Quantitative challenges in the reconstruction of language history

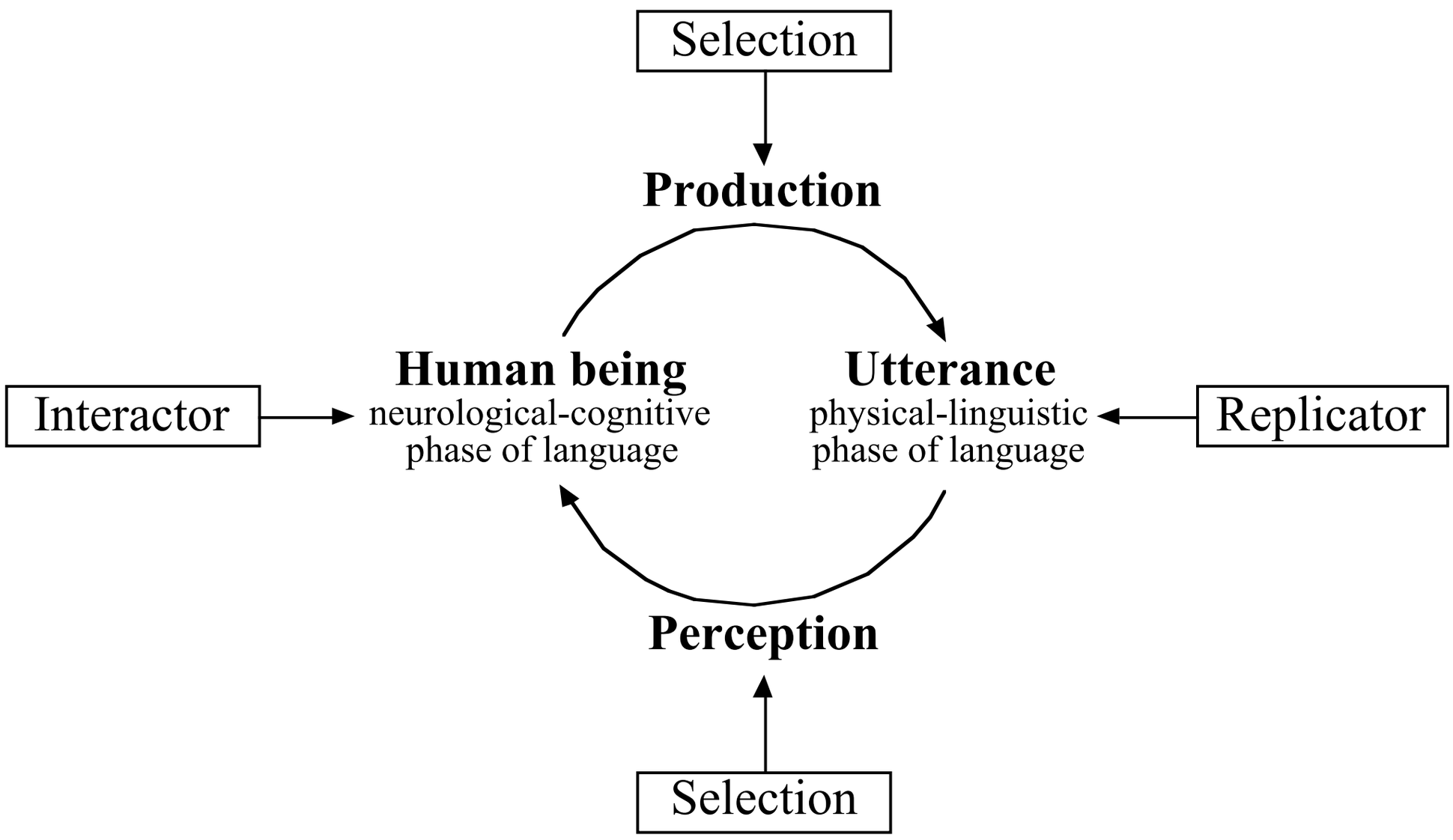
Michael Cysouw
MPI-EVA Leipzig, LMU München

Languages as object of phylogenetic study

- Rapidly changing
  at least potentially
- High within-species diversity
- Complex diachronic dynamics
  vertical descent and horizontal transfer are completely mixed
- Individual utterances (not whole languages) are more like biological organisms

Theory of Utterance Selection

(Croft 2000)



•Dawkins, Richard. 1976. The Selfish Gene. Oxford: Oxford University Press.
•Hull, David L. 1988. Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science. Chicago: University of Chicago Press.
•Croft, William. 2000. Explaining Language Change: An Evolutionary Approach. Harlow: Longman.
•Cysouw, Michael. 2006. review of "Explaining Language Change," Beiträge zur Geschichte der deutschen Sprache und Literatur 128(2): 298-306.

Replicator

Biology	Language
DNA	Utterance
Nucleotide	Phonemes
Gene	Morphemes
?	Words, Syntax
Genetic Information	Intrinsic Meaning
Gene Product	Expressed Meaning

Defining “Meaning”

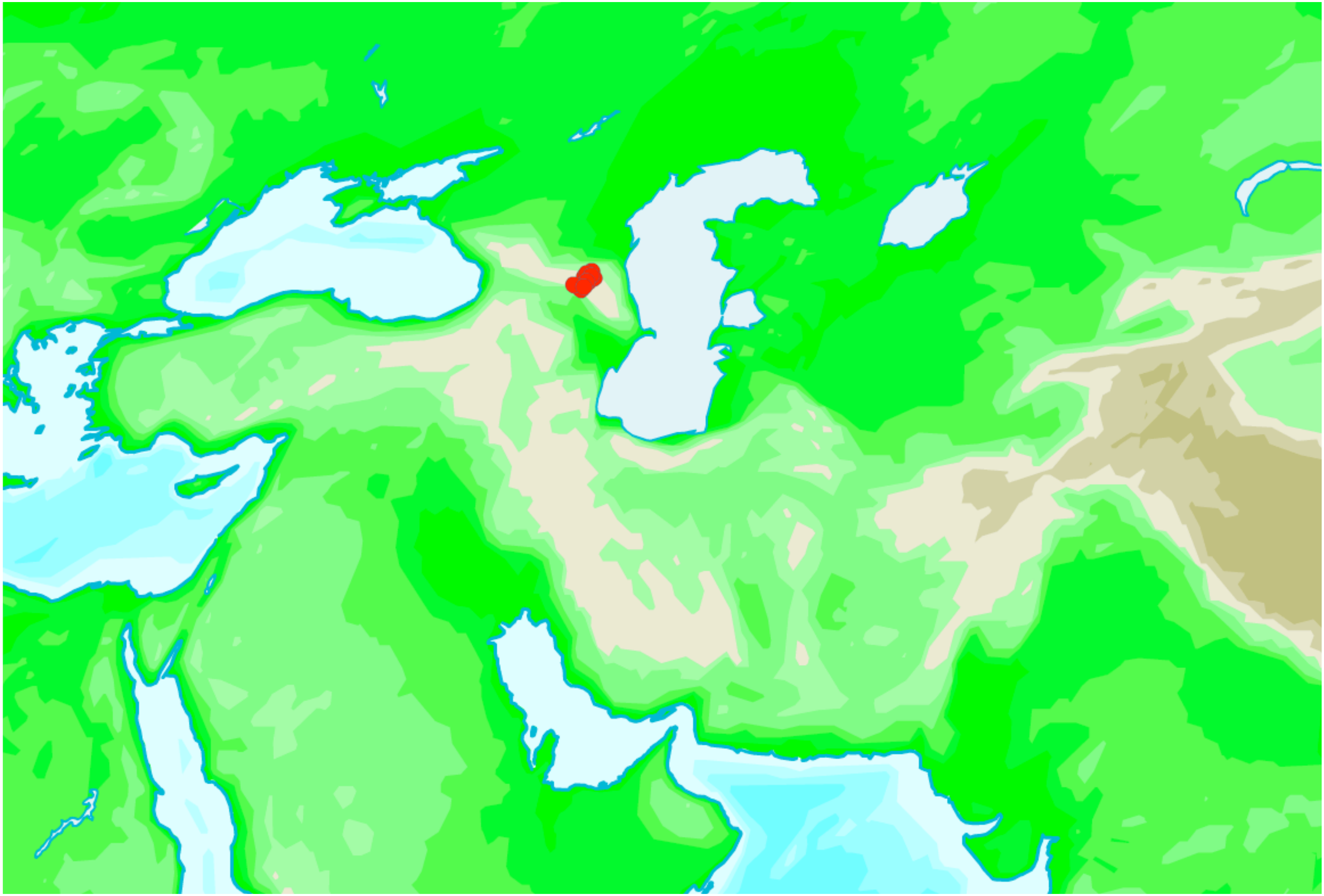
- Intrinsic Meaning is not tangible
  “Double articulation”: phonemes are arbitrarily linked to meaning
- Expressed Meaning can be fruitfully studied
  By looking at the context of individual utterances
- Define Intrinsic Meaning as
  the set of all Expressed Meanings
- So, the ‘meaning’ of a word is like
  a population of its uses

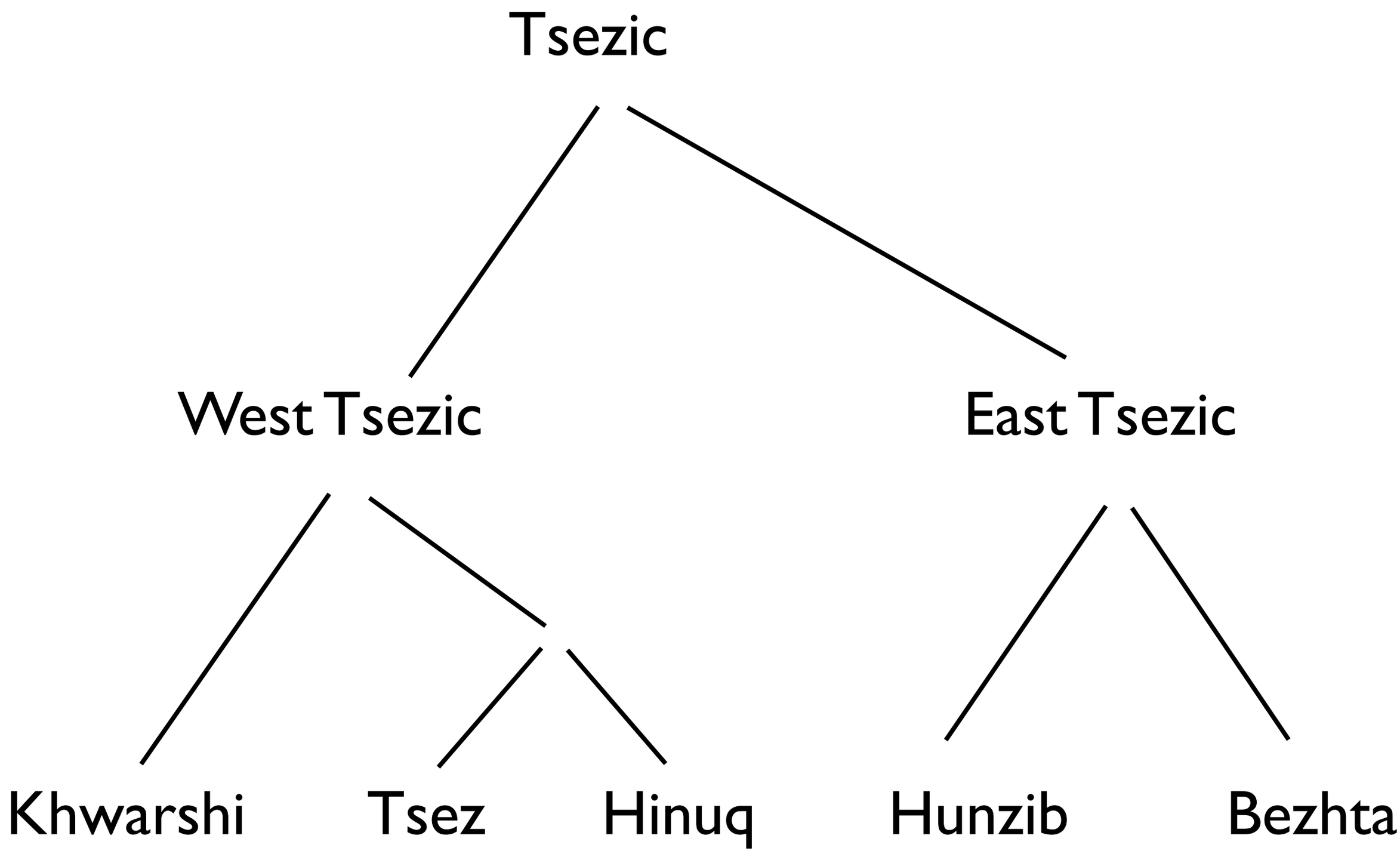
Reconstruction of function

Non-spatial usage of spatial case marking in Tsezic

Michael Cysouw & Diana Forker

Cysouw, Michael & Diana Forker. 2009. Reconstruction of function. Non-spatial usage of spatial case marking in Tsezic. Language 85.3: 588-617.




Tsezic
West Tsezic
East Tsezic
Khwarshi
Tsez
Hinuq
Hunzib
Bezhta

Hinuq

Hunzib

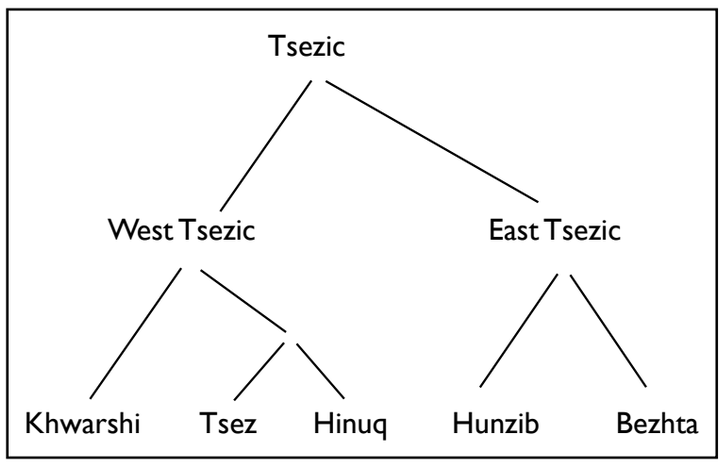
Tsez (Sagadin)

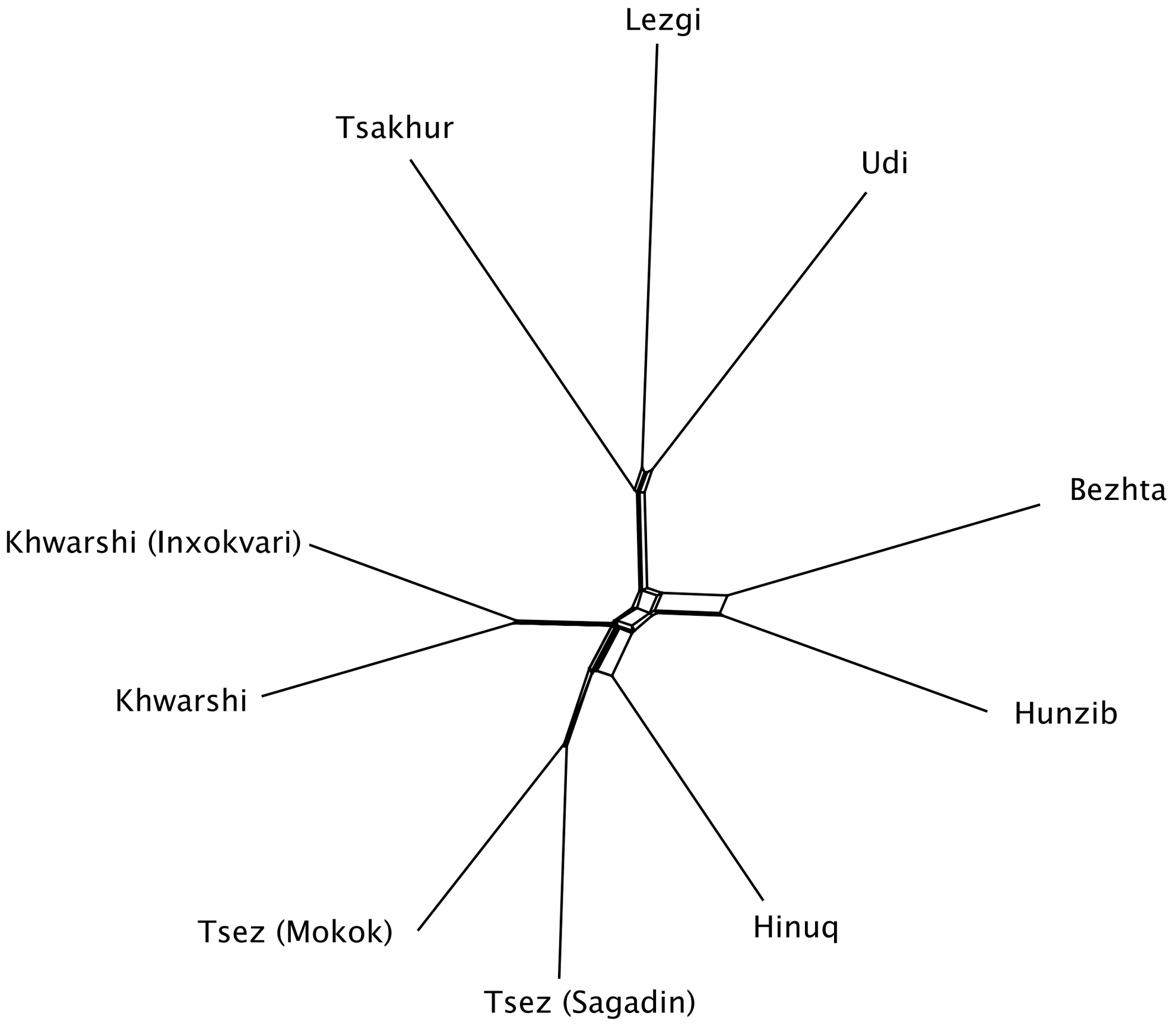
Bezhta

Tsez (Mokok)

Khwarshi (Inxokvari)

Khwarshi




Lezgi
Tsakhur
Udi
Bezhta
Khwarshi (Inxokvari)
Khwarshi
Hunzib
Tsez (Mokok)
Hinuq
Tsez (Sagadin)

Hinuq	CONT	IN	SUB	SPR	AT	APUD	AD	
Essive	-ł	-V/ma	-ƛ	-ƛ’o	-qo	-de	-ho	
Lative	-ł-er	-V/ma-r	-ƛ-er	-ƛ’o-r	-qo-r	-de-r	-ho-r	
Ablative/Genitive	-ł-es	-V/ma-s	-ƛ-es	-ƛ’o-s	-qo-s	-de-s	-ho-s	
Ablative	-ł-ezo	-V/ma-zo	-ƛ-ezo	-ƛ’o-zo	-qo-zo	-de-zo	-ho-zo	
Directive	-ł-edo	-V/ma-do	-ƛ-edo	-ƛ’o-do	-qo-do	-de-do	-ho-do	
Tsez	CONT	IN	SUB	SPR	AT	APUD	AD	
Essive	-ł	-ā	-ƛ	-ƛ’(o)	-q(o)	-de	-x(o)	
Lative	-ł-er	-ā-r	-ƛ-er	-ƛ’o-r	-qo-r	-de-r	-xo-r	
Ablative	-ł-āy	-ā-y	-ƛ-āy	-ƛ’-āy	-q-āy	-d-āy	-x-āy	
Versative	-ł-xor	-ā-ɣor	-ƛ-xor	-ƛ’-āɣor/’ār	-q-āɣor/ār	-d-āɣor/ār	-x-āɣor/ār	
Khwarshi	CONT	IN	SUB	SPR	AT		AD	NEAR
Essive	-ł	-(m)a	-ƛ	-ƛ’o	-qo		-ho	-ɣo
Lative	-ł-ul	-(m)a-l	-ƛ-ul	-ƛ’o-l	-qo-l		-ho-l	-ɣo-l
Ablative	-ł-zi	-(m)a-zi	-ƛ-zi	-ƛ’o-zi	-qo-zi		-ho-zi	-ɣo-zi
Versative	-ł-ɣul	-(m)a-ɣul	-ƛ-ɣul	-ƛ’o-ɣul	-qo-ɣul		-ho-ɣul	-ɣo-ɣul
Translative	-ł-ɣužaz	-(m)a-ɣužaz	-ƛ-ɣužaz	-ƛ’o-ɣužaz	-qo-ɣužaz		-ho-ɣužaz	-ɣo-ɣužaz
Terminative	-ł-q’a	-(m)a-q’a	-ƛ-q’a	-ƛ’o-q’a	-qo-q’a		-ho-q’a	-ɣo-q’a
Bezhta	CONT	IN	SUB	SPR	AT	APUD		NEAR
Essive	-ł	-ʔ	-ƛ	-ƛ’a	-qa	-doy		-ɣa
Lative		-ʔ-il		-ƛ’a-l		-doy-l		-ɣa-l
Ablative/Genitive	-ł-so	-ʔ-is	-ƛ-so	-ƛ’a-s	-qa-s	-doy-s		-ɣa-s
Ablative	-ł-la	-ʔ-la	-ƛ-la	-ƛ’a-la	-qa-la	-doy-la		-ɣa-la
Directive	-ł-daa	-ʔ-daa	-ƛ’a-la	-ƛ’a-daa	-qa-daa	-doy-da		-ɣa-daa
Translative	-ł-laƛ’ā	-ʔ-(la)ƛ’ā	-qa-la	-ƛ’a-laƛ’ā	-qa-laƛ’ā	-doy-		-ɣa-laƛ’ā
Hunzib	CONT	IN	SUB	SPR	AT	APUD		
Essive	-ł	-V	-ƛ	-ƛ’(o)	-g(o)	-dər		
Ablative/Genitive	-ł-sə	-V-s		-ƛ’o-s	-go-s	-dər-sə		
Translative	-ł-ƛ’i	-V-ƛ’		-ƛ’o-ƛ’		-dər-ƛ’i		
Directive	-ł-do	-dα-α				-dər-do		

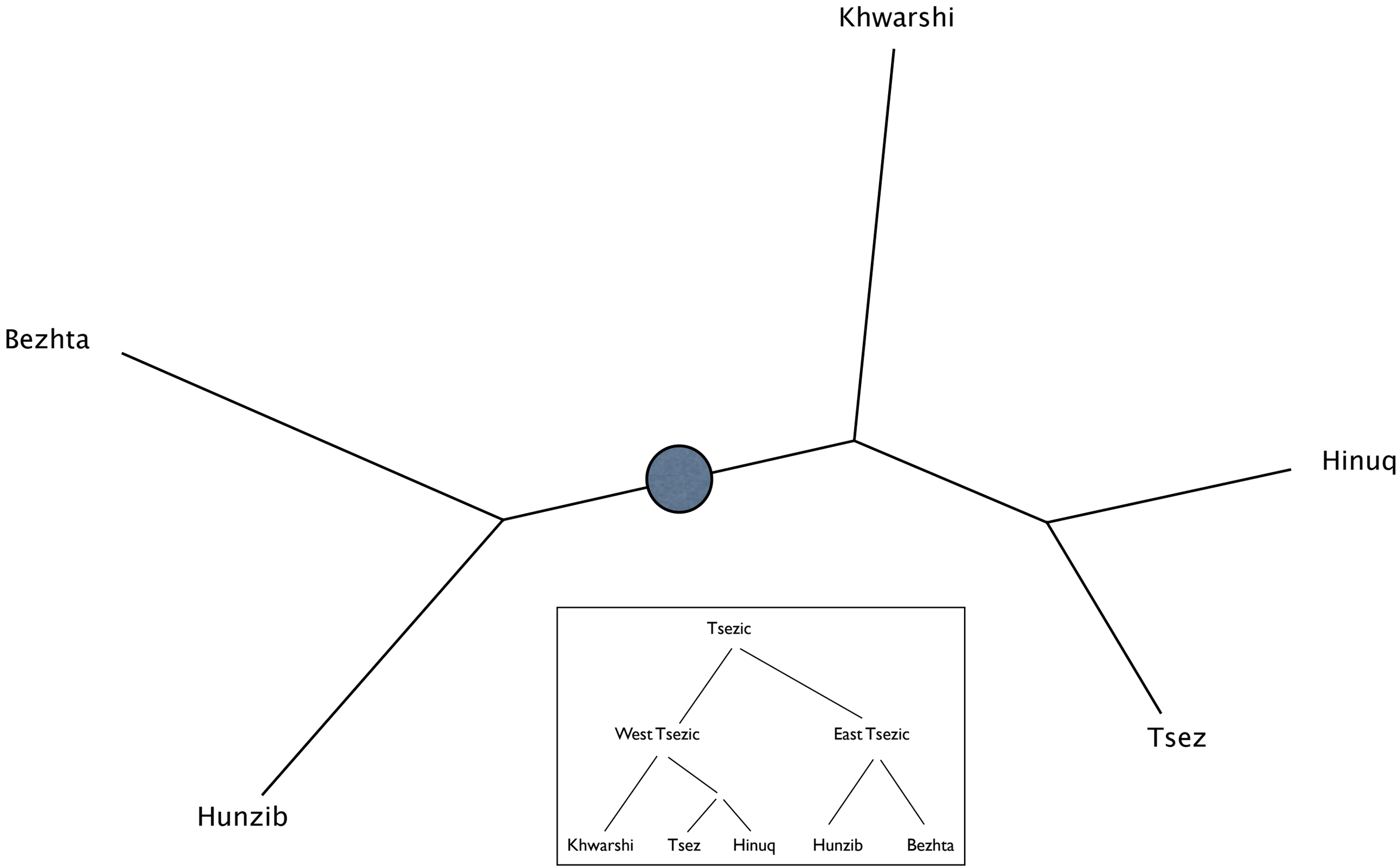
Proto-Tsezic reconstruction	Approximate meaning in spatial domain	Gloss
*-ø	essive (‘location’)	NULL
*-*r/l*	lative (‘orientation towards a goal’)	LAT
*-*s*	ablative (‘movement from, or out of’)	GEN

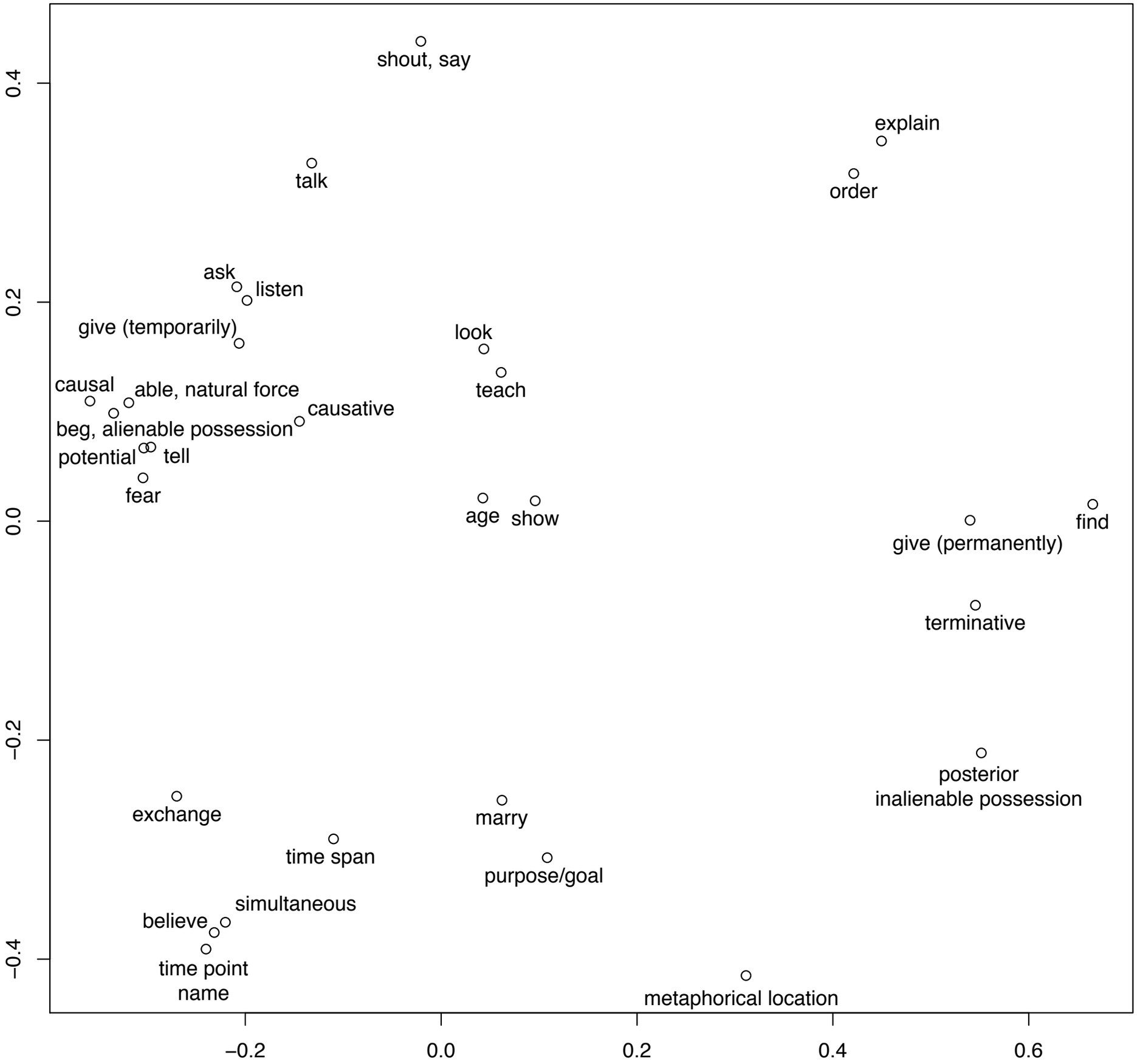
Proto-Tsezic reconstruction	Approximate meaning in spatial domain	Gloss
*-*ɬ*-	contact	CONT
*-*ʔa*-	in, inside	IN
*-*ƛ*-	under	SUB
*-*ƛ'o*-	on top, above	SPR
*-*qo*-	attached to	AT
*-*de*-	near, by (animates)	APUD
*-*ho*-	near, by (inanimates)	AD
*-*ɣo*	near, close	NEAR

Role	Label	Construction
Addressees	TALK	talk *[to somebody]*
	TELL	tell *[to somebody]*
	SHOUT	shout *[at somebody]*
	SAY	say *[to somebody]*
	ASK	ask *[somebody]*
	BEG	beg *[somebody]*
	EXPLAIN	explain *[to somebody]*
	TEACH	teach *[to somebody]*
	ORDER	order *[somebody]*
Human objects	MARRY	marry *[a man]*
	FEAR	fear *[somebody]*
	BELIEVE	believe *[somebody]*
	LISTEN	listen *[to somebody]*
Recipients	GIVE (PERMANENTLY)	give X as a gift *[to somebody]*
	GIVE (TEMPORARILY)	give X to hold *[to somebody]*
	SHOW	show X *[to somebody]*
Possessors	INALIENABLE POSSESSION	*[somebody]* has a daughter
	ALIENABLE POSSESSION	*[somebody]* has money
Subjects	AGE	*[somebody]* is X years old
	NAME	*[somebody]* is called X
	FIND	*[somebody]* finds X
Objects	EXCHANGE	give X away *[for something else]*
	PURPOSE/GOAL	go with the goal to get *[something]*
	LOOK	look *[at something]*
	METAPHORICAL LOCATION	talk *[about something]*
Causees	CAUSATIVE	cause *[somebody]* to do X
Potential agents	POTENTIAL	*[somebody]* can do X
	ABLE	*[somebody]* is able to do X
Reasons	NATURAL FORCE	*[because of something]* X happened
Converbs	SIMULTANEOUS	*[while X happened]* …
	TERMINATIVE	*[until/before X happened]* …
	CAUSAL	*[because X happened]* …
	POSTERIOR	*[after X happened]*…
Time indications	TIME SPAN	something happened *[during period X]*
	TIME POINT	something happened *[at time X]*

Function	Hinuq		Tsez		Khwarshi		Bezhta		Hunzib	
	Loc	Dir	Loc	Dir	Loc	Dir	Loc	Dir	Loc	Dir
able	AT	NULL	AT	NULL	AT	NULL	AT	NULL	AT	NULL
age	AT	NULL	AT	NULL	AT	NULL	NULL	LAT	NULL	GEN
alienable possession	AT	NULL	AT	NULL	AT	NULL	AT	NULL	–	–
ask	AT	NULL	AT	NULL	AT	LAT	AT	NULL	AT	NULL
beg	AT	NULL	AT	NULL	AT	NULL	AT	NULL	–	–
believe	SPR	NULL	SPR	NULL	SPR	NULL	SPR	NULL	SPR	NULL
causal	–	–	AT	NULL	–	–	AT	NULL	AT	NULL
causative	AT	NULL	AT	NULL	AT	NULL	NULL	INSTR	AT	NULL
exchange	SUB	NULL	SUB	NULL	SUB	NULL	SUB	NULL	SUB	NULL
explain	AT	LAT	NULL	LAT	AT	LAT	NULL	LAT	–	–
fear	AT	NULL	AT	NULL	AT	NULL	AT	NULL	SPR	NULL
find	NULL	DAT	NULL	LAT	NULL	LAT	NULL	LAT	–	–
give (permanently)	NULL	DAT	NULL	LAT	NULL	LAT	NULL	LAT	IN	NULL
give (temporarily)	AT	NULL	AT	NULL	NEAR	LAT	AT	NULL	AT	NULL
inalienable possession	NULL	GEN	NULL	GEN	NULL	GEN	NULL	GEN	NULL	GEN
listen	AT	NULL	AT	LAT	AT	NULL	AT	NULL	–	–
look	SPR	LAT	SPR	LAT	AT	LAT	AT	NULL	AT	NULL
marry	AD	NULL	AD	NULL	AD	NULL	NULL	LAT	IN	NULL
metaphorical location	SPR	NULL	CONT	ABL	NULL	GEN	NULL	GEN	SPR	GEN
name	SPR	NULL	SPR	NULL	SPR	NULL	SPR	NULL	–	–
natural force	AT	NULL	AT	NULL	AT	NULL	AT	NULL	AT	NULL
order	AT	LAT	AT	LAT	NULL	LAT	NULL	LAT	–	–
posterior	NULL	GEN	NULL	GEN	NULL	GEN	NULL	GEN	NULL	GEN
potential	AT	NULL	AT	NULL	AT	NULL	–	–	–	–
purpose/goal	SPR	NULL	SPR	LAT	AD	NULL	IN	LAT	SPR	NULL
say	AT	LAT	AT	LAT	AT	LAT	AT	NULL	AT	NULL
shout	AT	LAT	AT	LAT	AT	LAT	AT	NULL	AT	NULL
show	NULL	DAT	AT	NULL	AT	NULL	NULL	LAT	IN	NULL
simultaneous	SPR	NULL	SPR	NULL	SPR	NULL	IN	NULL	IN	NULL
talk	AT	LAT	AT	LAT	AT	NULL	AT	NULL	AT	NULL
teach	AT	NULL	AT	NULL	AT	LAT	NULL	LAT	IN	NULL
tell	AT	NULL	AT	NULL	AT	NULL	AT	NULL	IN	NULL
terminative	SPR	LAT	SPR	LAT	AD	LAT	NULL	LAT	NULL	LAT
time point	SPR	NULL	SPR	NULL	SPR	NULL	SPR	NULL	–	–
time span	CONT	NULL	CONT	NULL	NULL	INSTR	IN	NULL	CONT	NULL

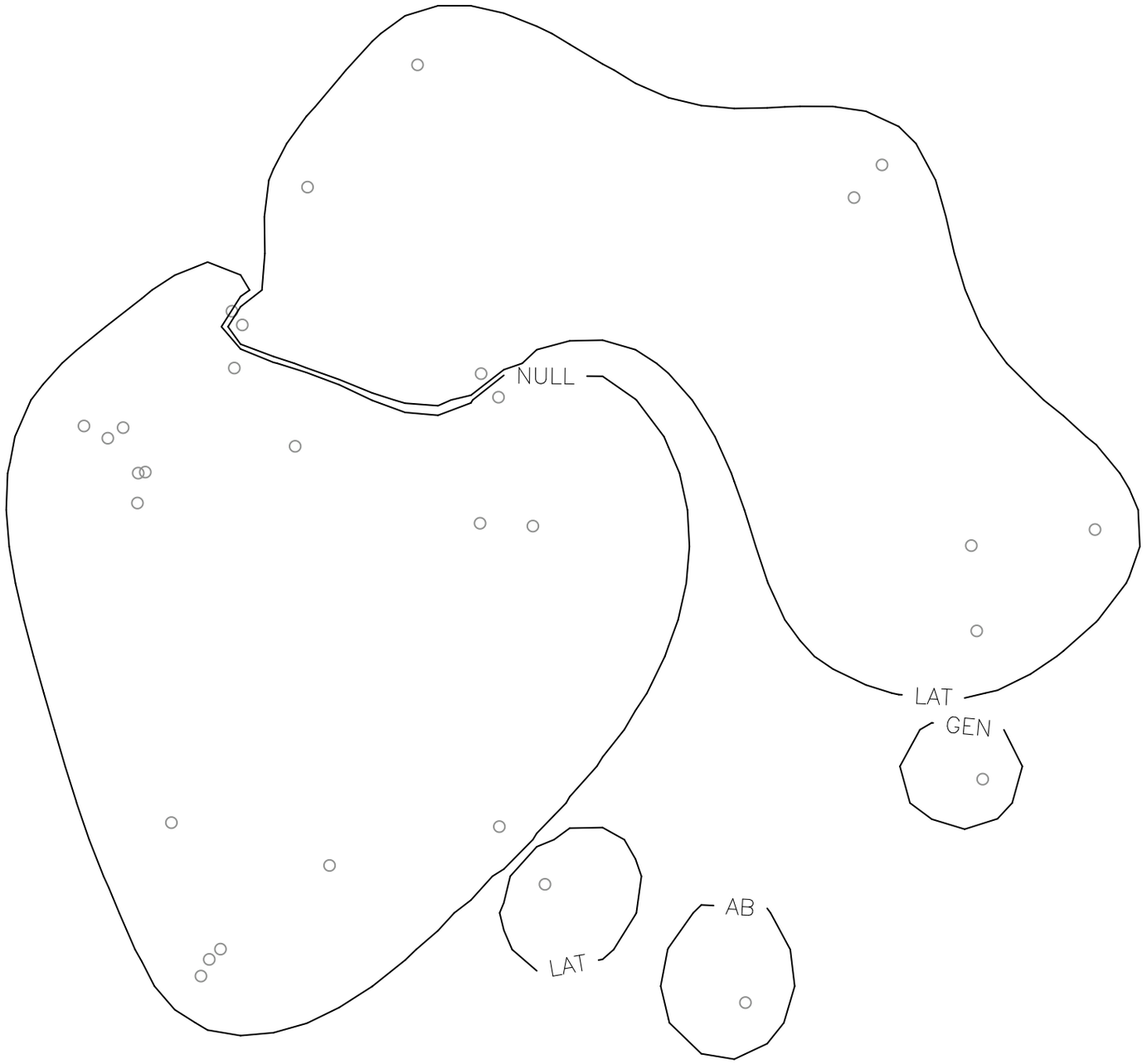
Optimal maximum parsimony tree for case marking




shout, say
talk
explain
order
ask
listen
give (temporarily)
look
teach
causal
able, natural force
causative
beg, alienable possession
potential
tell
fear
age
show
find
give (permanently)
terminative
posterior
inalienable possession
exchange
marry
time span
purpose/goal
simultaneous
believe
time point
name
metaphorical location
0.4
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
0.4
0.6

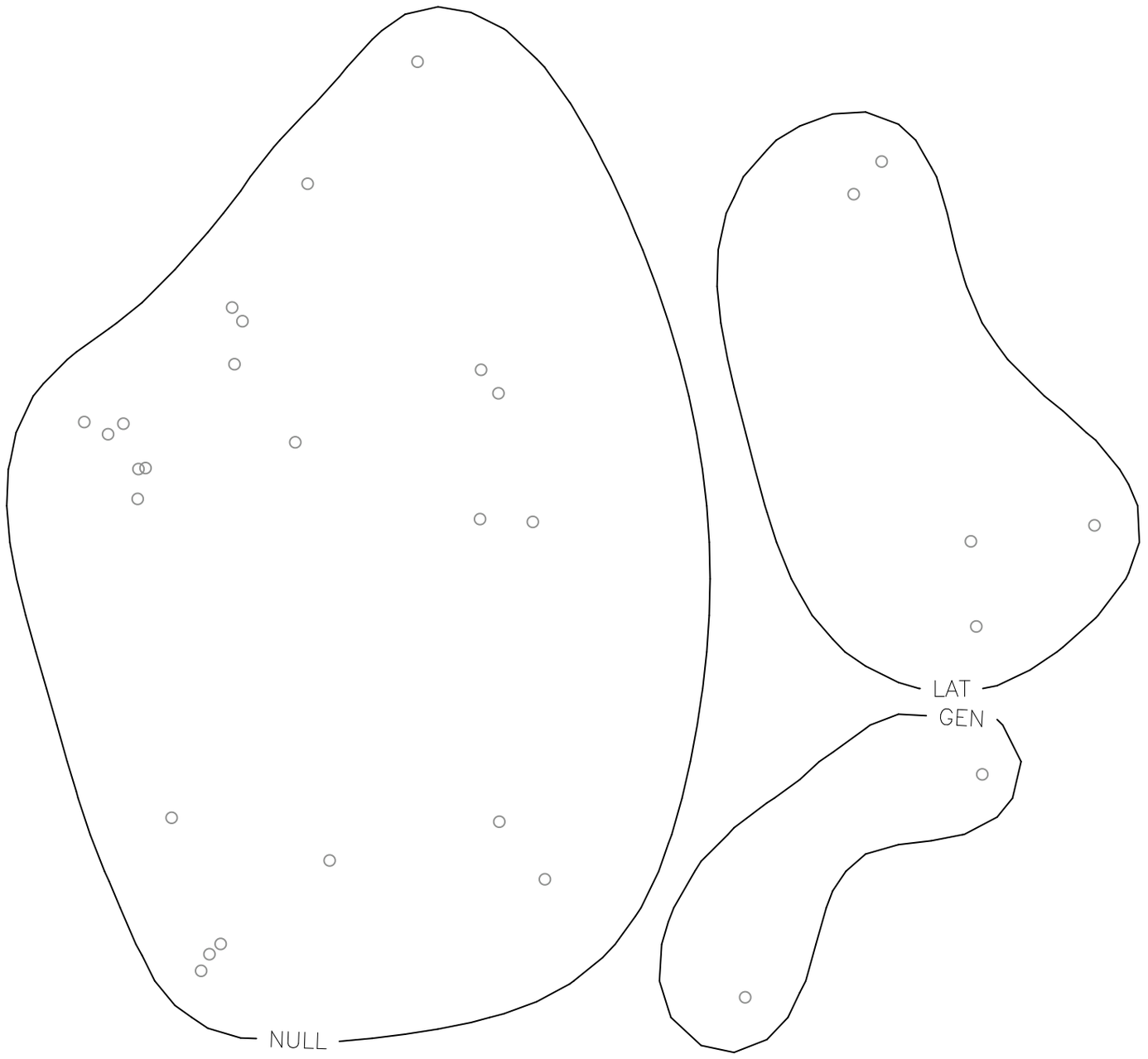
Tsez

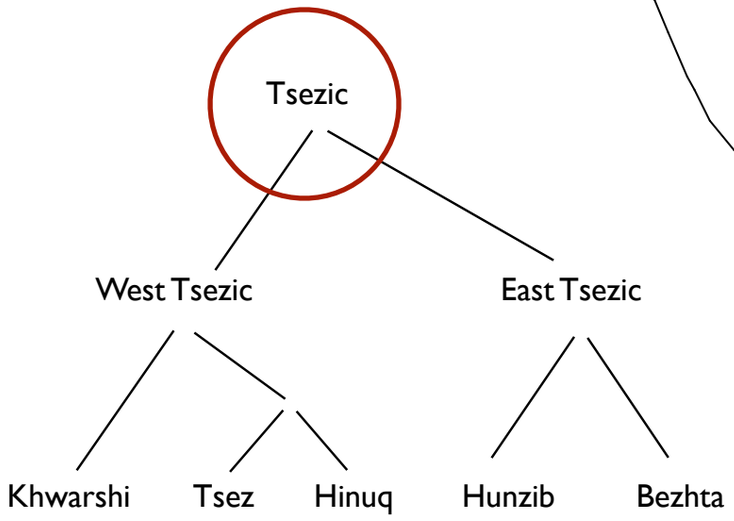
Directional Case Marking



Proto-Tsezic

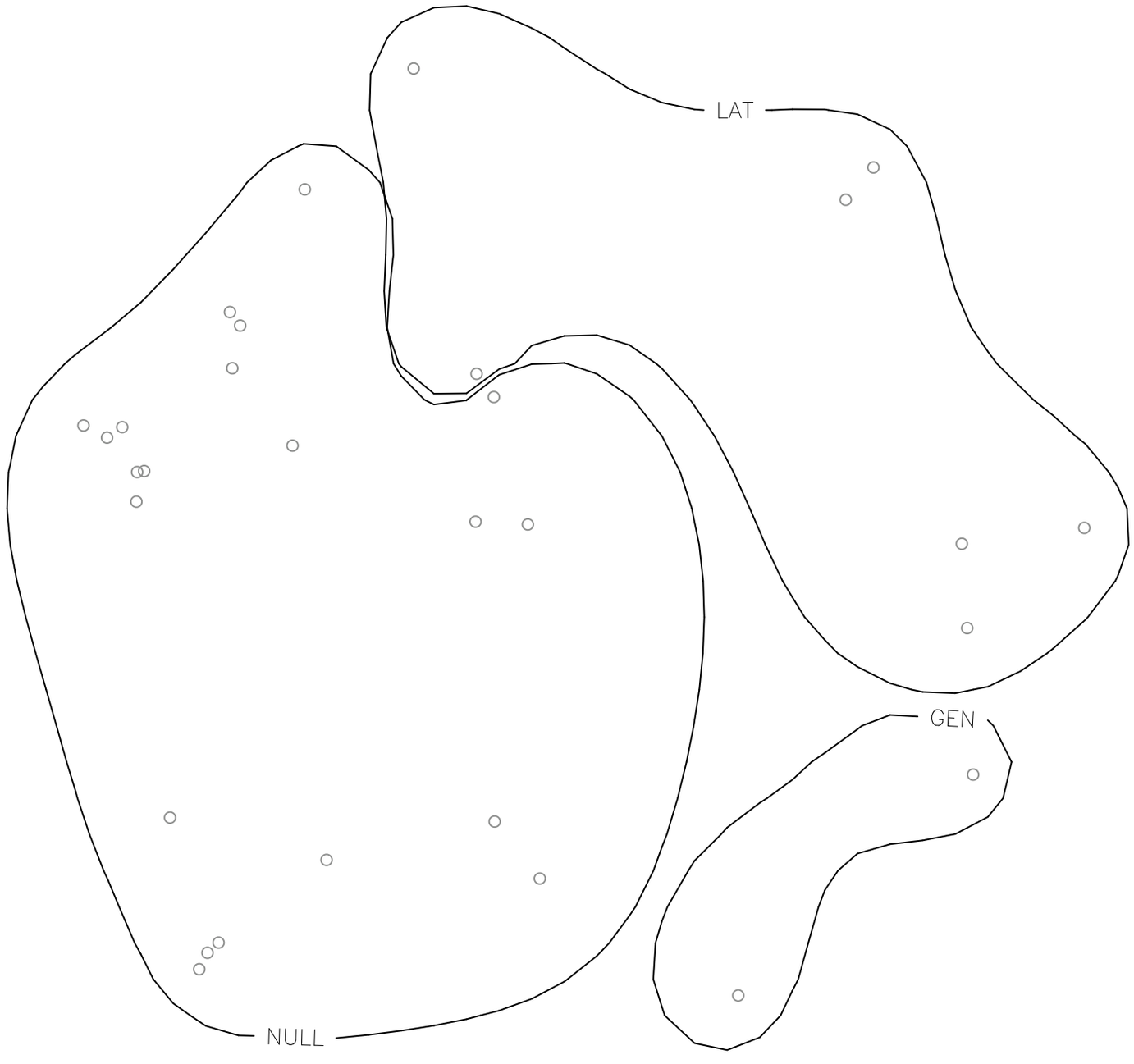
“directional”

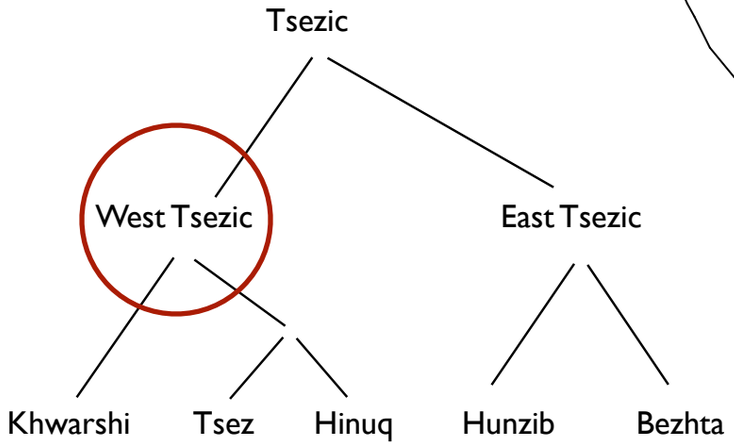




Proto-West-Tsezic

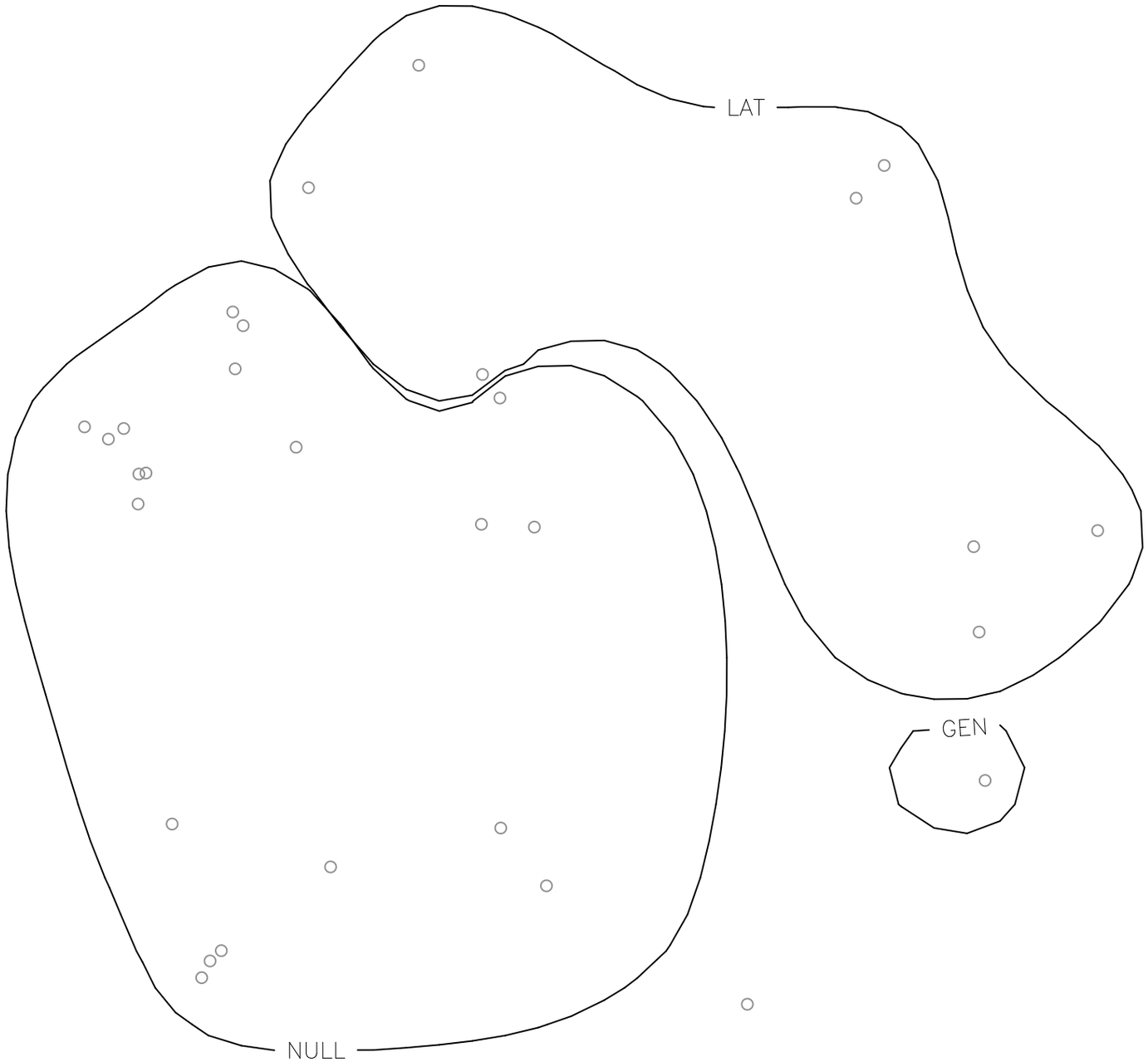
"directional"

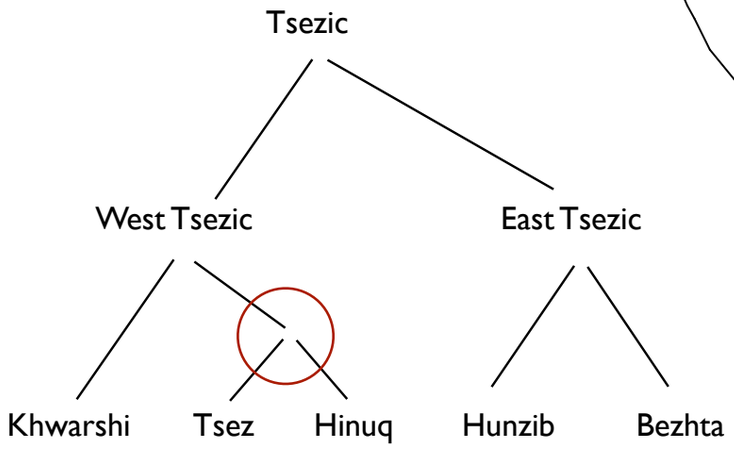




Proto-Hinuq-Tsez

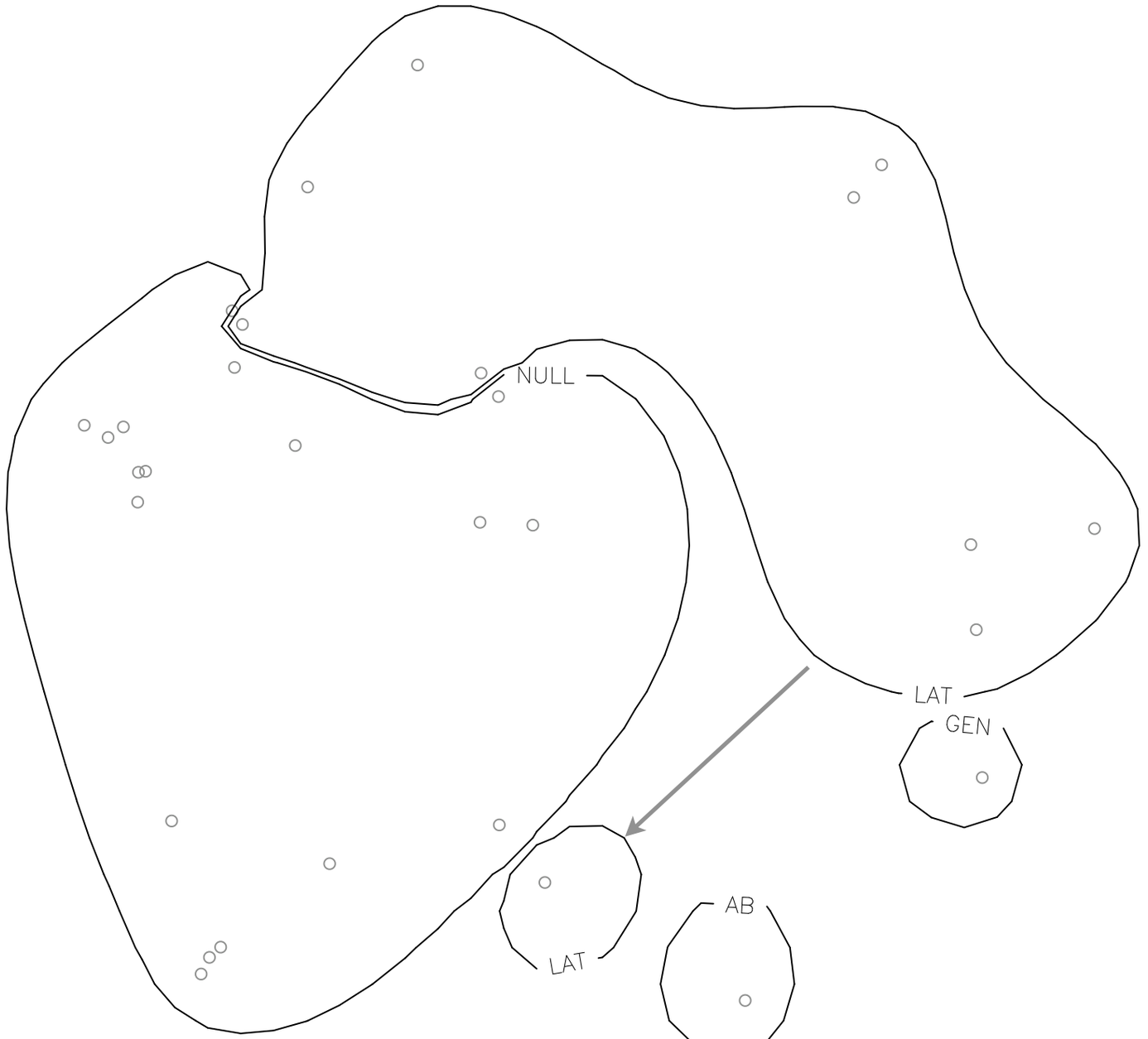
"directional"

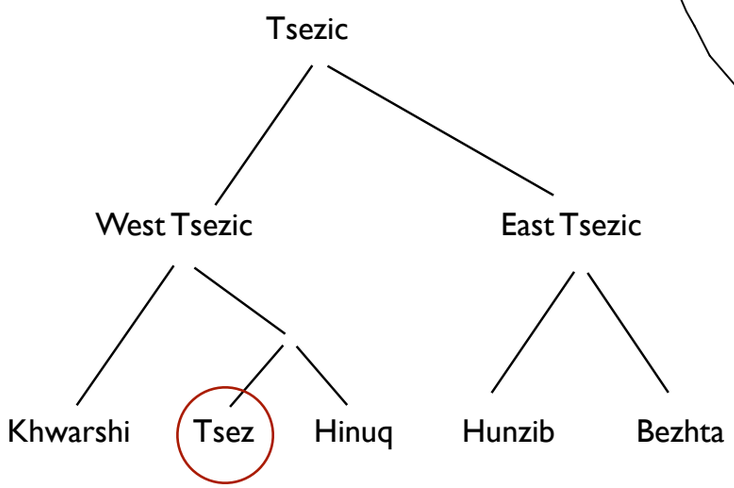




Tsez

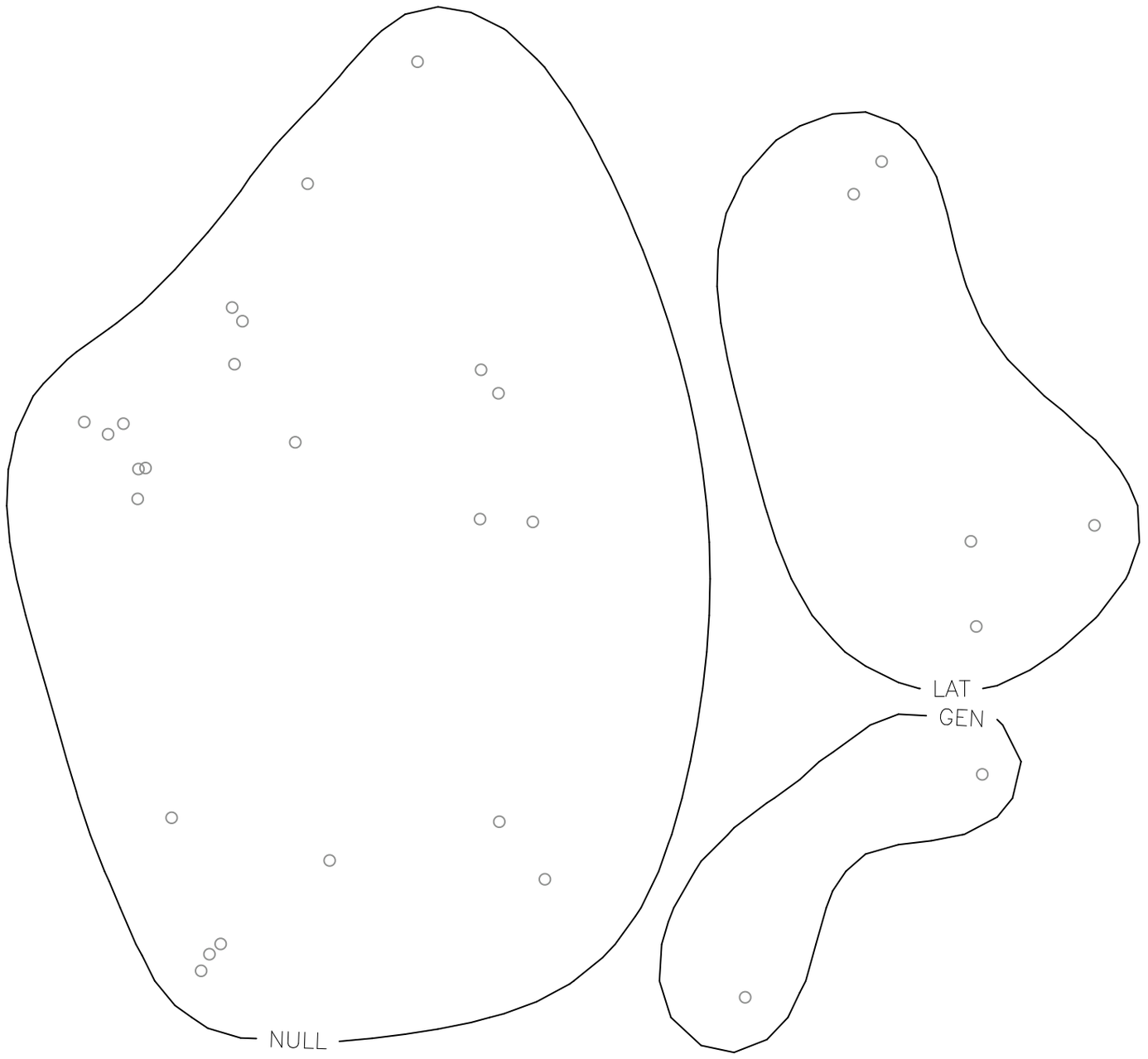
"directional"

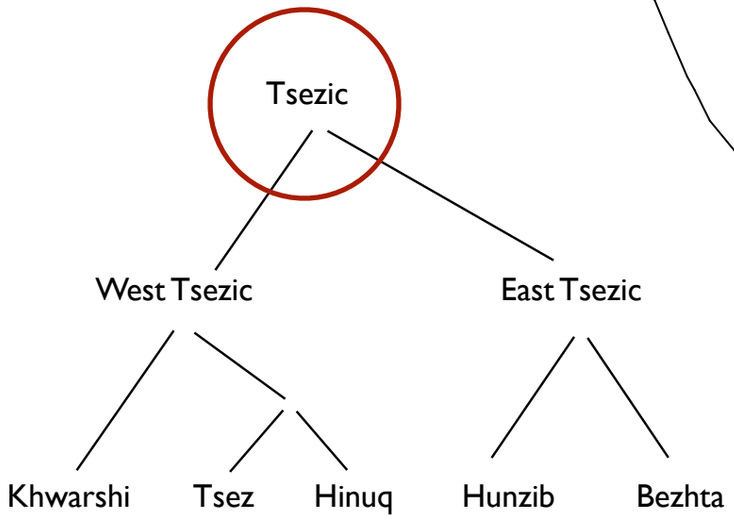




Proto-Tsezic

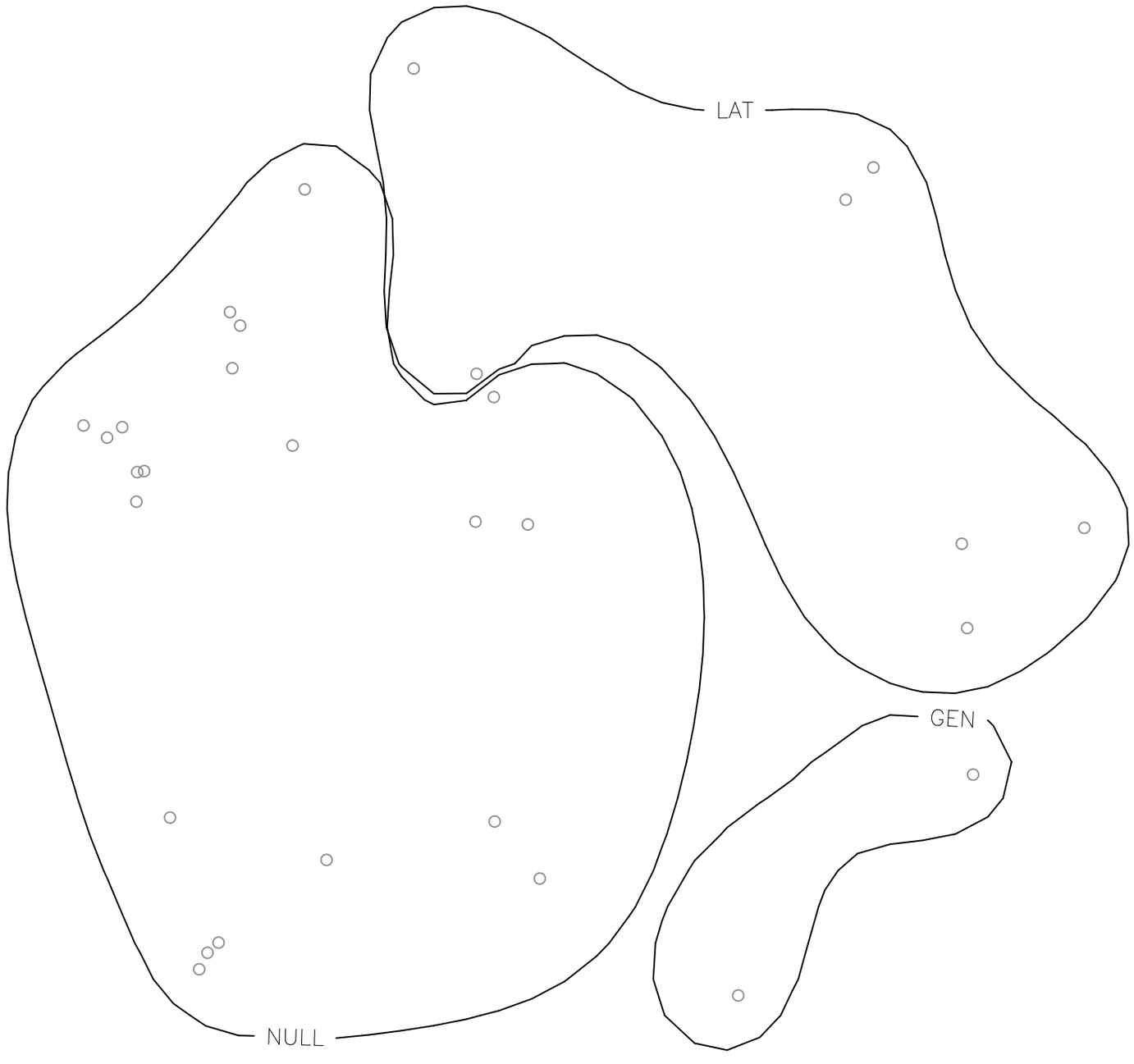
“directional”

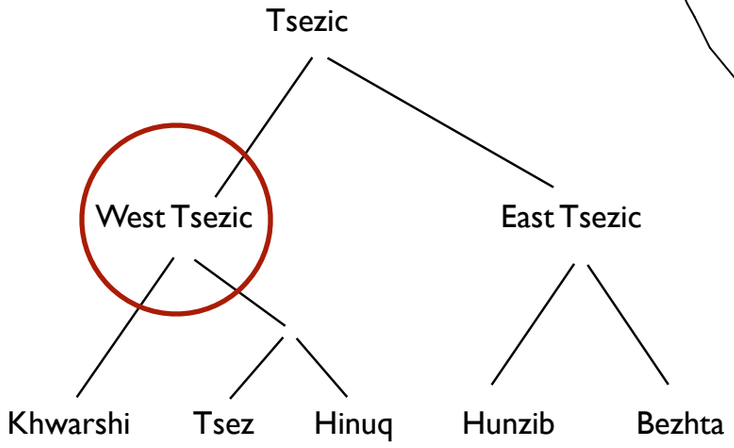




Proto-West-Tsezic

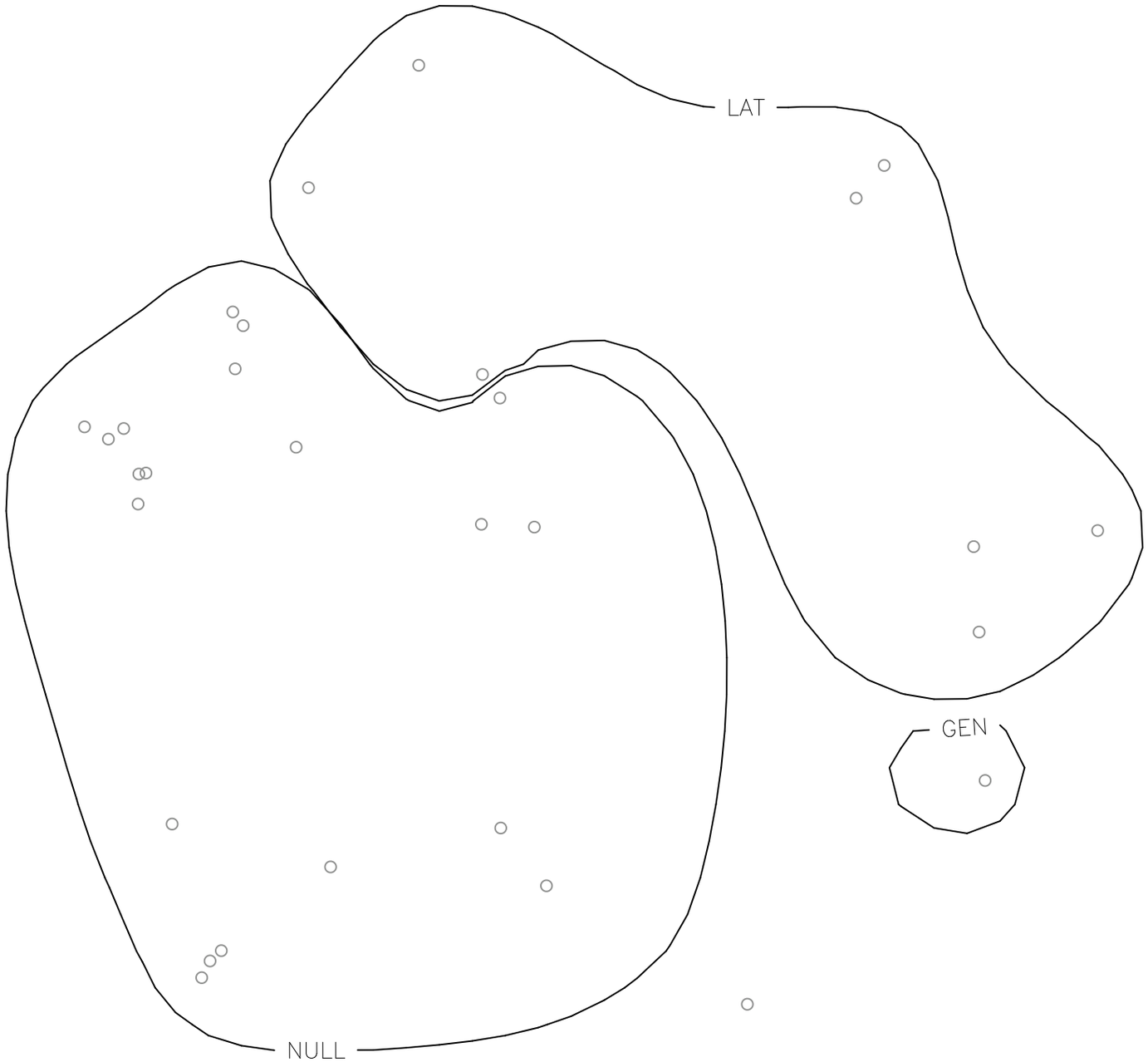
"directional"

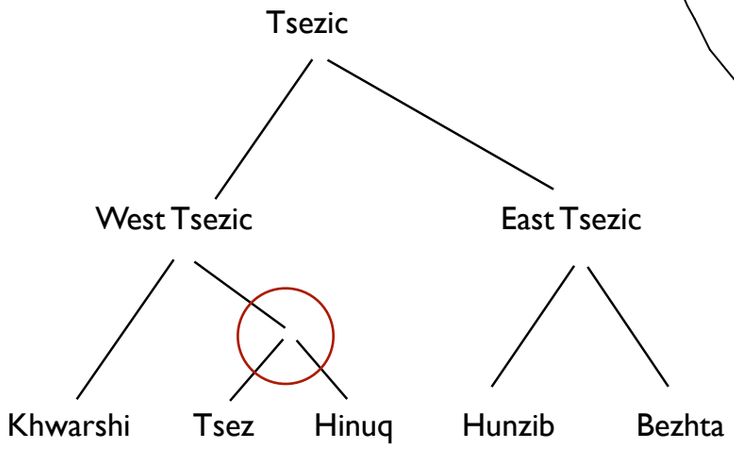




Proto-Hinuq-Tsez

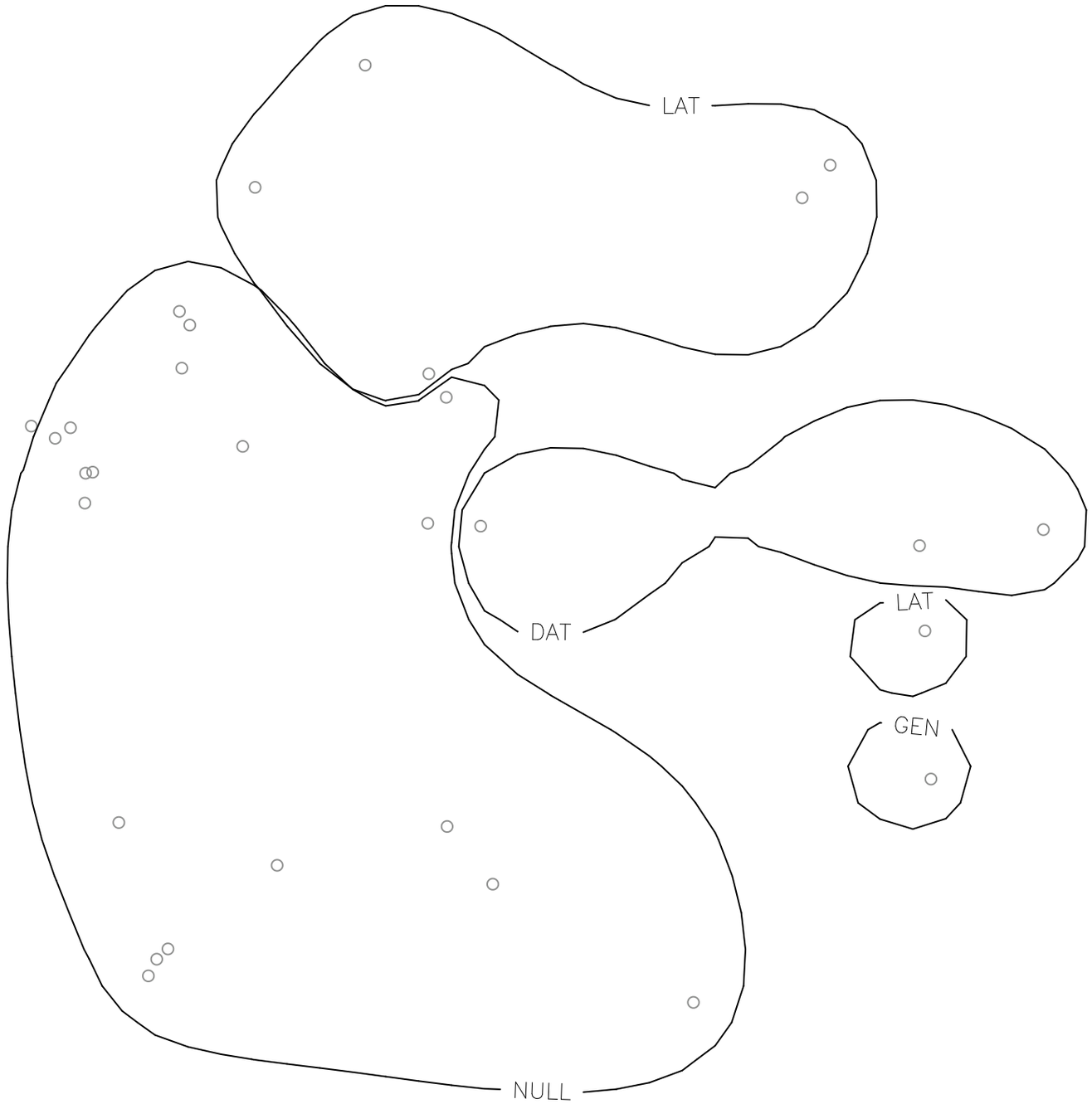
"directional"

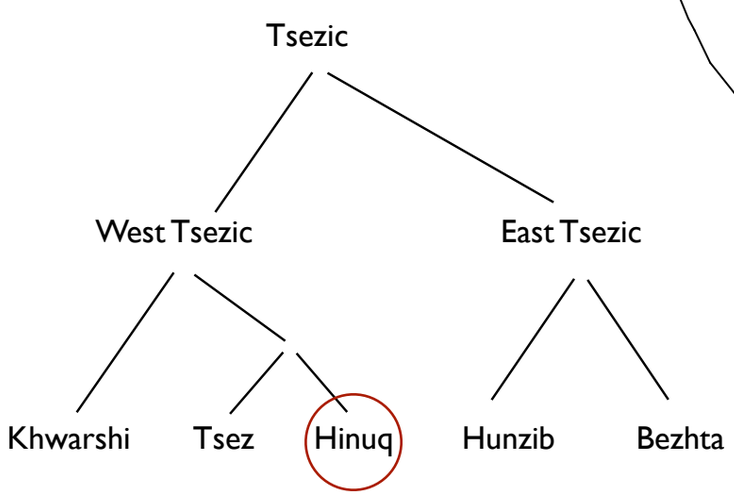




Hinuq

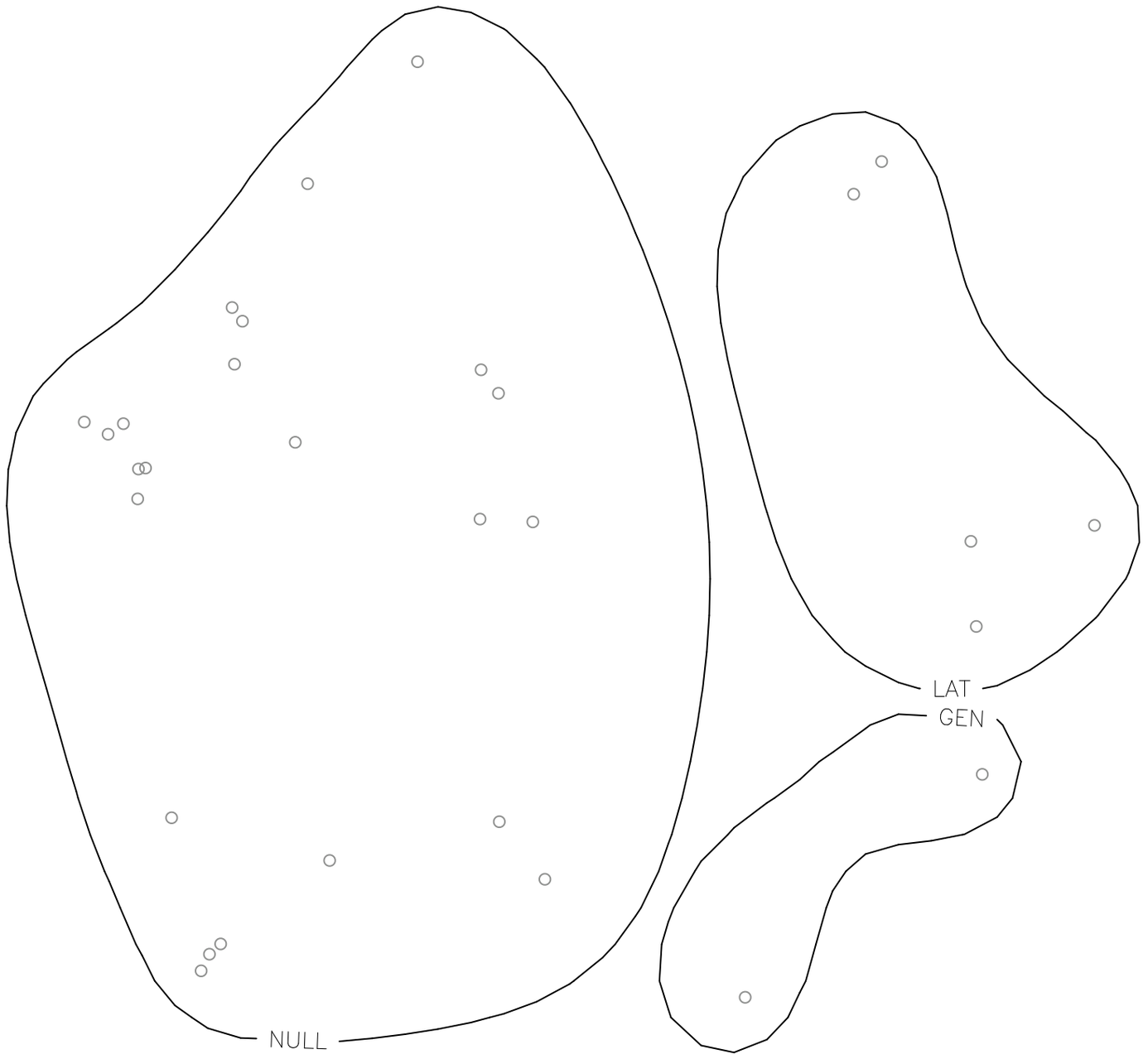
"directional"

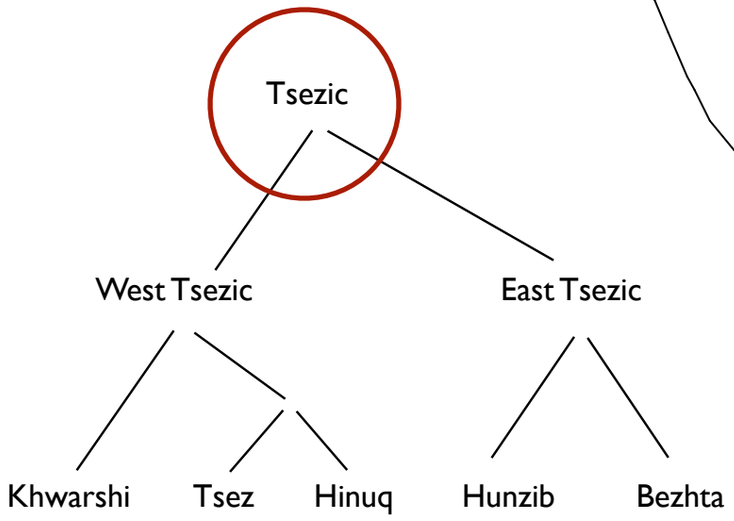




Proto-Tsezic

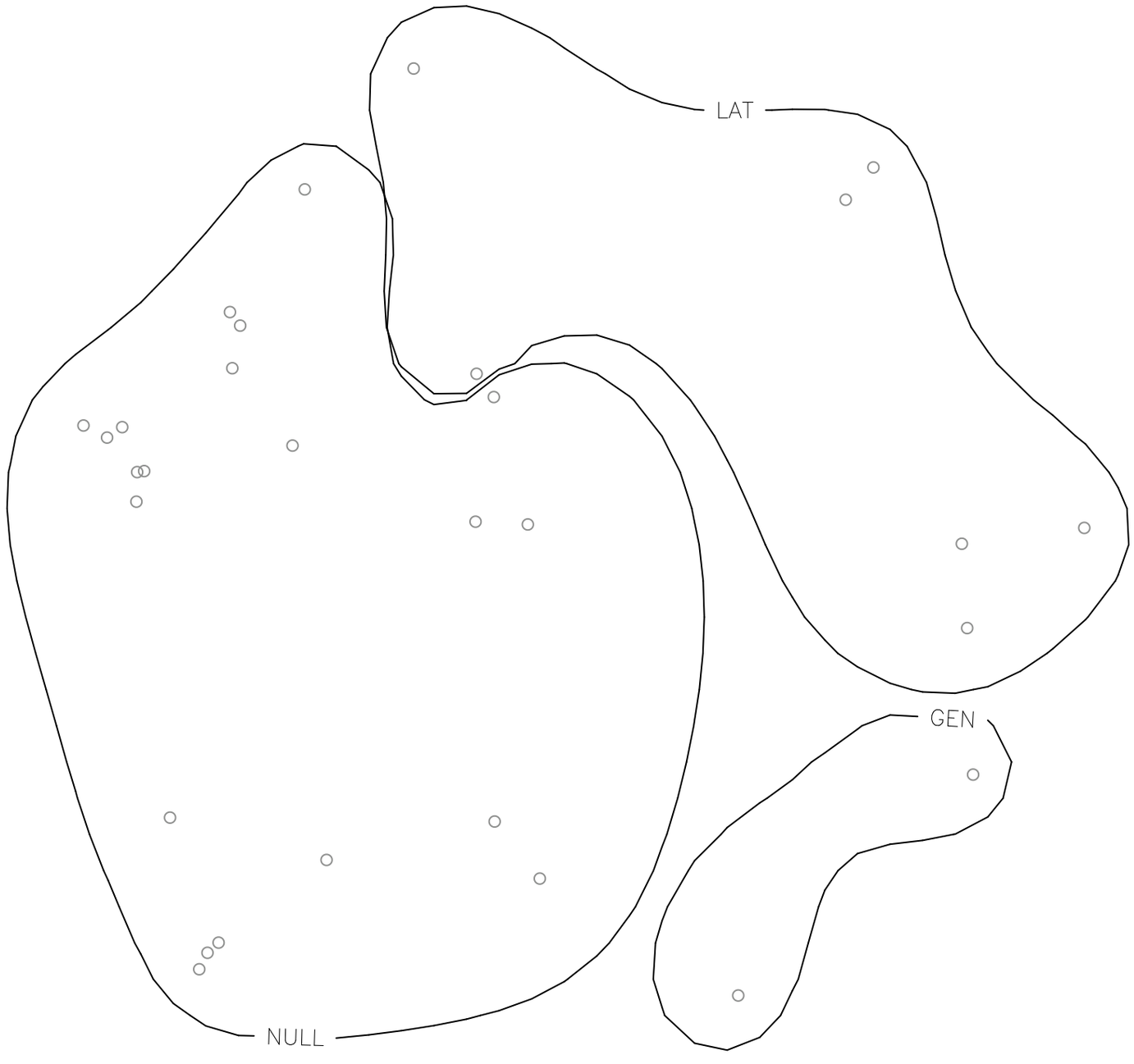
“directional”

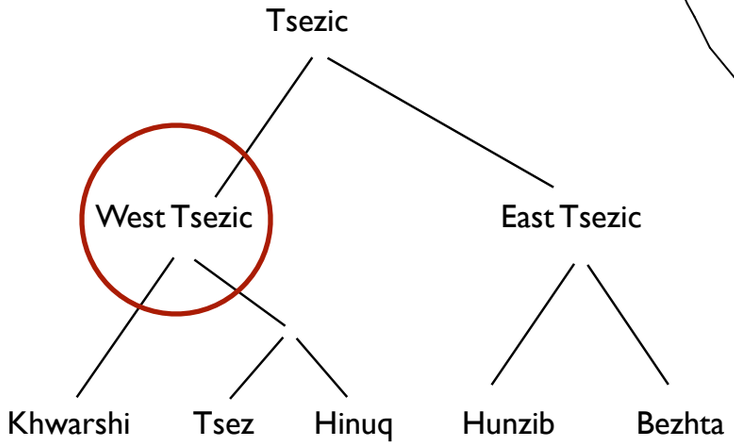




Proto-West-Tsezic

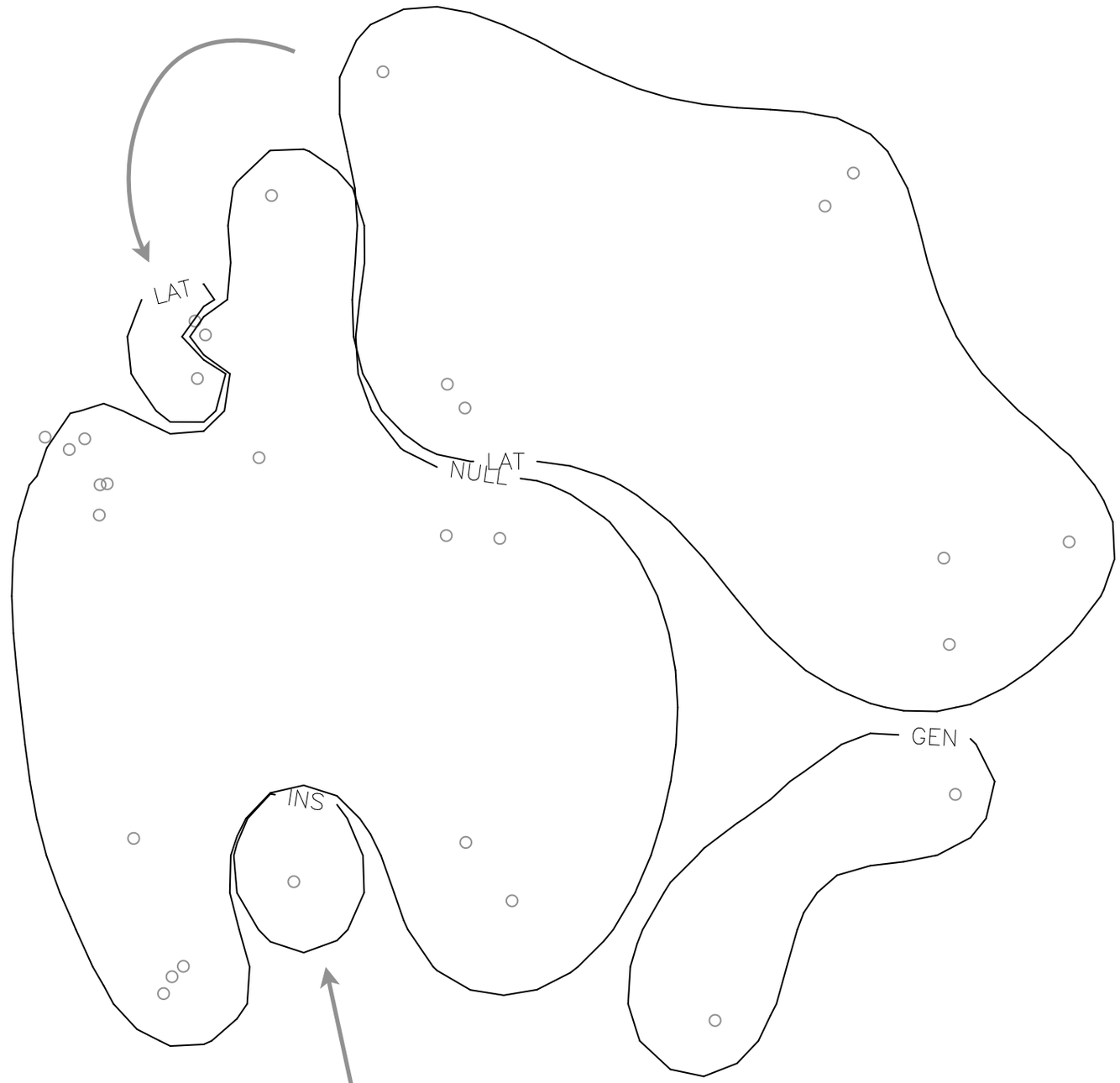
"directional"

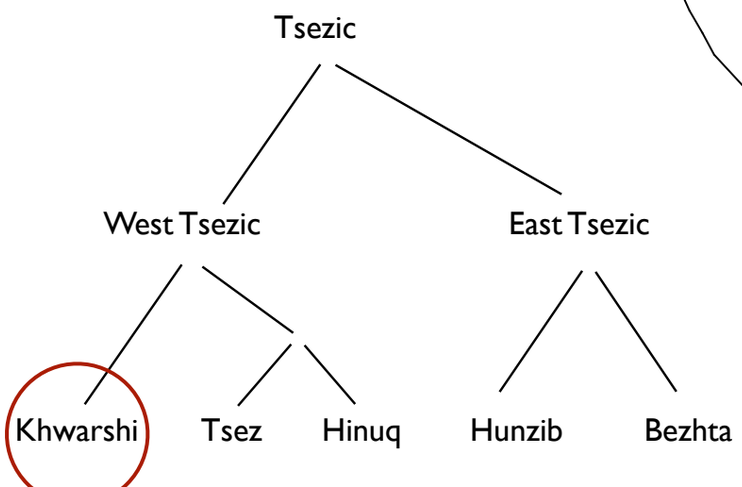




Khwarshi

“directional”



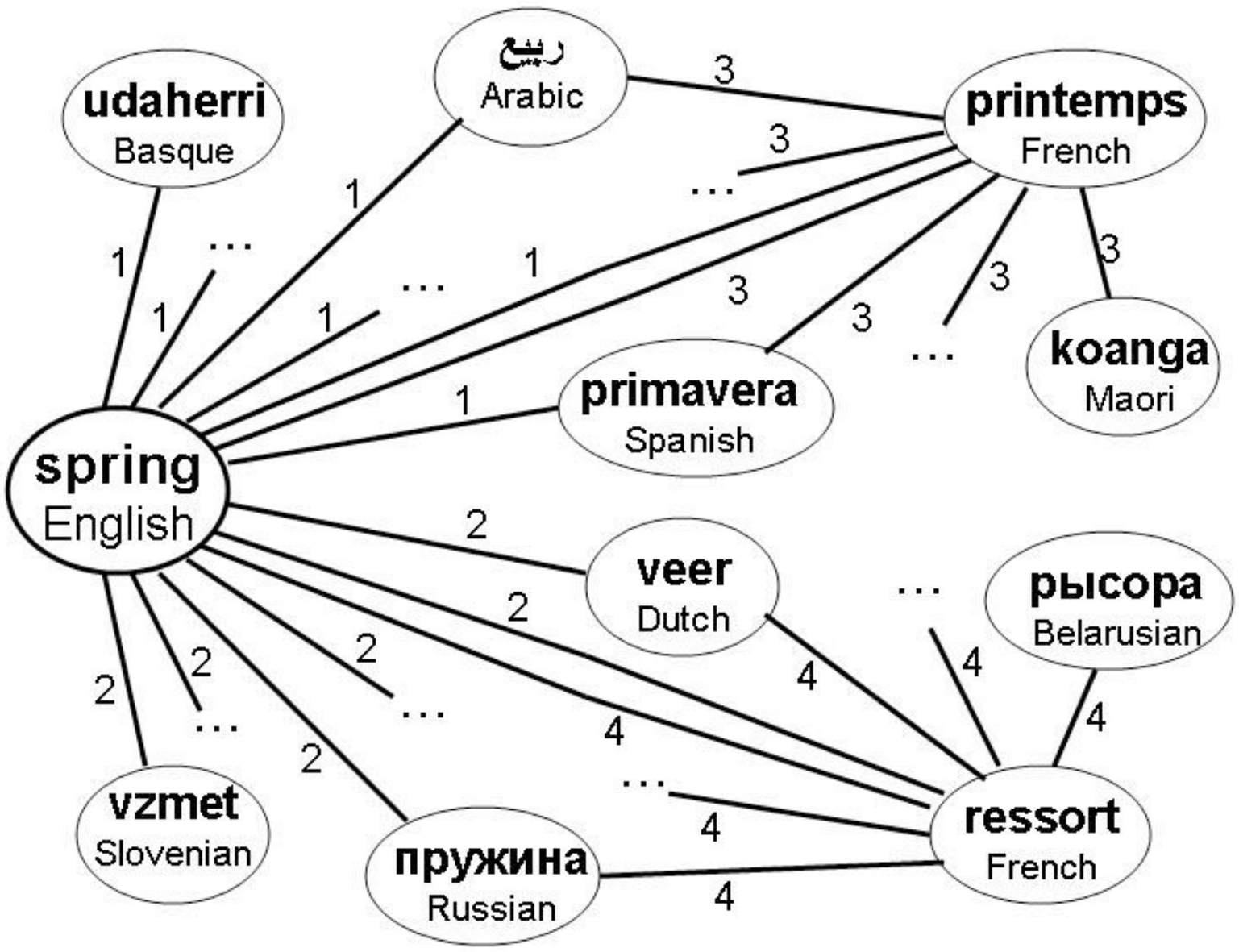


Challenges generalizing this approach

- How to find similar meanings across languages
- How to establish homologues (“cognates”) across languages

Translation Graph

(Etzioni et al. 2007)



Nodes are words, vertices are senses (intrinsic meanings)

Oren Etzioni, Kobi Reiter, Stephen Soderland, and Marcus Sammer. 2007. Lexical Translation with Application to Image Search on the Web. Proceedings of Machine Translation Summit XI.

Thinking further

- Consider word meanings as set of uses
- Translations show that sets overlap
- Make inferences about the structure of uses-space through available translations
- Extend this approach to higher structures
  phrases, sentences, paragraphs, texts

Finding and using homologues ("cognates")

- Traditional "historical-comparative method"
- Ignore changes in meaning
- Look for forms that have the same origin
- Changes in form to infer a phylogeny
- Similar to nucleotide-alignment

German *Hund*
Dutch *hond*
English *hound* (not *dog* !)

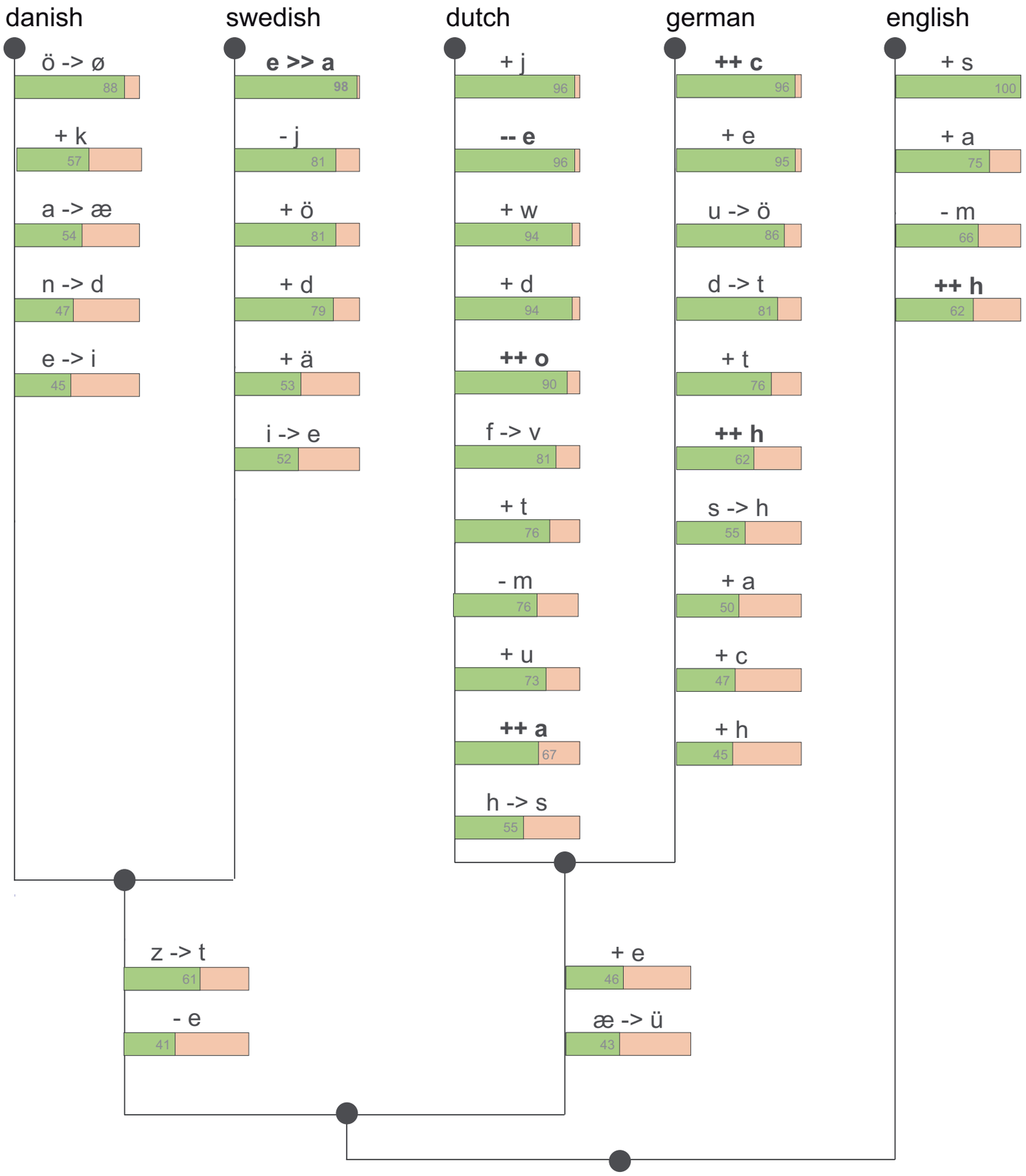
h	u	n	d
h	o	n	d
h	ou	n	d

Challenges

- How to find (possible) homologues
  meanings often have changed
- Alignment difficult
  strings are short
- But: letters are more informative than nucleotides
- Interpretation different
  Recurrent changes only count as one change, though the more frequent, the better the argument for phylogenetic structure

Hagen Jung

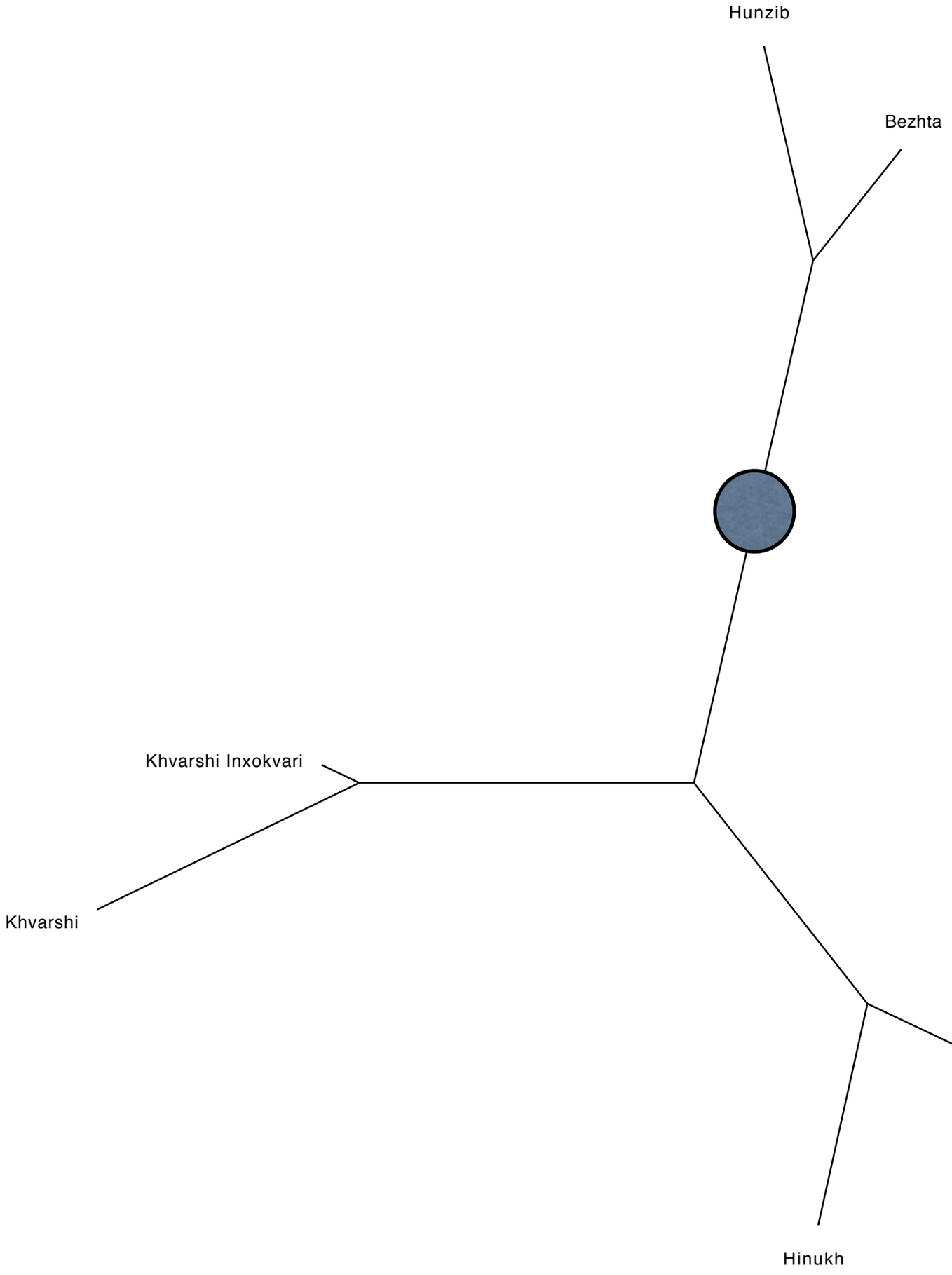
- Search for minimally necessary edits on a tree
- Evaluate cognates by their fit on a tree
- This results in a phylogeny with sound changes and cognate sets in one process

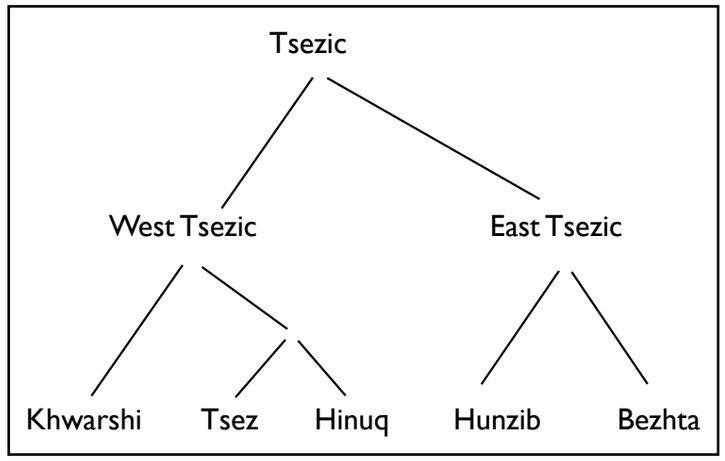

danish
ö -> ø
88
+ k
57
a -> æ
54
n -> d
47
e -> i
45
swedish
e >> a
98
- j
81
+ ö
81
+ d
79
+ ä
53
i -> e
52
dutch
+ j
96
-- e
96
+ w
94
+ d
94
++ o
90
f -> v
81
+ t
76
- m
76
+ u
73
++ a
67
h -> s
55
german
++ c
96
+ e
95
u -> ö
86
d -> t
81
+ t
76
++ h
62
s -> h
55
+ a
50
+ c
47
+ h
45
english
+ s
100
+ a
75
- m
66
++ h
62
z -> t
61
- e
41
+ e
46
æ -> ü
43

Lydia Steiner

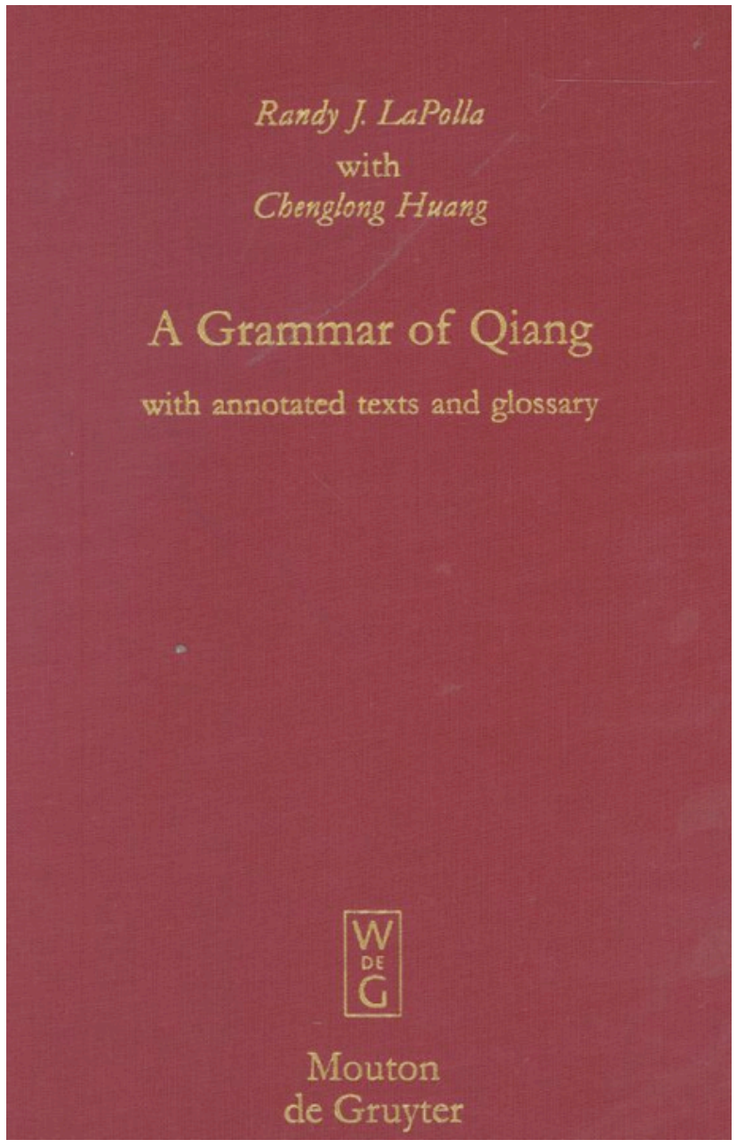
- Follow traditional linguistic methodology as closely as possible
- Guess cognates, refine through learning algorithm
- Use changes in alignment to infer tree

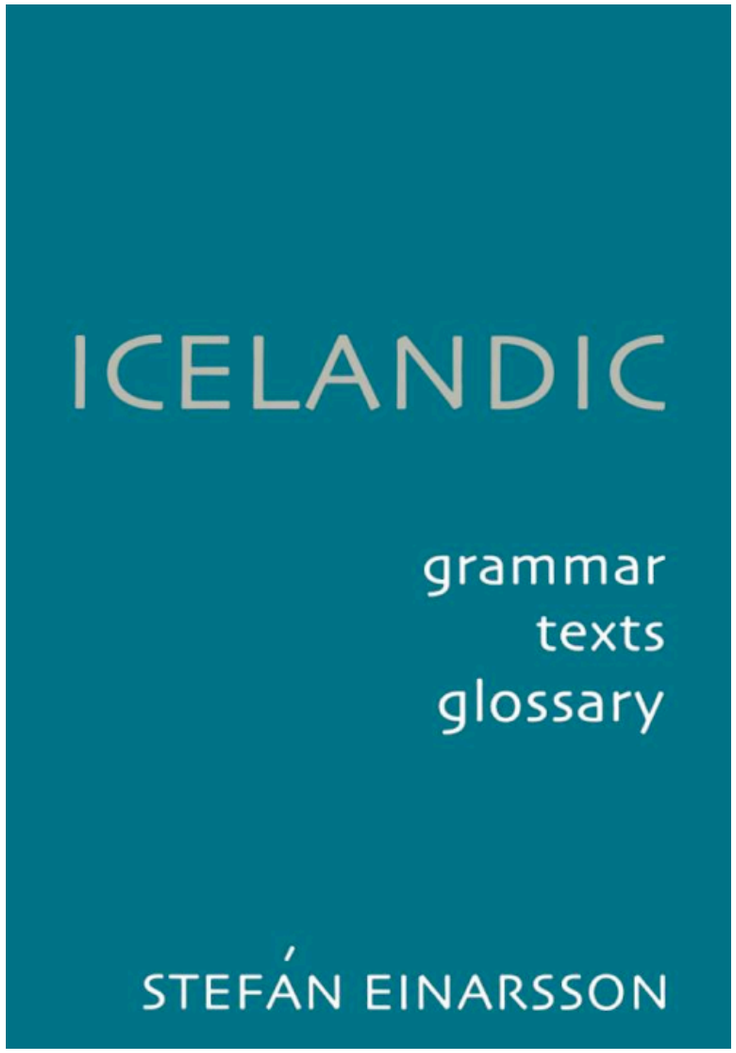
Hunzib	u	č	u	z	a	b
Bezhta	u	č	u	z	a	b
Khvarshi	u	č	u	z	a	r
Khvarshi Inxokvari	u	č	u	z	a	r
Hinukh	u	č	u	z	a	w
Tsez Sagadin	u	č	u	z	a	w
Tzes Mokok	u	č	u	z	a	w

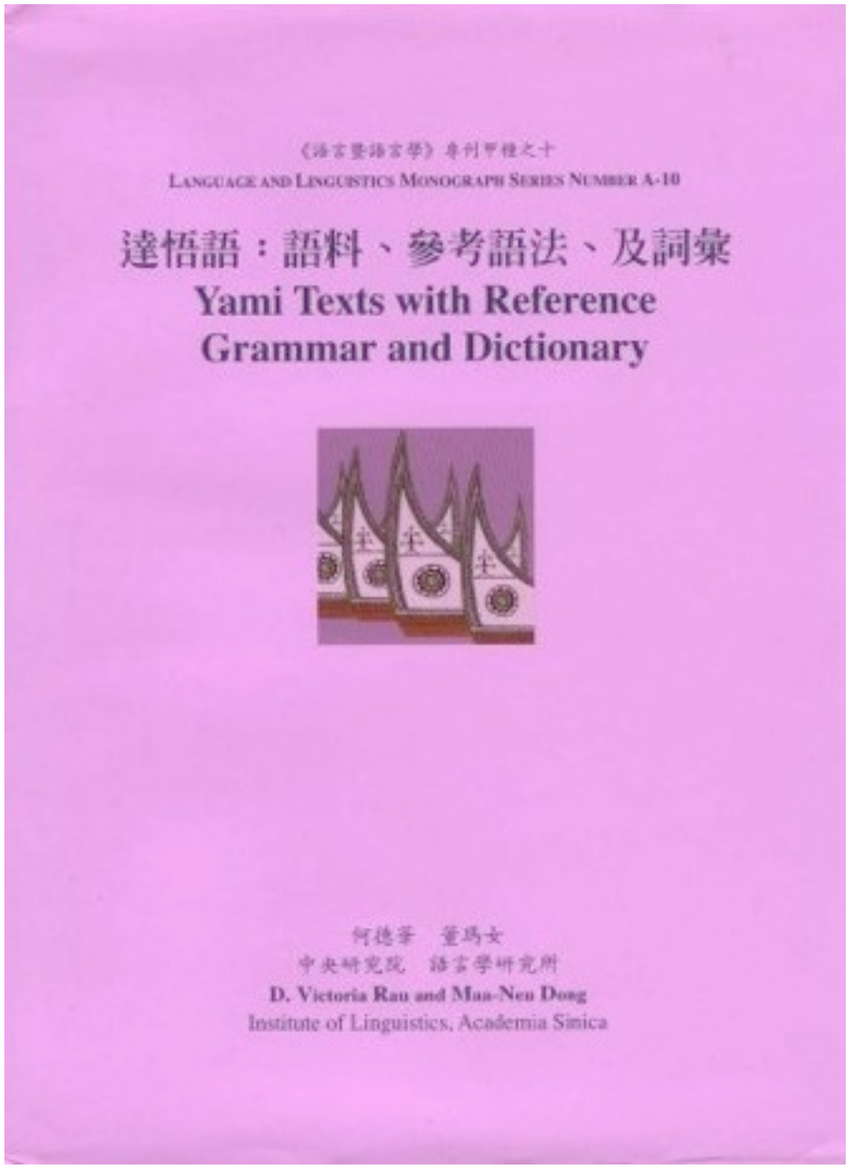

Hunzib
Bezhta
Khvarshi Inxokvari
Khvarshi
Hinukh
Tsez Sagadin
Tsez Mokok


Tsezic
West Tsezic
East Tsezic
Khwarshi
Tsez
Hinuq
Hunzib
Bezhta

The Descriptional Trinity:







- Grammar
- Text
- Dictionary