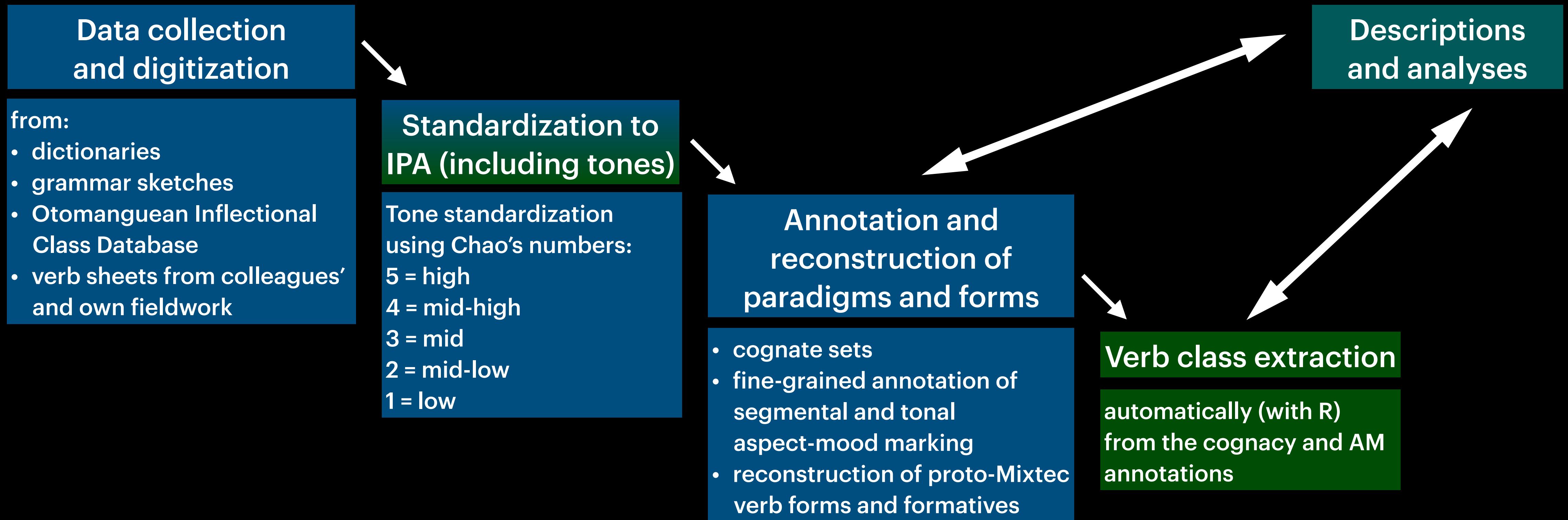
- modularity and connectivity
- **AUTOTYP**ology
- definition vs. data files
- late aggregation
- exemplar-based method

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

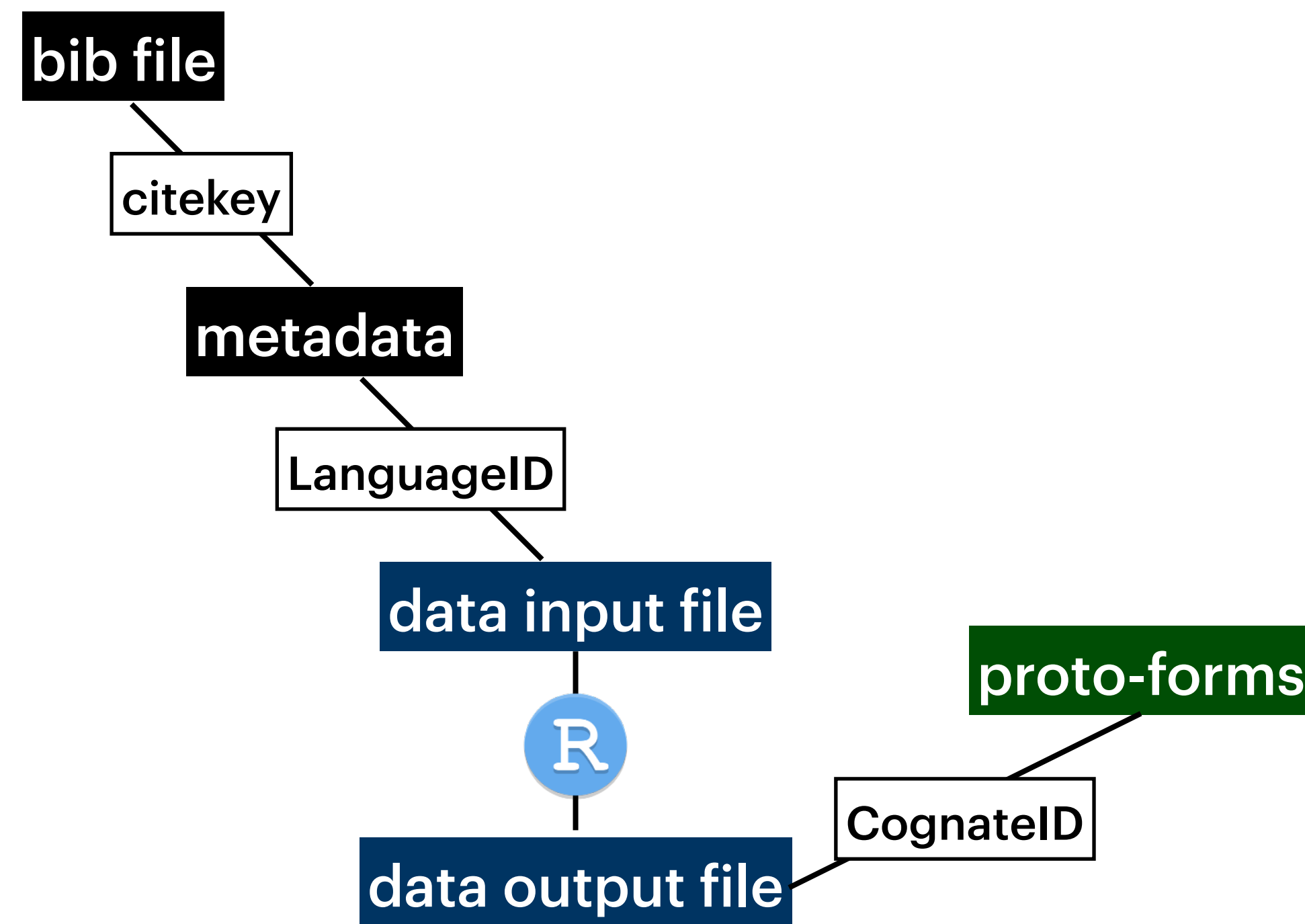
diachrony

- cognacy
- reconstruction

Workflow



Architecture



Current state of the database

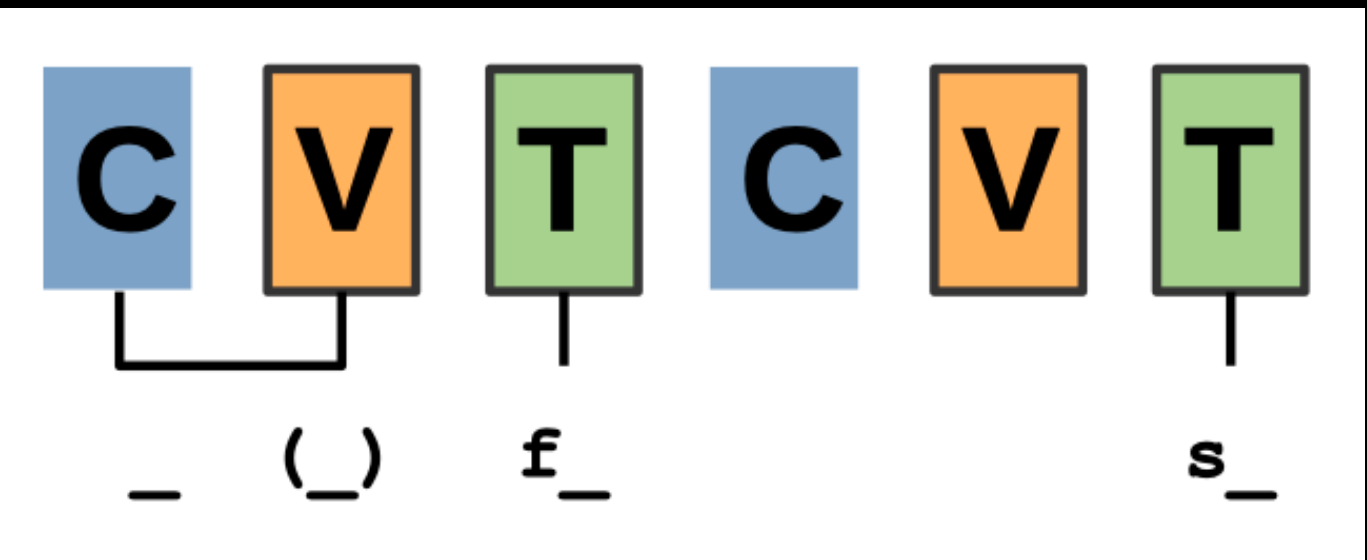
- > 21'000 data points (verb forms)
- 6419 paradigms
 - 3679 annotated paradigms
 - spread over 578 cognate sets
- 15 varieties, but 4 lack important information:
 - no tone marking and only citation form: Chayuco and Yosondua
 - low coverage: S. Seb. Monte (40), Yucunani (83)

Annotation Schema - Input

anchor points/‘base’ forms

	Irrealis	Imperfective	Perfective
	tonal anchor	segmental anchor	
segments			
tones			

template for form annotation

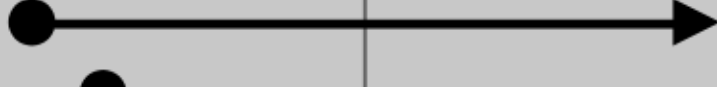
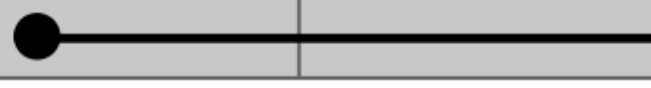


POT_IPA	INCPL_IPA	COMPL_IPA	Gloss
ɕa³ɕã³	ɕa⁵ɕã⁵	i¹+ɕa³ɕã³	give as a gift
ka¹ʔnu¹	ɕa⁵ʔnu¹	i¹+ɕa¹ʔnu¹	cut sth.
ku³+si⁵ki⁵	si⁵+si⁵ki⁵	i¹+si³+si⁵ki⁵	play

Annotation Schema - Cognacy

Language	Gloss	COG_ID	POT_IPA	INCPL_IPA	COMPL_IPA
MagdalenaPenascoMix	quemarse; arder; marchitarse	57	ka ³ ʒu ¹	ka ⁵ ʒu ³	
SanAndresYutatioMixt	arder, quemarse; originar; empeorar	57	ke ³ i ¹	ke ⁵ i ¹	ni ¹ +ke ³ i ¹
SanJuanColoradoMixte	quemarse	57	ka ¹ ju ¹	ka ³ ju ⁵	ka ¹ ju ¹
SanMartinDuraznosMix	quemar	57	ka ¹ i ¹	ka ³ i ¹	i ¹ +ka ¹ i ¹
AlacatlatzalaMixtec	quemar	58	ka ³ ?mi ³	ʃa ⁵ ?mi ³	ʃa ¹ ?mi ³
PiedraAzulMixtec	fumar	58	ka ³ ?mi ³	ʃa ⁵ ?mi ³	ʃa ¹ ?mi ³
PiedraAzulMixtec	quemar, fumar	58	ka ³ mi ³	ka ⁵ mi ³	ka ¹ mi ³
SanJuanColoradoMixte	quemar	58	ka [?] ³ mi ³	tʃa [?] ⁵ mi ⁵	tʃa [?] ³ mi ³
SanMartinDuraznosMix	fumar	58	ka ³ ?mi ³	ʃa ⁵ ?mi ³	i ¹ +ʃa ³ ?mi ³
TlahuapaMixtec	quemar, fumar	58	ka ³ ?mi ³	ʃa ⁵ ?mi ³	ʃa ¹ ?mi ²
XochapaMixtec	quemar	58	ka ³ ?mi ³	ʃa ⁵ ?mi ⁵	ʃa ² ?mi ²
YoloxochitlMixtec	quemar	58	ka [?] ³ mi ³	ʃa [?] ⁵ mi ³	ʃa [?] ¹³ mi ³

Annotation Schema - Input

	Irrealis	Imperfective	Perfective
	tonal anchor	segmental anchor	
segments			
tones	 		

POT_IPA	INCPL_IPA	COMPL_IPA	Gloss
ɬa³ɬã³	ɬa⁵ɬã⁵	i¹+ɬa³ɬã³	give as a gift
70	fH-70	l+70	
ka¹ʔnu¹	ɬa⁵ʔnu¹	i¹+ɬa¹ʔnu¹	cut sth.
K-51	fH-51	l+51	
ku³+si⁵ki⁵	si⁵+si⁵ki⁵	i¹+si³+si⁵ki⁵	play
KU+207	Xl+207	l+Xl+207	

Annotation schema - Output

- sorry, real output too big to fit on slide

Language	Meaning	Forms in IPA	Cognate IDs	Affixes	Stem Alternation	Tone Alternations

Quick Demo

Applications

Marking of AM categories

	Form	ToneAlt	n	rel.freq
	<chr>	<chr>	<int>	<dbl>
1	COMPL	L	245	0.4
2	INCPL	H	<u>1887</u>	0.81
3	POT	M	53	0.43

	Form	StemAlt	n	rel.freq
	<chr>	<chr>	<int>	<dbl>
1	POT	NA	<u>5264</u>	0.87
2	POT	KU-	426	0.07
3	POT	K-	349	0.06
4	POT	U-	17	0
5	INCPL	NA	<u>5590</u>	0.99
6	INCPL	K-	37	0.01
7	INCPL	KU-	17	0
8	INCPL	U-	2	0
9	COMPL	NA	<u>2746</u>	0.99
10	COMPL	K-	20	0.01
11	COMPL	KU-	5	0
12	COMPL	U-	1	0

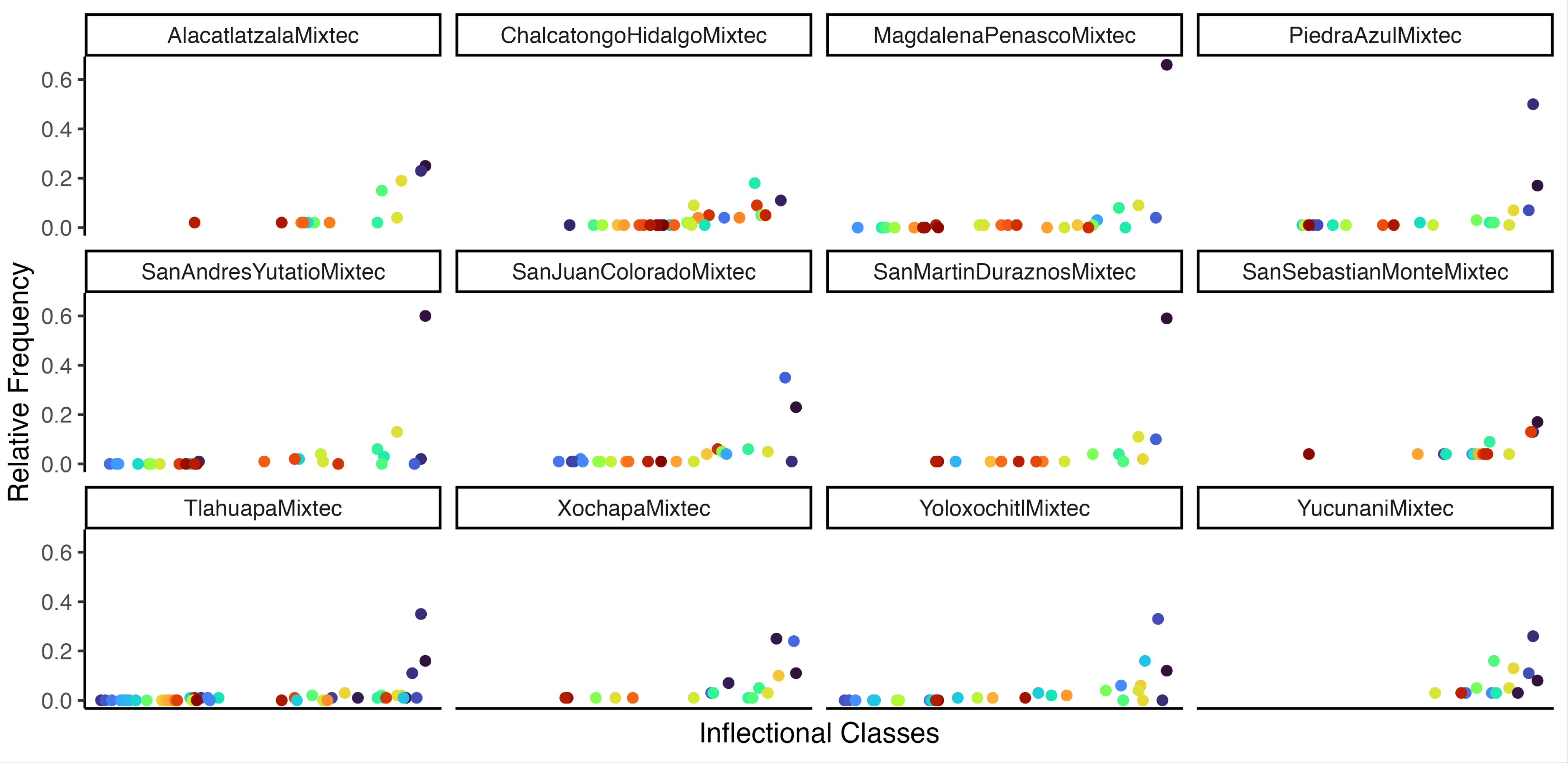
Description of varieties

- comprehensive overviews of classes and their distribution in a variety

Language_ID	Class_Description	n	rel.freq
PiedraAzulMixtec	H for INCLP and L for COMPL	80	0.30
PiedraAzulMixtec	Prefix KU	33	0.12
PiedraAzulMixtec	H for INCLP	28	0.10
PiedraAzulMixtec	Prefix SA	25	0.09
PiedraAzulMixtec	KU- in POT, H for INCLP and L for COMPL	14	0.05
PiedraAzulMixtec	H for INCLP and R for COMPL	12	0.04
PiedraAzulMixtec	Prefix KU, H for INCLP	7	0.03
PiedraAzulMixtec	Prefix NTV	9	0.03
PiedraAzulMixtec	K- in POT, H for INCLP and R for COMPL	5	0.02
PiedraAzulMixtec	Prefix CH	5	0.02

Language_ID	Class_Description	n	rel.freq
SanAndresYutatioMixtec	H for INCLP	156	0.28
SanAndresYutatioMixtec	Prefix KU	80	0.15
SanAndresYutatioMixtec	Prefix SA	53	0.10
SanAndresYutatioMixtec	KU- in POT, H for INCLP	33	0.06
SanAndresYutatioMixtec	Prefix NTV	32	0.06
SanAndresYutatioMixtec	Prefix KU, H for INCLP	20	0.04
SanAndresYutatioMixtec	Prefix SA, H for INCLP and H for COMPL and H for POT	22	0.04
SanAndresYutatioMixtec	K- in POT, H for INCLP	15	0.03
SanAndresYutatioMixtec	Prefix CH	17	0.03
SanAndresYutatioMixtec	KU- in all	9	0.02

Comparison between varieties



Comparison in cognate sets

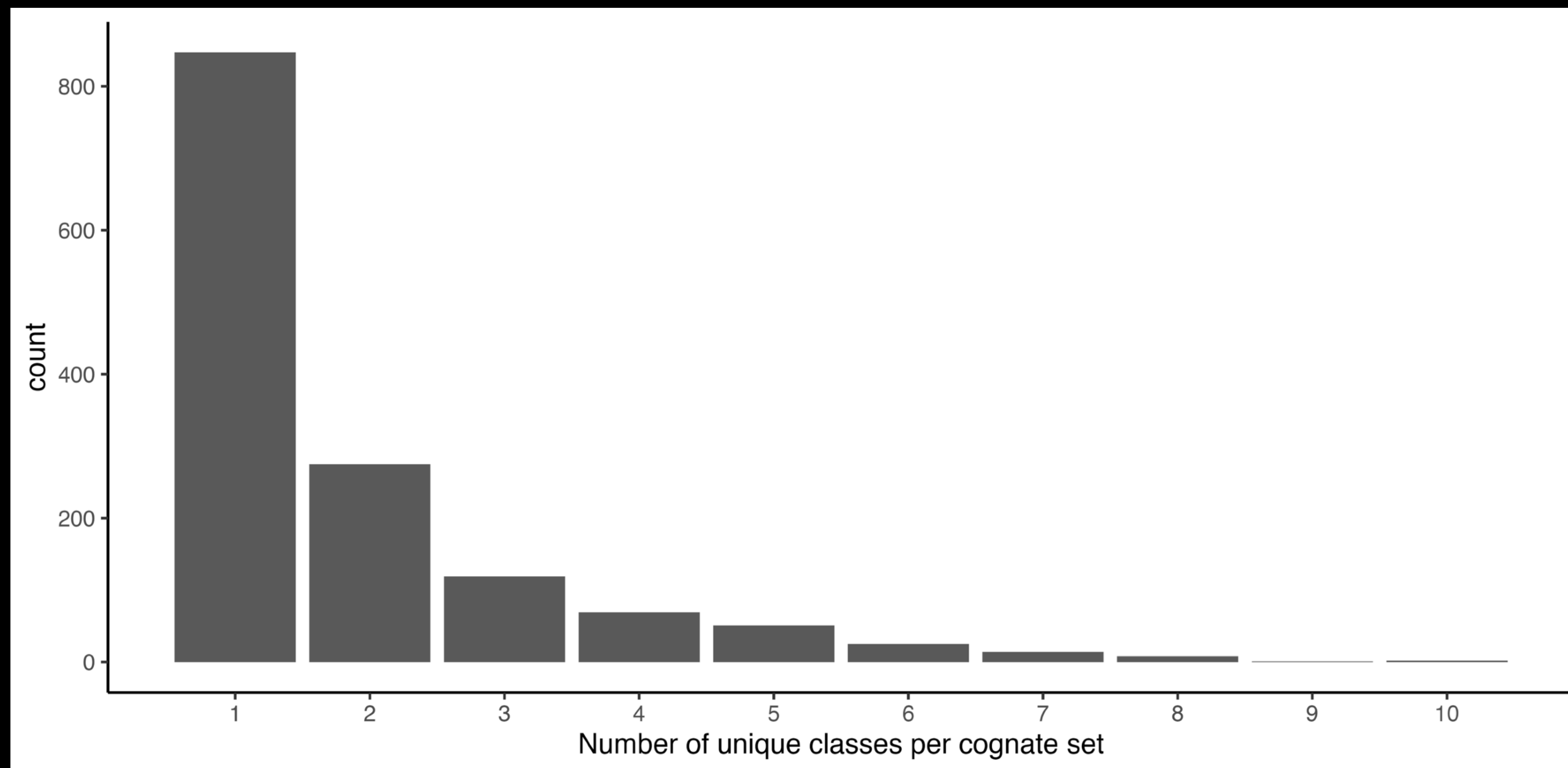
Language_ID ↕	Spanish ↕	Paradigm_ID_New ↕	POT_IPA ↕	INCPL_IPA ↕	COMPL_IPA ↕	CogID_POT ↕	CogID_INCPL ↕	CogID_COMPL ↕
ChalcatongoHidalgoMixtec	quebrar	P0510	ka ⁵ ʔnu ³	xa ⁵ ʔnu ³	xa ⁵ ʔnu ³	510	510	510
MagdalenaPenascoMixtec	quebrar, cortar (contra la veta)	P0510	ka ³ ʔnu ¹	ha ⁵ ʔnu ¹	NA	510	510	NA
PiedraAzulMixtec	quebrarlo	P0510	ka ¹ ʔnu ¹	ʃa ⁵ ʔnu ¹	ni ¹ +ʃa ¹ ʔnu ¹	510	510	510
SanAndresYutatioMixtec	quebrar, cortar (sin herramienta); doblar	P0510	ka ³ ʔa ¹ no ¹	ka ⁵ ʔa ¹ no ¹	ni ¹ +ka ³ ʔa ¹ no ¹	510	510	510
SanMartinDuraznosMixtec	quebrar (rama), doblar (papel)	P0510	ka ¹ ʔnu ¹	ɬa ⁵ ʔnu ¹	i ¹ +ɬa ¹ ʔnu ¹	510	510	510
TlahuapaMixtec	quebrarlo	P0510	ka ³ ʔnu ¹	ʃa ⁵ ʔnu ¹	ʃa ¹ ʔnu ²	510	510	510
TlahuapaMixtec	quebrarlo	P0510	ka ³ ʔnu ¹	ka ⁵ ʔnu ¹	ka ¹ ʔnu ²	510	510	510
XochapaMixtec	quebrar	P0510	ka ¹ ʔnu ¹	ʃa ⁵ ʔnu ¹	ⁿti ¹ +ʃa ¹ ʔnu ¹	510	510	510
YoloxochitlMixtec	partir en dos	P0510	kaʔ ¹ nu ¹	ʃaʔ ⁵ nu ¹	ʃaʔ ¹ nu ¹	510	510	510
MagdalenaPenascoMixtec	impedir	P0510	ka ³ ʔnu ¹ +ⁿte ³ e ¹	ha ⁵ ʔnu ¹ +ⁿte ³ e ¹	NA	NA	NA	NA

Affixes_POT ↕	Affixes_INCPL ↕	Affixes_COMPL ↕	StemAlt_POT ↕	StemAlt_INCPL ↕	StemAlt_COMPL ↕	ToneAlt_POT ↕	ToneAlt_INCPL ↕	ToneAlt_COMPL ↕
NA	NA	NA	K-	NA	NA	NA	NA	NA
NA	NA	NA	K-	NA	NA	NA	H	NA
NA	NA	NI	K-	NA	NA	NA	H	L
NA	NA	NI	K-	K-	K-	NA	H	NA
NA	NA	I	K-	NA	NA	NA	H	NA
NA	NA	NA	K-	NA	NA	NA	H	L-B
NA	NA	NA	K-	K-	K-	NA	H	L-B
NA	NA	NDI	K-	NA	NA	NA	H	NA
NA	NA	NA	K-	NA	NA	NA	H	L
NA	NA	NA	K-	NA	NA	NA	H	NA

Comparison in cognate sets

Language_ID ↕	Spanish ↕	English ↗	Paradigm_ID ↕	Paradigm_ID_New ↕	POT_IPA ↕	INCPL_IPA ↕	COMPL_IPA ↕	CogID_POT ↕	CogID_INCPL ↕	CogID_COMPL
ChalcatongoHidalgoMixtec	morir	die	P0290	P0290	ku ³ u ¹	xi ⁵ ʔi ³	xi ⁵ ʔi ³	310	290	290
MagdalenaPenascoMixtec	morir; eclipsarse (el sol o la luna), estar en eclipse	die	P0290	P0290	ku ³ u ¹	hi ⁵ ʔi ³	NA	310	290	NA
PiedraAzulMixtec	morirse	die	P0290	P0290	ku ³ u ¹	ʃi ⁵ ʔi ¹	ni ¹ +ʃi ¹ ʔi ¹	310	290	290
SanJuanColoradoMixtec	morir	die	P0290	P0290	ku ¹ u ¹ ~ku ¹ βi ¹	tsi ³ ʔi ⁵	tsi ⁷ i ¹	310	290	290
SanMartinDuraznosMixtec	morir	die	P0290	P0290	ki ³ βi ³	ɕi ⁵ ʔi ¹	i ¹ +ɕi ¹ ʔi ¹	310	290	290
TlahuapaMixtec	morirse	die	P0290	P0290	ku ³ βi ³	ʃi ⁵ ʔi ¹	ʰti ¹ +ʃi ¹ ʔi ¹	310	290	290
XochapaMixtec	morir	die	P0290	P0290	ku ³ βi ³	ʃi ⁵ ʔi ³	ʰti ¹ +ʃi ¹ ʔi ³	310	290	290
YoloxochitlMixtec	morir	die	P0290	P0290	ku ³ u ²	ʃi ⁵ i ¹	ʃi ⁷ i ¹	310	290	290
SanAndresYutatioMixtec	morir; estar enfermo	die, be ill	P0290	P0290	ku ³ u ¹	ʃi ⁵ ʔi ¹	ni ¹ +ʃi ³ ʔi ¹	310	290	290
AlacatzlatzalaMixtec	morir	NA	P0290	P0290	ki ³ βi ¹	ʃi ⁵ ʔi ¹	ni ¹ +ʃi ¹ ʔi ¹	310	290	290

Comparison across cognate sets



Wrap up

Some dis/advantages

- (very) time consuming - a lot of manual labor
- language(-family) specific



Some dis/advantages

- (very) time consuming - a lot of manual labor
- language(-family) specific
- easily extendable and adaptable
- true bottom-up, instead of shallow across
- integration of diachrony into typology
- integration of description/documentation



Future work and ideas

- finish annotations and reconstructions, i.e. release a first 'finished' version of the database
- annotate tone alternations on non-verb elements
- find a better way to compare and visualize cognate sets and their reflex classes (computationally)
- update/rethink proto-Mixtec tone reconstruction in light of the verb data
- map this on a phylogenetic tree somehow
- etc.

Future work and ideas

- extend to other language families; build an 'inflectional class collective'
- link with PANGLOSS (example sentences, dictionaries)
- add other modules
 - argument structure
 - phonetic variation
 - etc.

-

Future ideas

- link with PANGLOSS (example sentences, dictionaries)
- add other modules
 - argument structure
 - phonetic variation
 - etc.
- phylogenetics

An aerial photograph of a village built on a steep, forested hillside. The houses are clustered together, with some featuring corrugated metal roofs. A prominent white building with a dome is visible in the center. The foreground shows a dark, corrugated metal roof. The background is a dense forest covering the hillside.

Shìntèé ini-ntò!

thank

inside-2PL