

# Historical reconstruction through comparison of meaning

Michael Cysouw  
LMU München

# Lexical Reconstruction

- Sound change
- Meaning change



# The Intercontinental Dictionary Series

*Founding Editor:*

Mary Ritchie Key (University of California, Irvine)

*General Editor:*

Bernard Comrie (Max Planck Institute for Evolutionary Anthropology, Leipzig)

**Purpose:** The purpose of the IDS is to establish a database where lexical material across the continents is organized in such a way that comparisons can be made. Historical studies, comparative, and theoretical linguistic research can be based on this documentation. This is a long-term cooperative project that will go on for the next generation or so and will involve linguists all over the world. It is aimed towards international understanding and cooperation. This is a pioneering effort that will have global impact. The purpose also contributes to preserving information on the little-known and "non-prestigious" languages of the world, many of which are becoming extinct.

**Rationale:** Information on languages of the world is scattered over all the continents and islands and published in dozens of languages and scripts. There is need of a database where one can find comparable material to formulate hypotheses and test and validate those theories. For example, theories on intercontinental connections have been proposed on the basis of the distribution of 'sweet potato' and yet there is no single source, where words with this meaning can be found in many languages. Good quantitative and statistical studies are almost impossible to do now in non-Western languages. The IDS will provide a quantitative base for a scientific approach to language analysis and comparisons. The IDS will provide the research tools necessary for expanding studies such as phonological theory, word formation, language change, lexical distribution, symbolism and onomatopoeia, classification, and other ideas that have to do with history of people and migrations. The IDS will serve not only as a synonym dictionary but as an index to meaning and to cultures of various people around the earth.

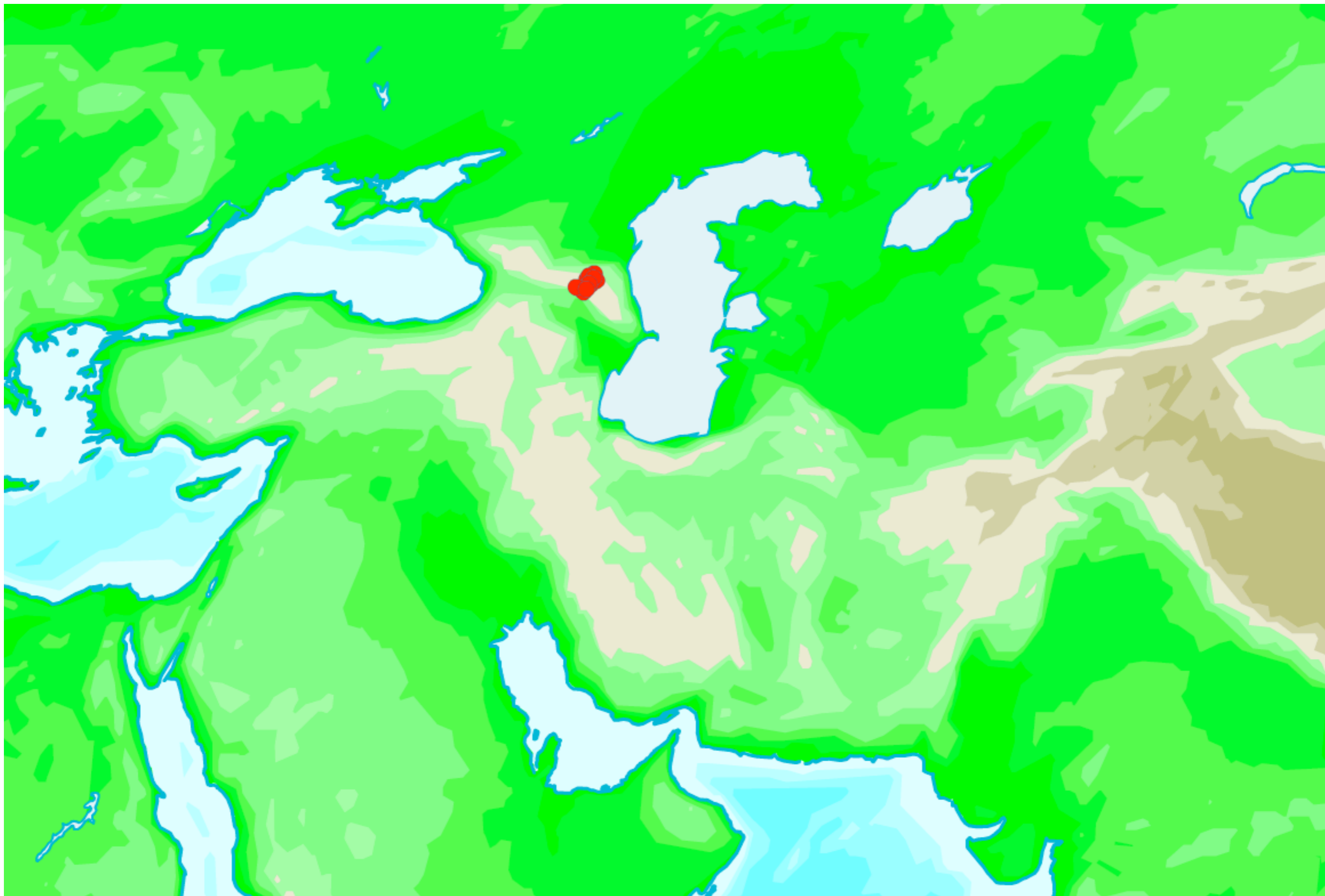
**Plan of Series:** The IDS series may appear as 1) a volume with 25 or more languages recorded; 2) a fascicle with 5 to 10 languages recorded; 3) in single WordLists, which are archived until enough are gathered to make up a fascicle or volume. A list of fascicles and volumes in progress is available from the general editor.

**Procedure:** The IDS is developed in cooperation and complementation with other research projects. Throughout the world there are linguistic activities from establishing of databases in universities and think-tanks to publishing grammar series and literacy materials, to individual projects such as the Tibetan dictionary project. Many projects seek to make linguistic data accessible in a format that will allow generalizations to be made. The computer now gives us the potential for tying together linguistic databases. The IDS editors will continue to monitor linguistic activity around the world, both for choosing the languages for forthcoming compilations and for collaboration with other research teams.

[IDS Main Page](#)  
[Simple Browsing](#)  
[Advanced Browsing](#)  
[Download Data](#)  
[Technical Background](#)

<http://lingweb.eva.mpg.de/ids/>





Tsezic

West Tsezic

East Tsezic

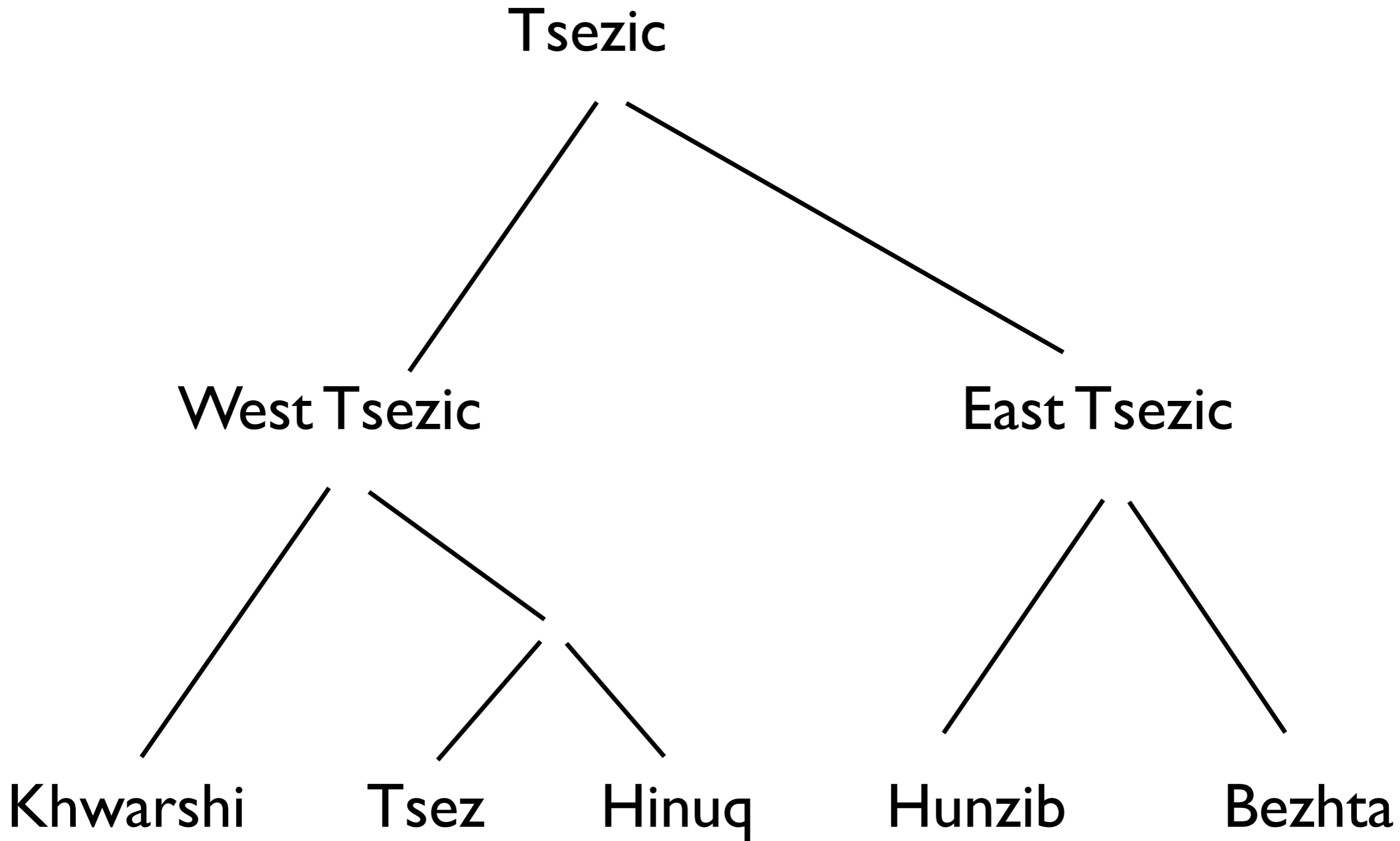
Khwarshi

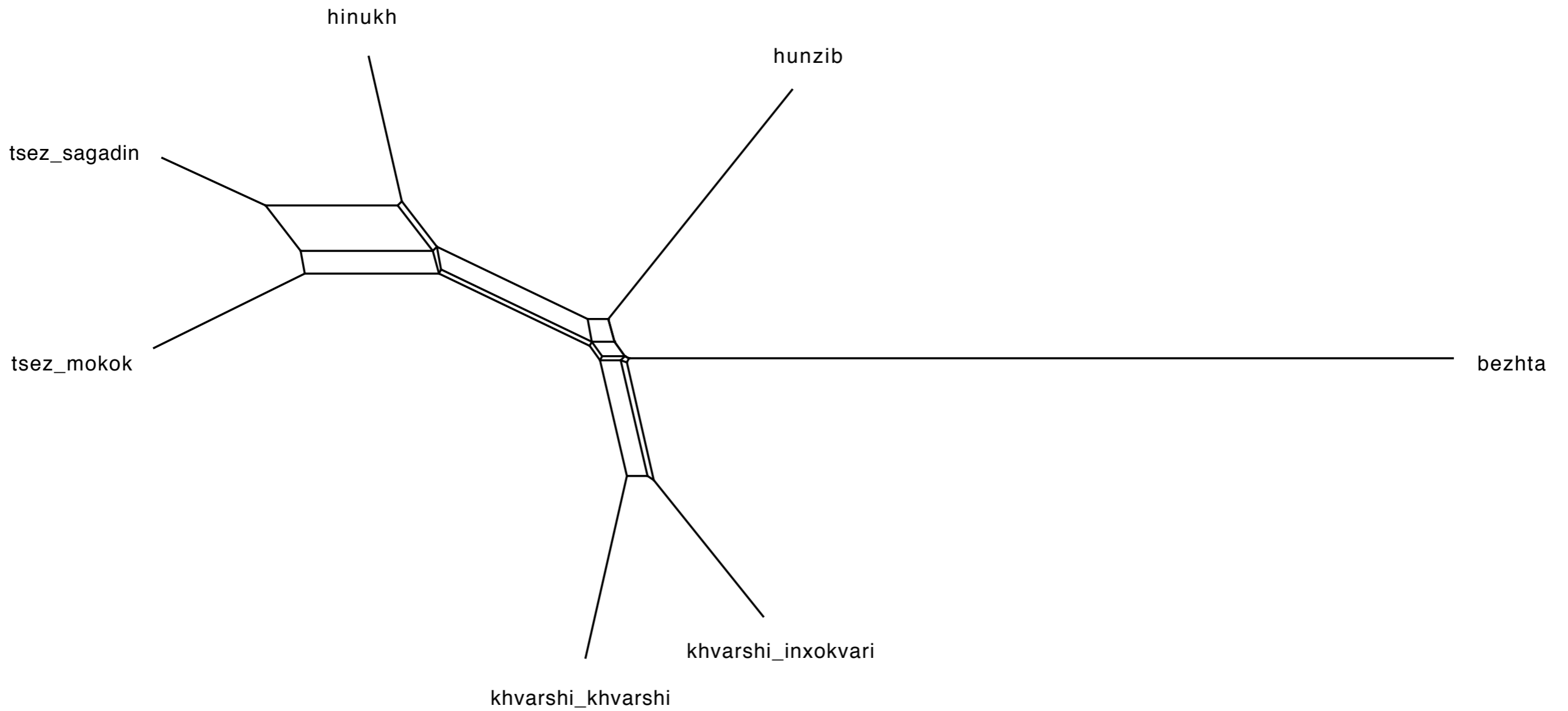
Tsez

Hinuq

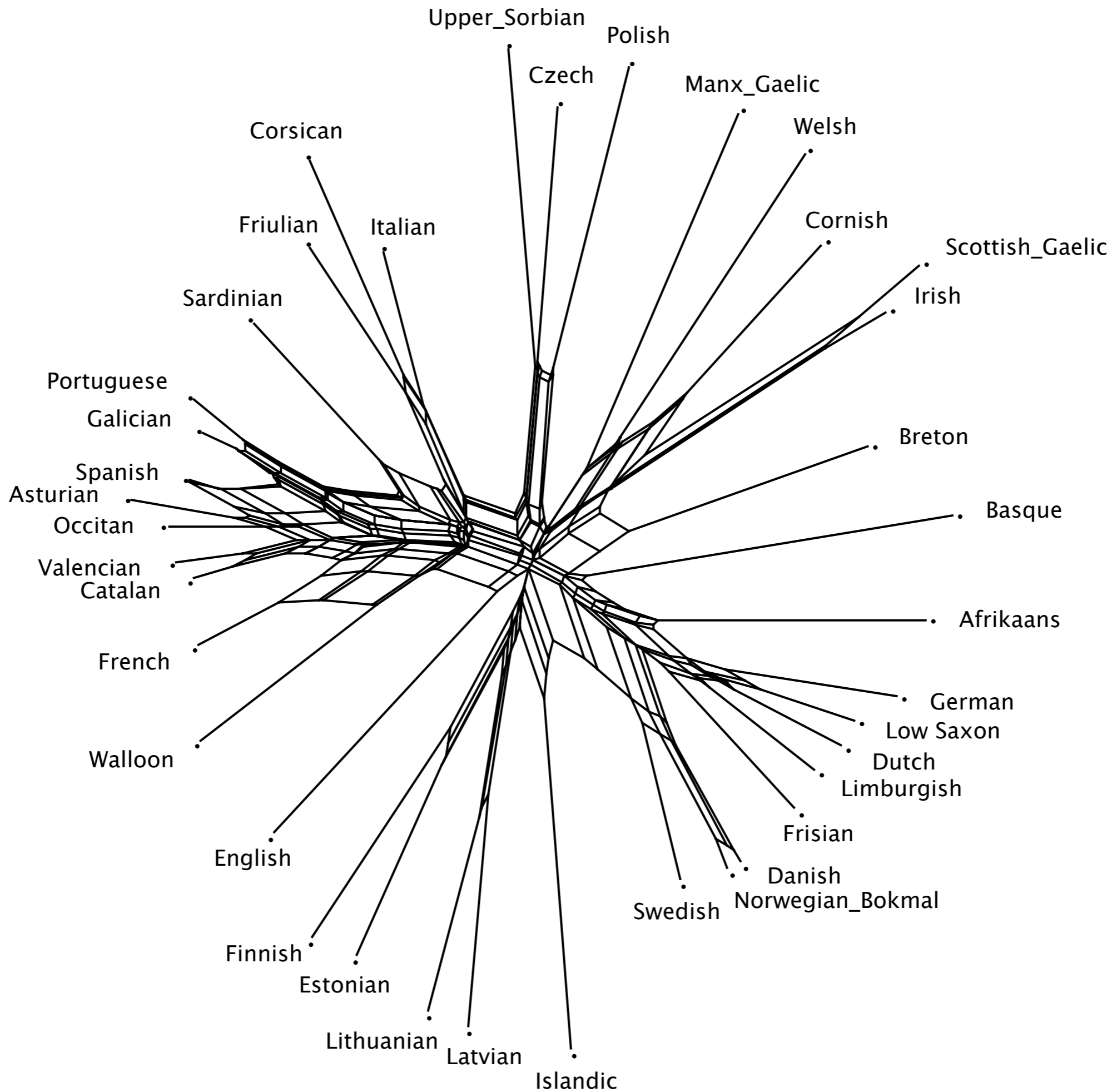
Hunzib

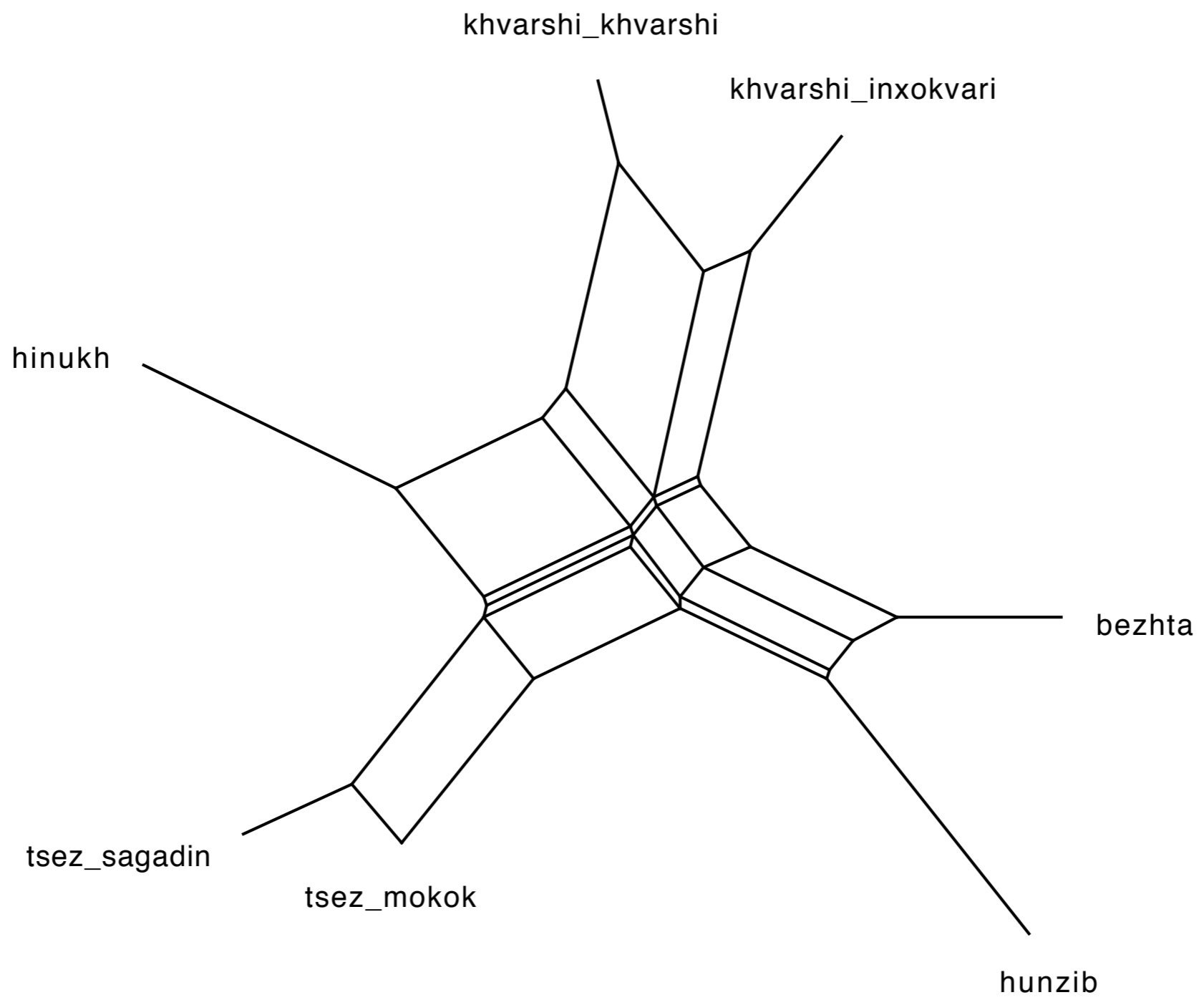
Bezhta





# 3-gram vector comparison



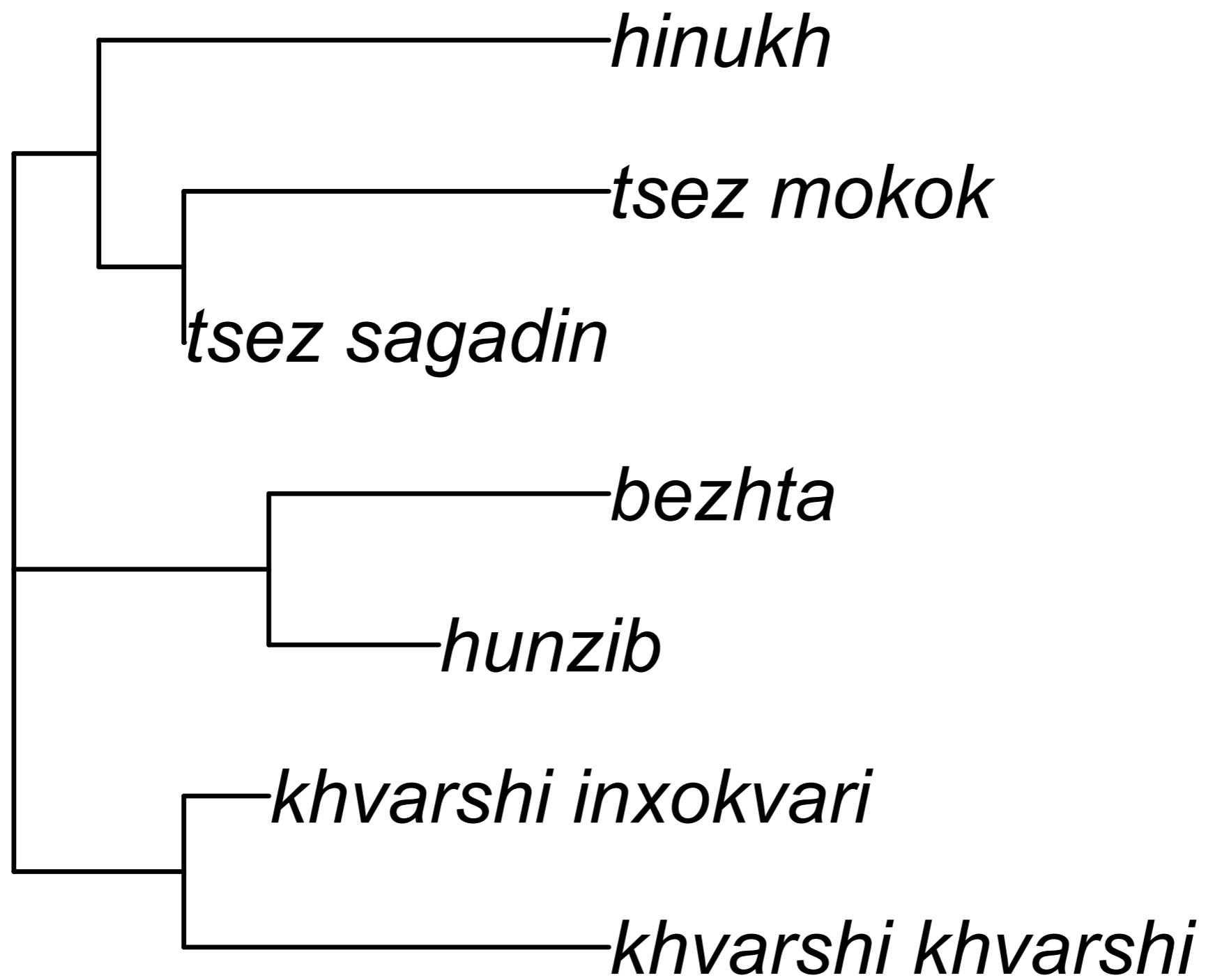


# Levenshtein distances



<b>IDS</b>	<b>English</b>	<b>Hinukh</b>
10.470	go	iλ'a
10.471	go up	šid iλ'a
2.330	marry	xoddoho yiλ'a
...	...	...

bezhta	0	1	0	0	1	0	0	0	0	1	1	1	0	1	0	0	0	1	0	0	1	1	0	0	0	0	0	1	1	
hinukh	0	1	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1	1	1	0	1	1	1	0	1	1	0	1	0	0
hunzib	0	0	0	0	1	0	0	0	0	1	1	1	1	1	0	0	1	0	0	0	1	1	0	0	0	0	0	1	1	0
khvarshi inxokvari	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	1	0	0
khvarshi khvarshi	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tsez mokok	1	1	0	0	0	0	1	1	1	0	1	1	0	1	0	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0
tsez sagadin	0	1	0	0	0	0	0	1	0	0	1	1	0	1	0	0	1	1	0	0	1	1	0	0	0	0	0	1	0	0



# Meaning/Function-space

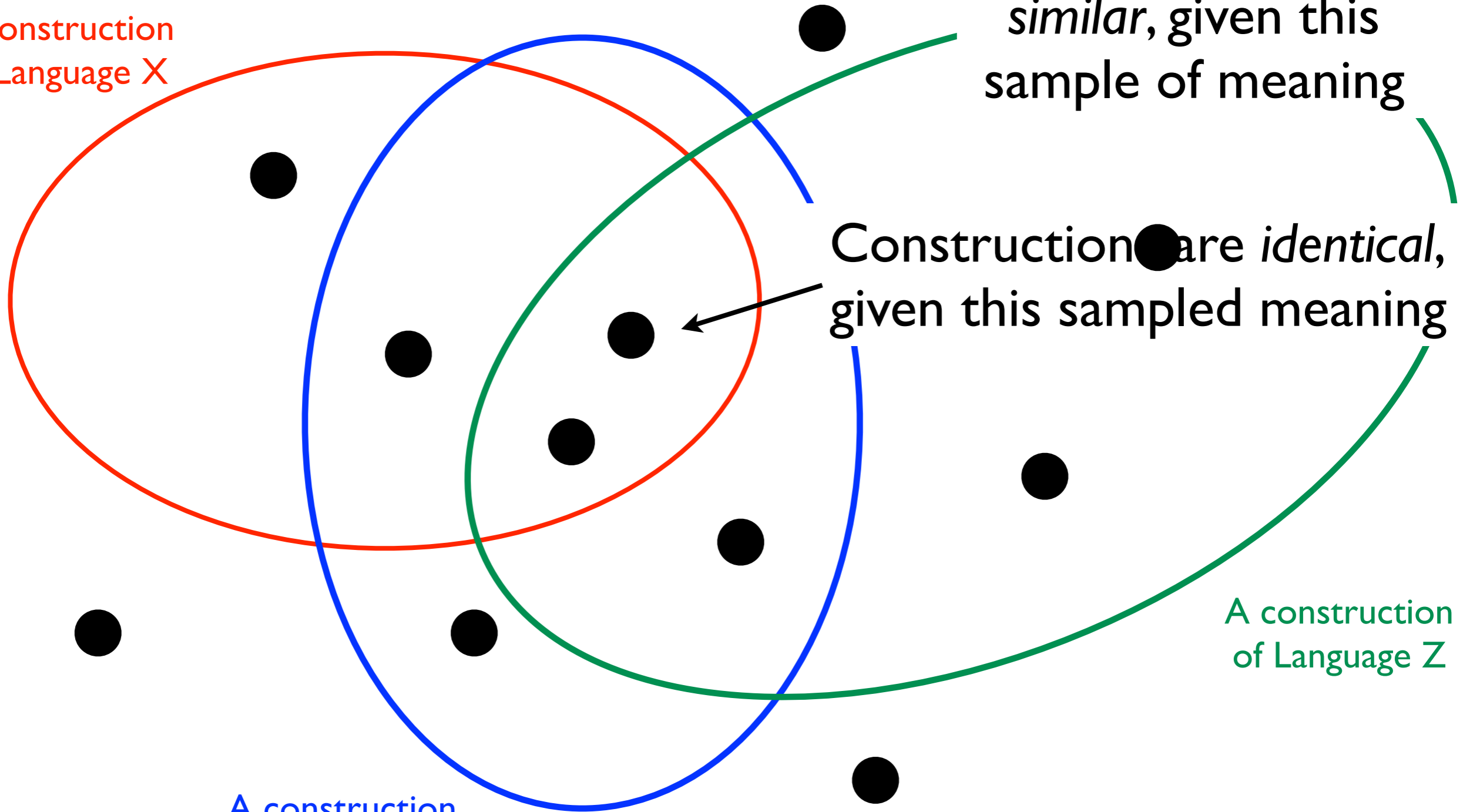
A construction  
of Language X

Constructions are  
*similar*, given this  
sample of meaning

Constructions are *identical*,  
given this sampled meaning

A construction  
of Language Z

A construction  
of Language Y

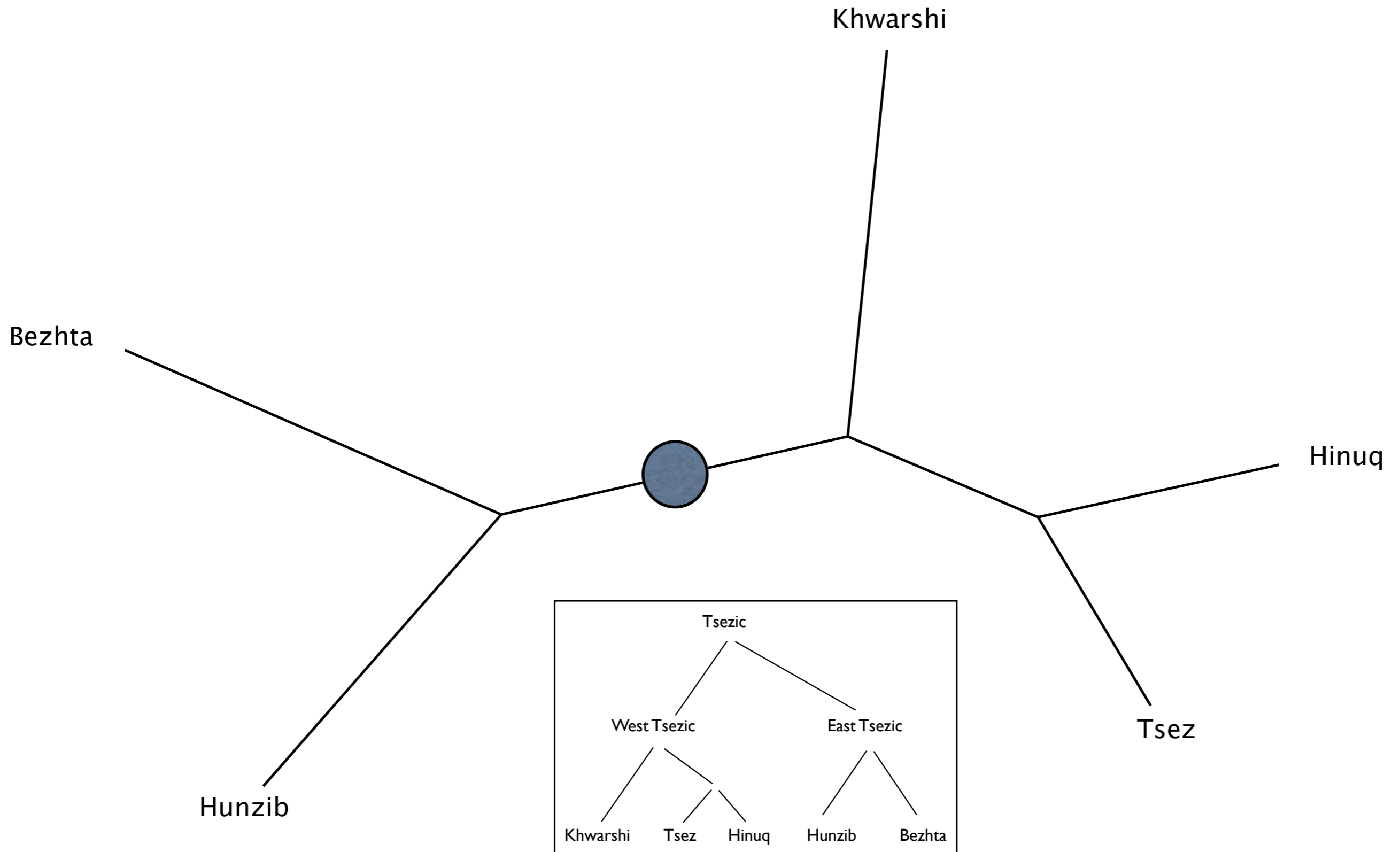


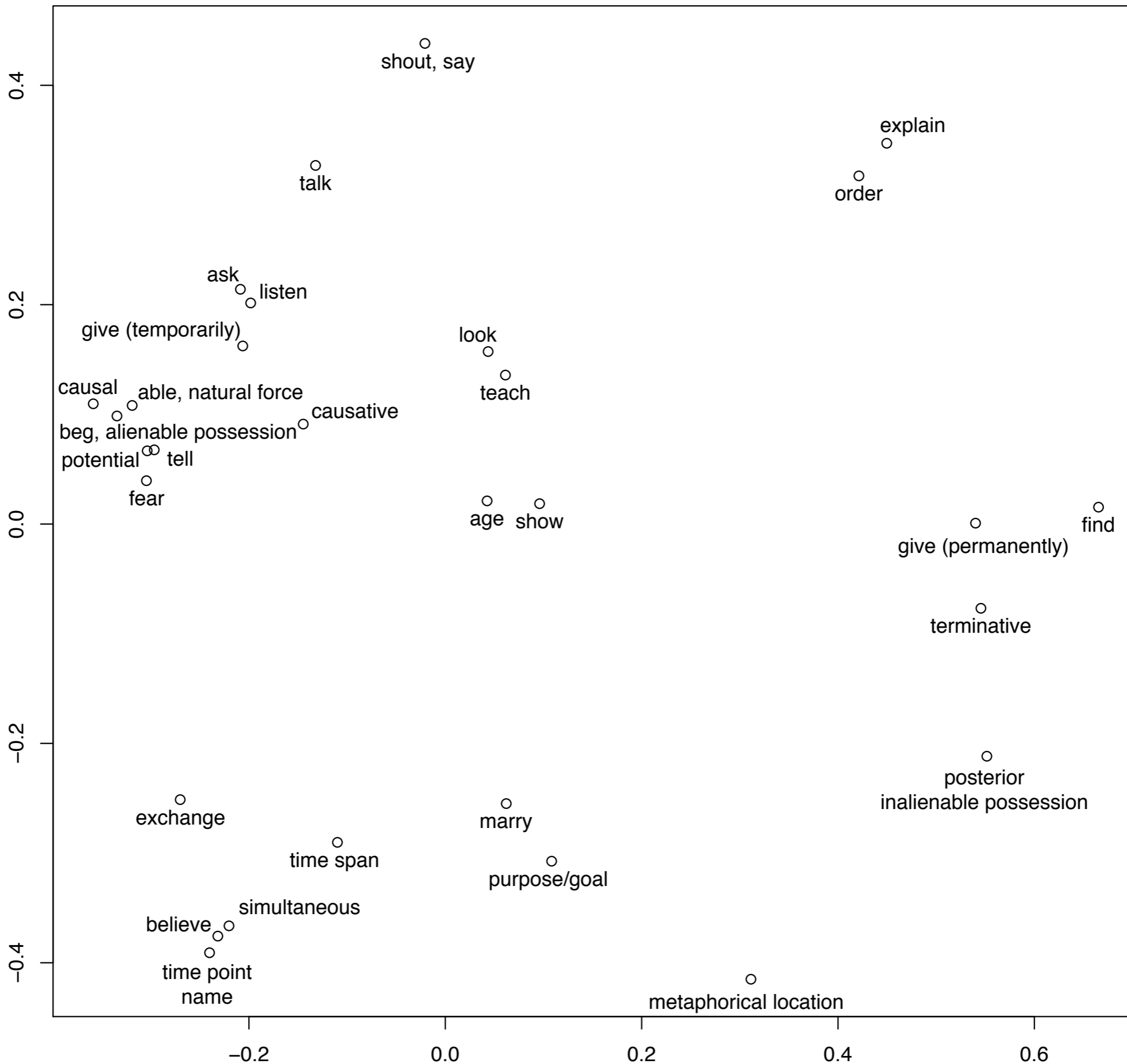
Role	Label	Construction
Addressees	TALK	talk <i>[to somebody]</i>
	TELL	tell <i>[to somebody]</i>
	SHOUT	shout <i>[at somebody]</i>
	SAY	say <i>[to somebody]</i>
	ASK	ask <i>[somebody]</i>
	BEG	beg <i>[somebody]</i>
	EXPLAIN	explain <i>[to somebody]</i>
	TEACH	teach <i>[to somebody]</i>
	ORDER	order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

Function	Hinuuq		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



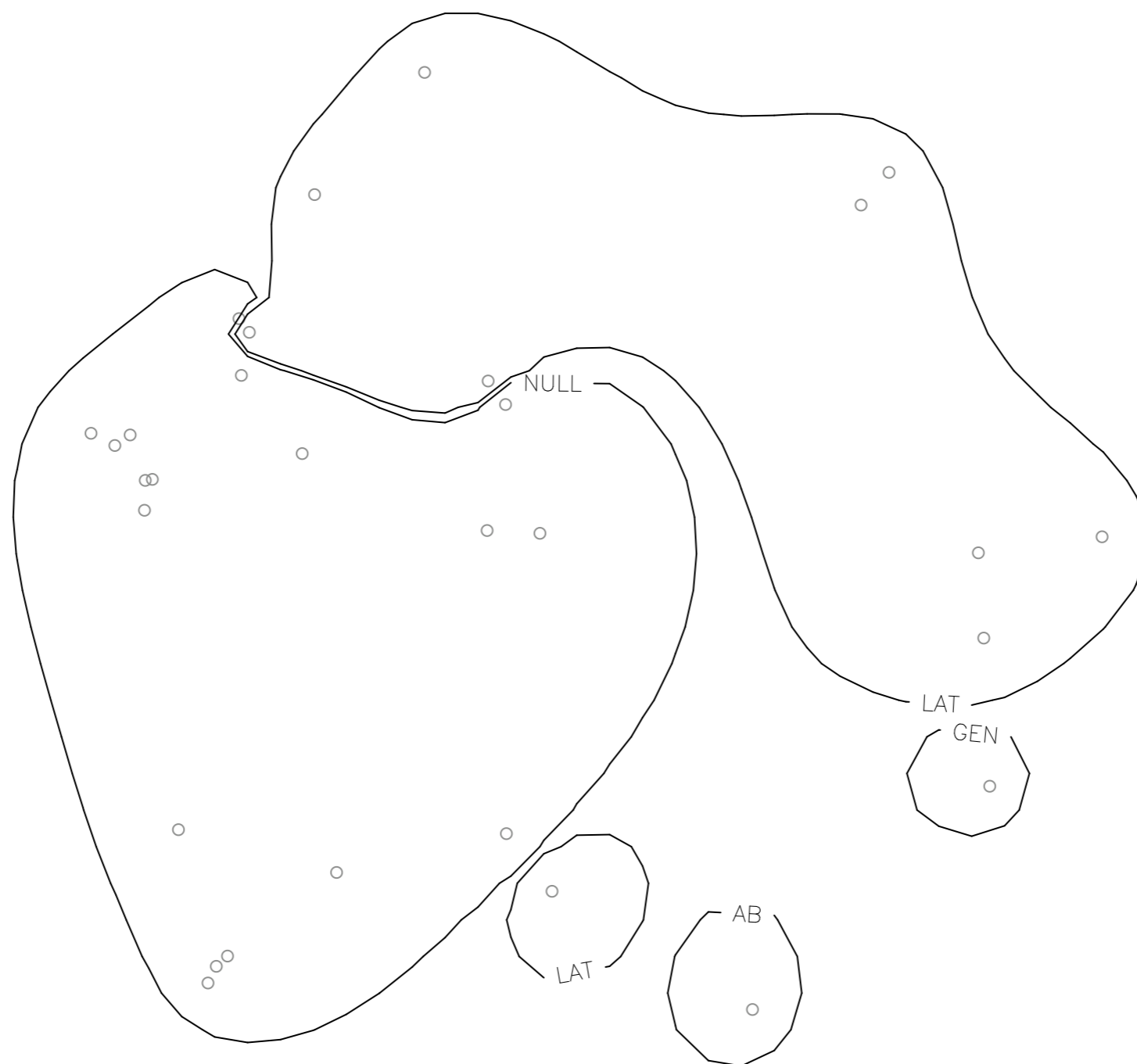
# Optimal maximum parsimony tree for case marking





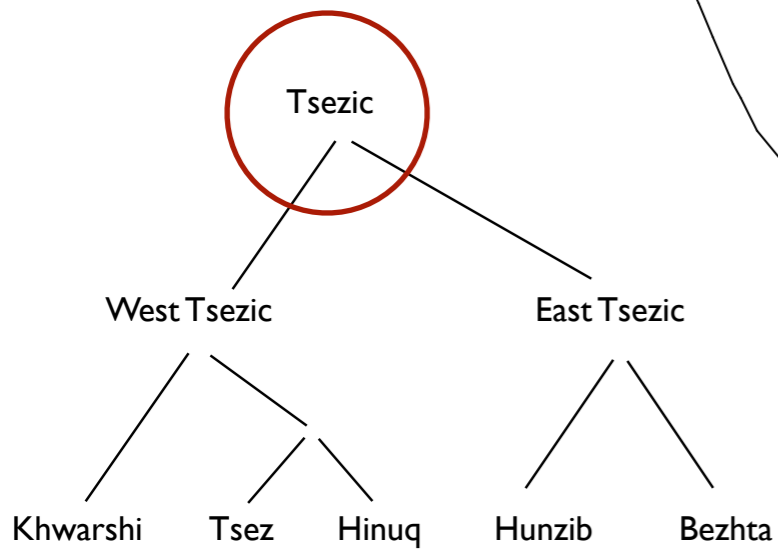
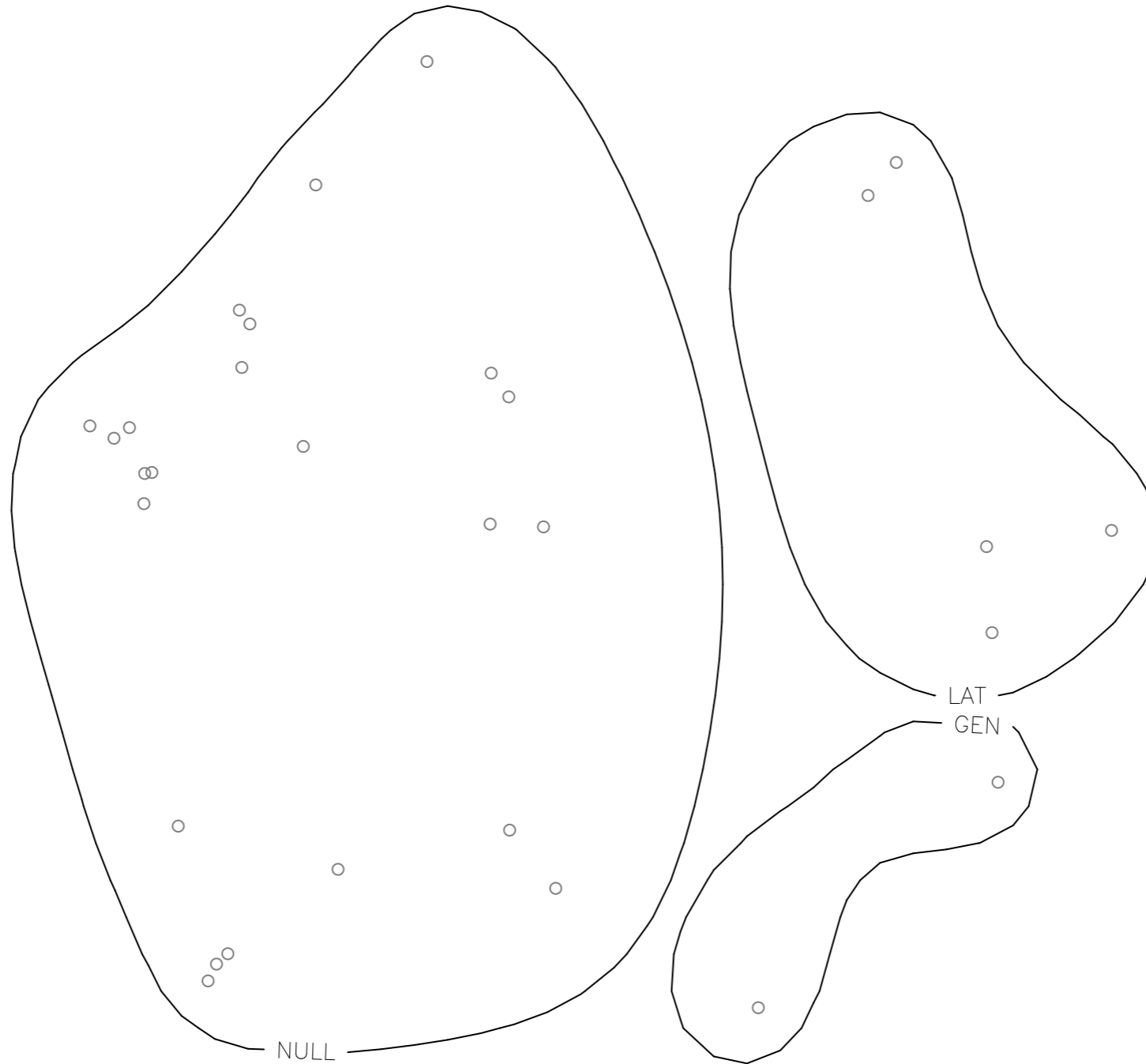
# Tsez

## Directional Case Marking



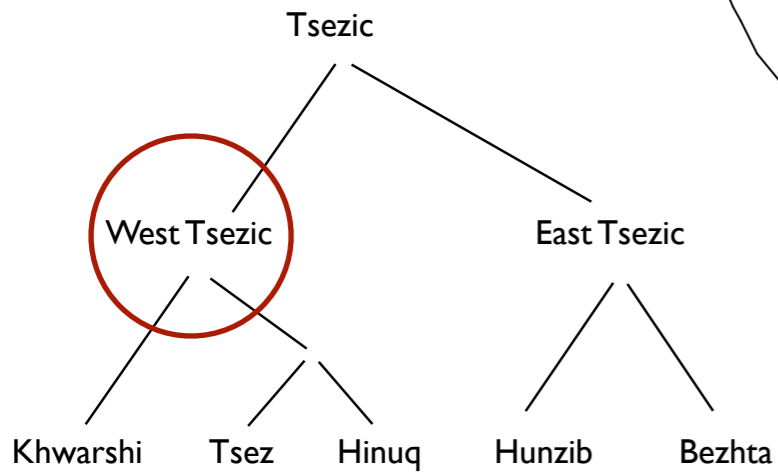
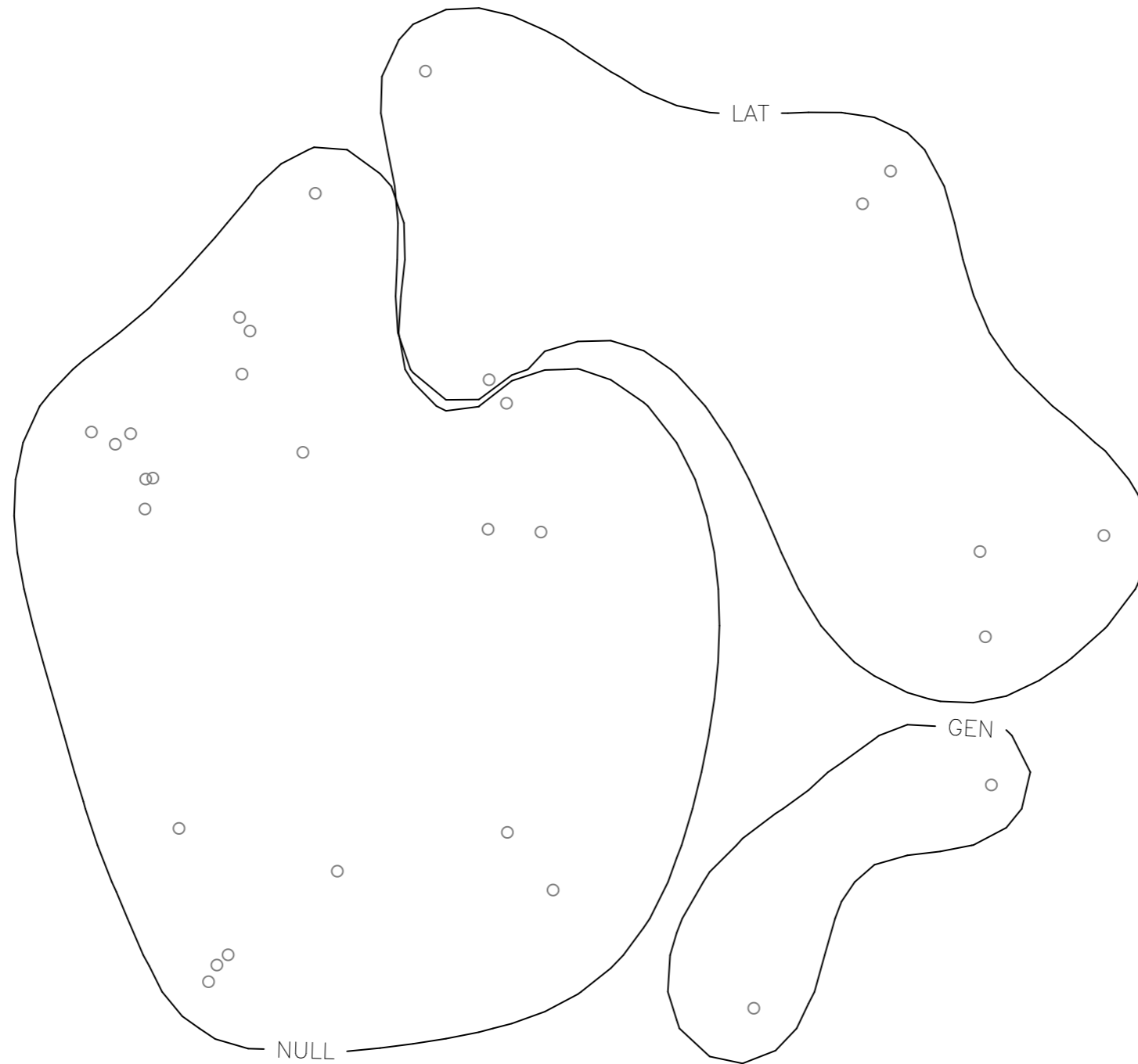
# Proto-Tsezic

“directional”



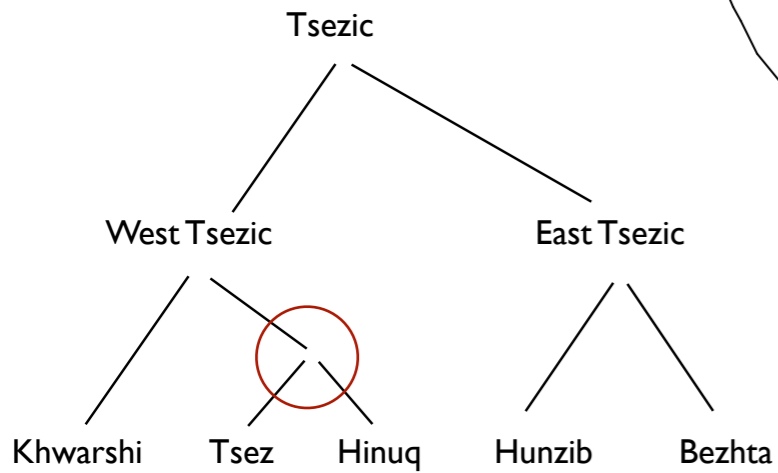
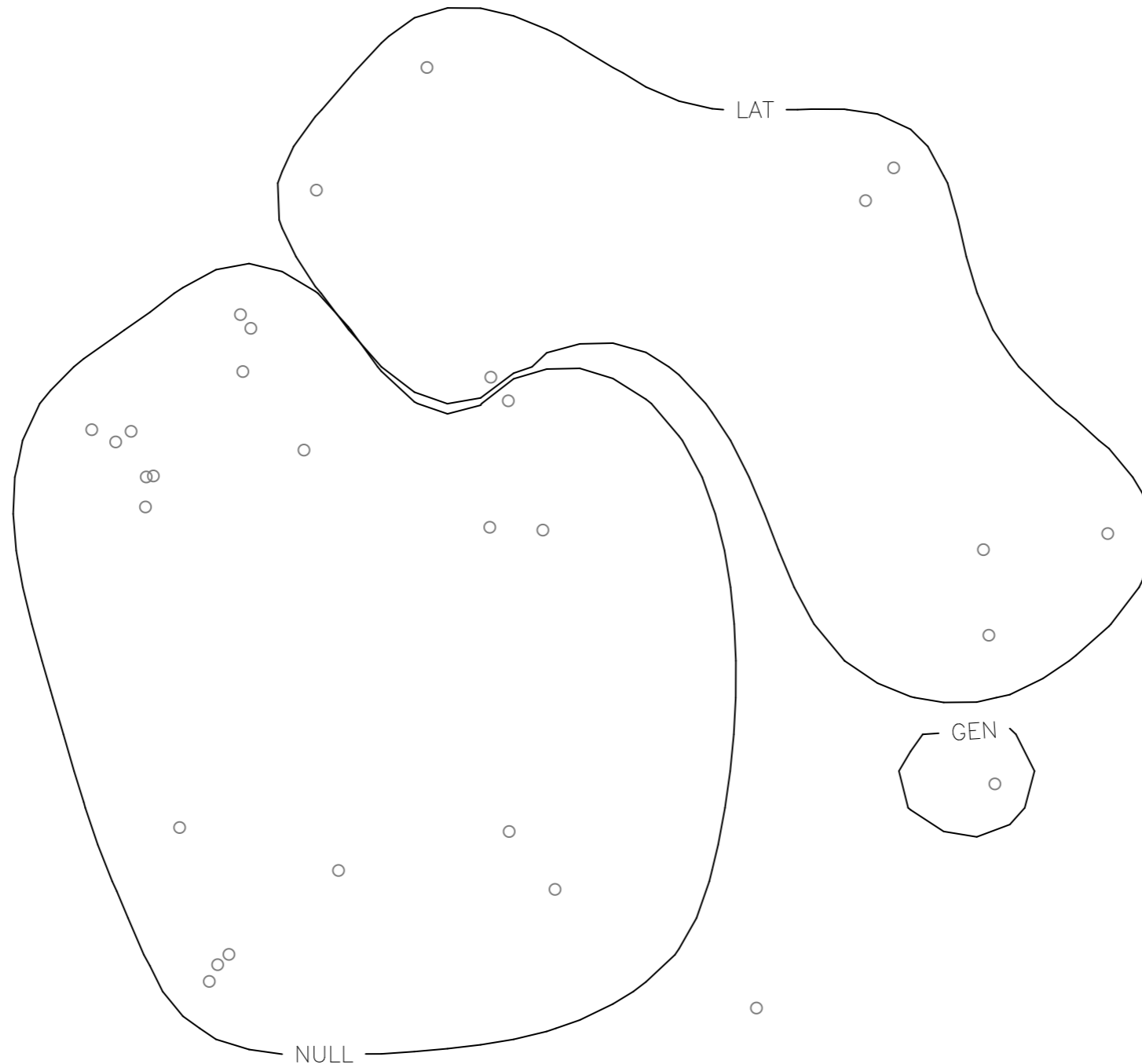
# Proto-West-Tsezic

“directional”



# Proto-Hinuq-Tsez

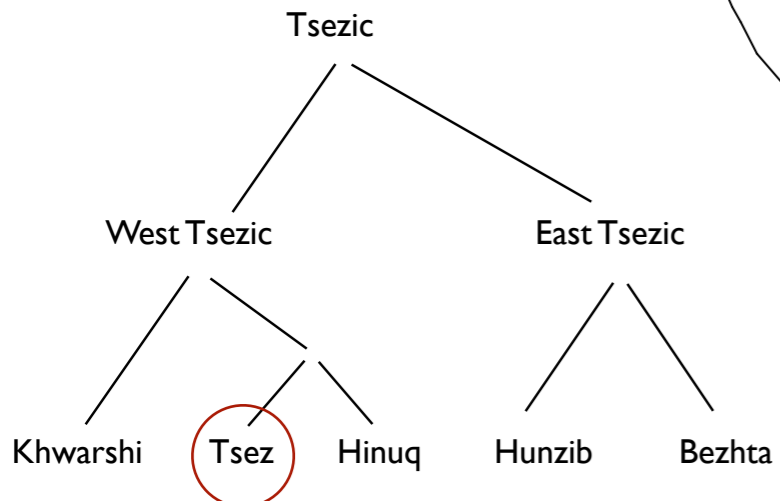
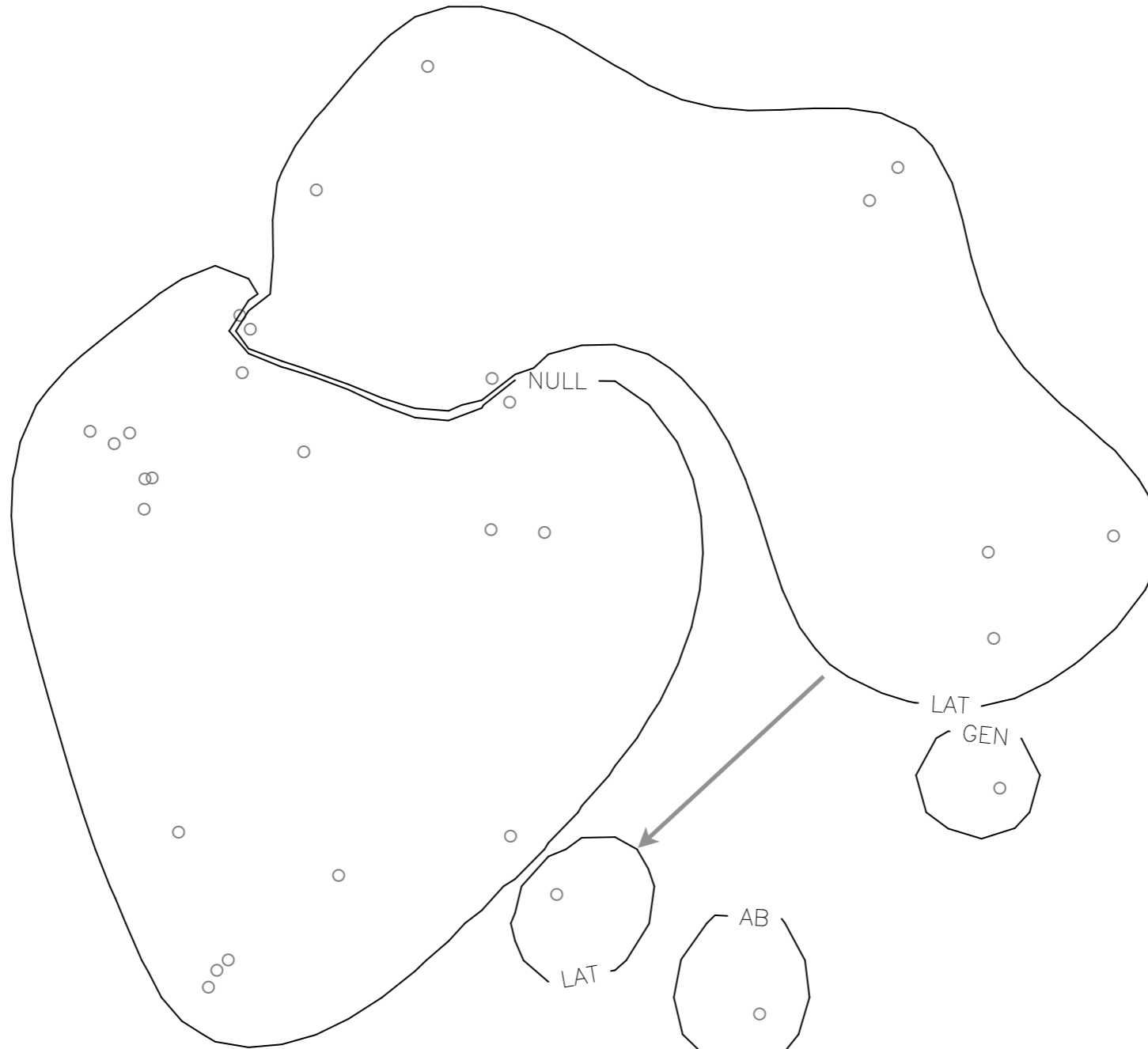
“directional”





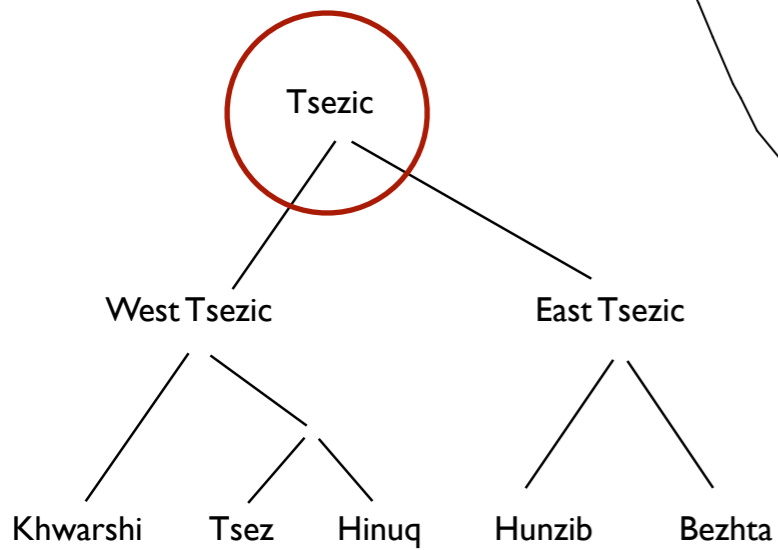
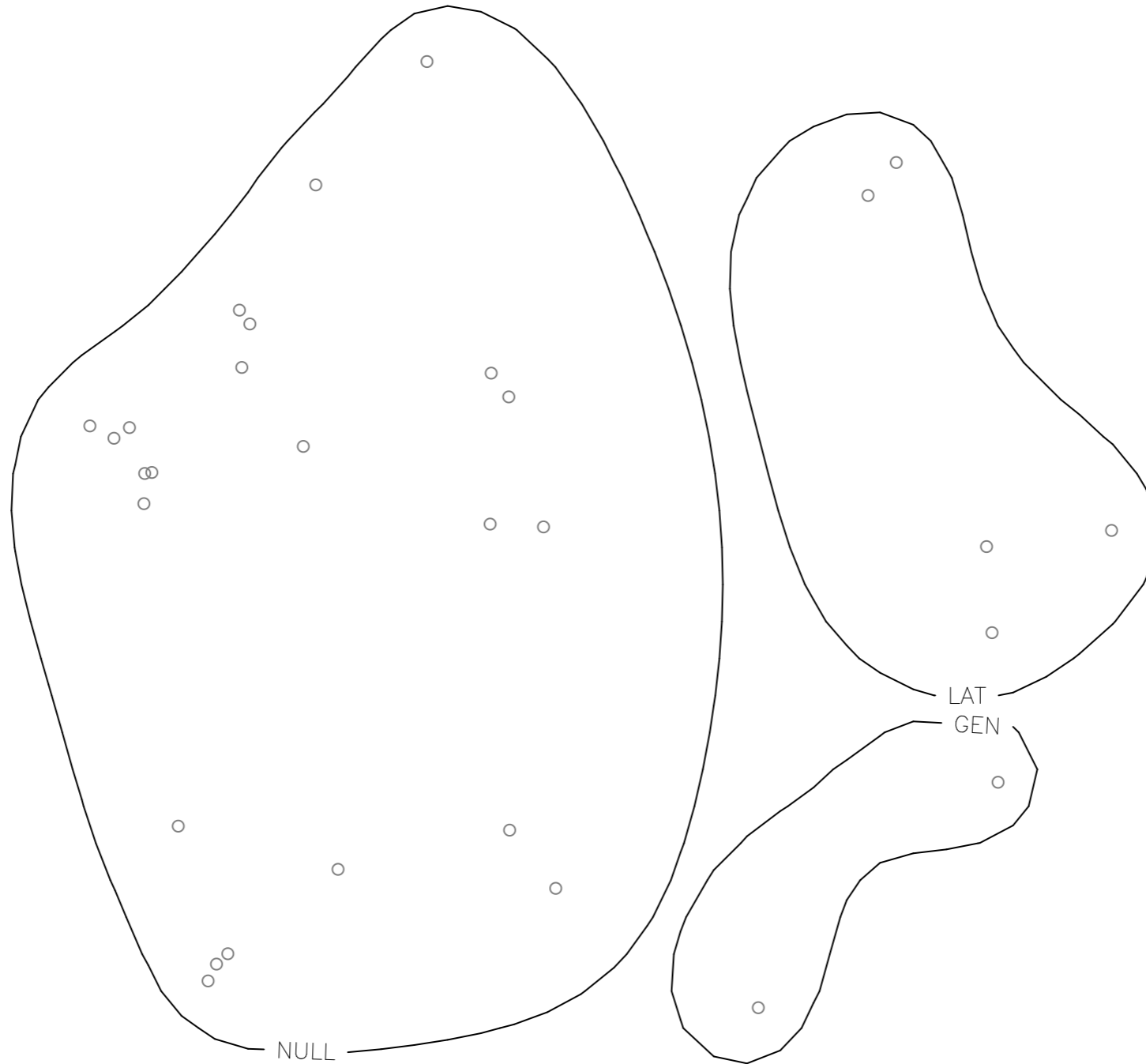
# Tsez

“directional”



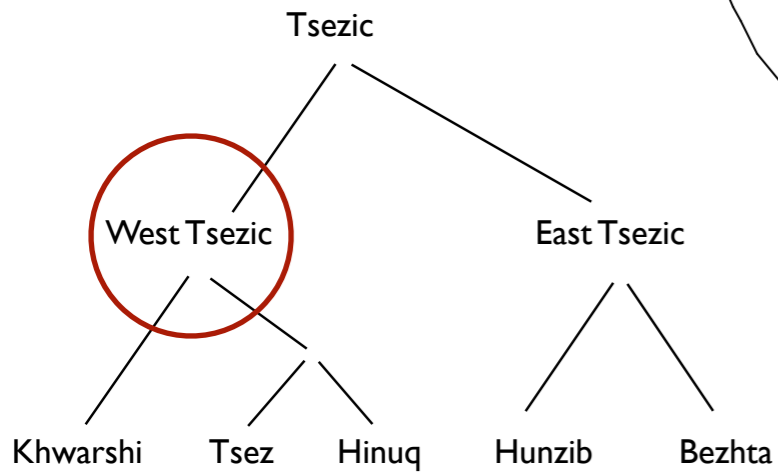
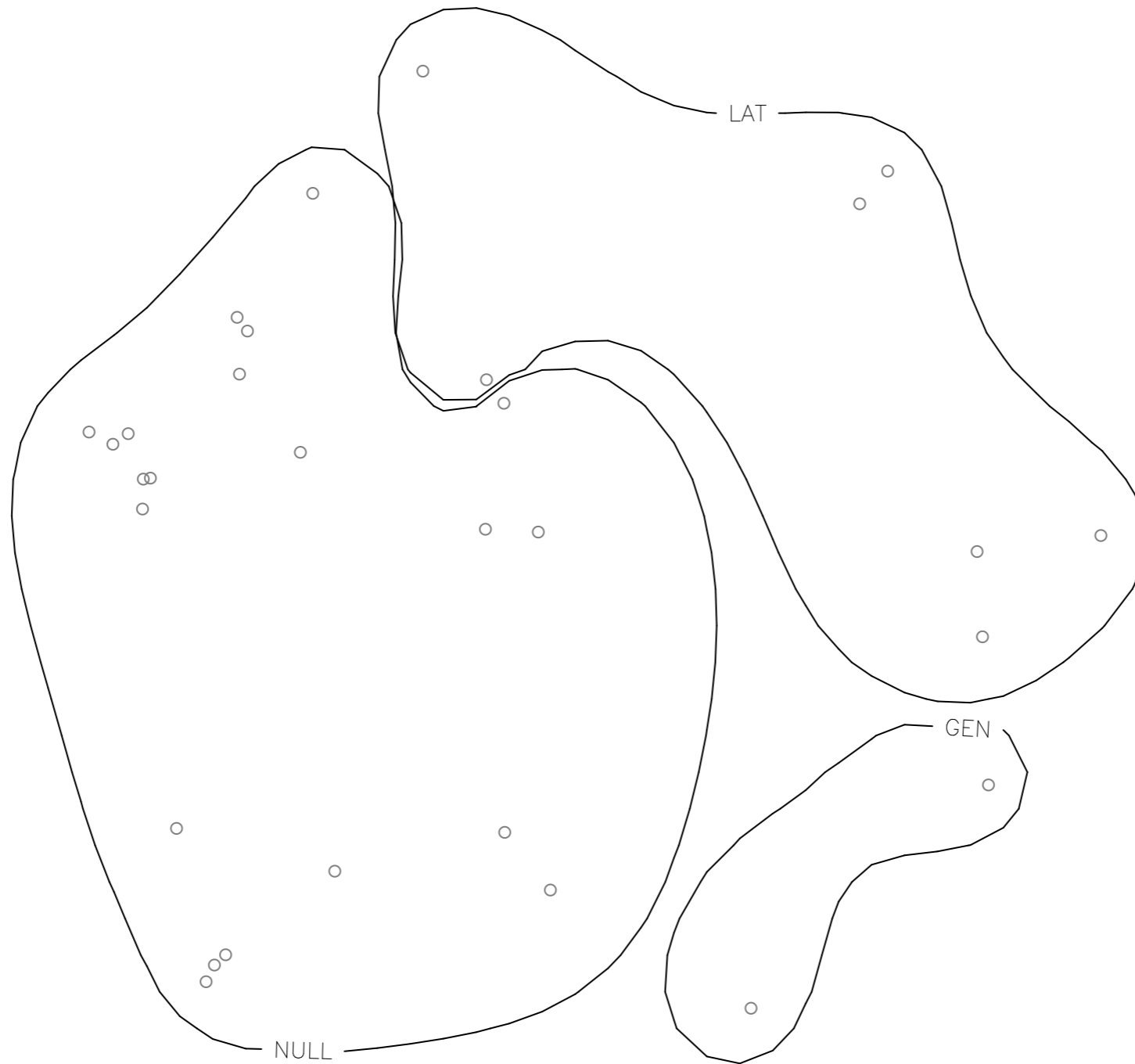
# Proto-Tsezic

“directional”



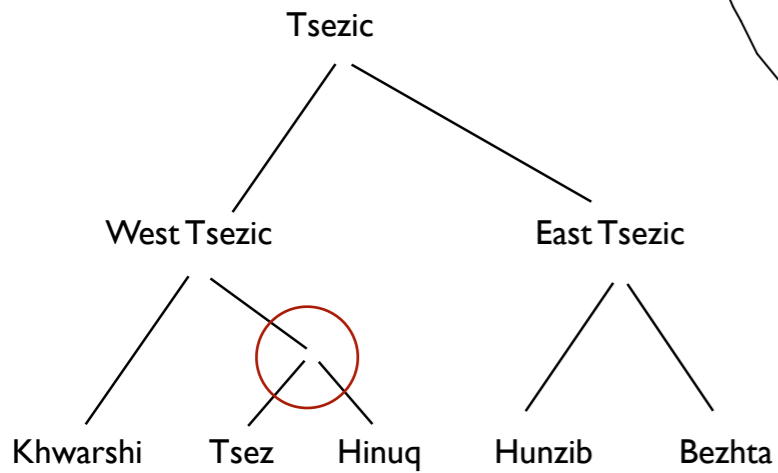
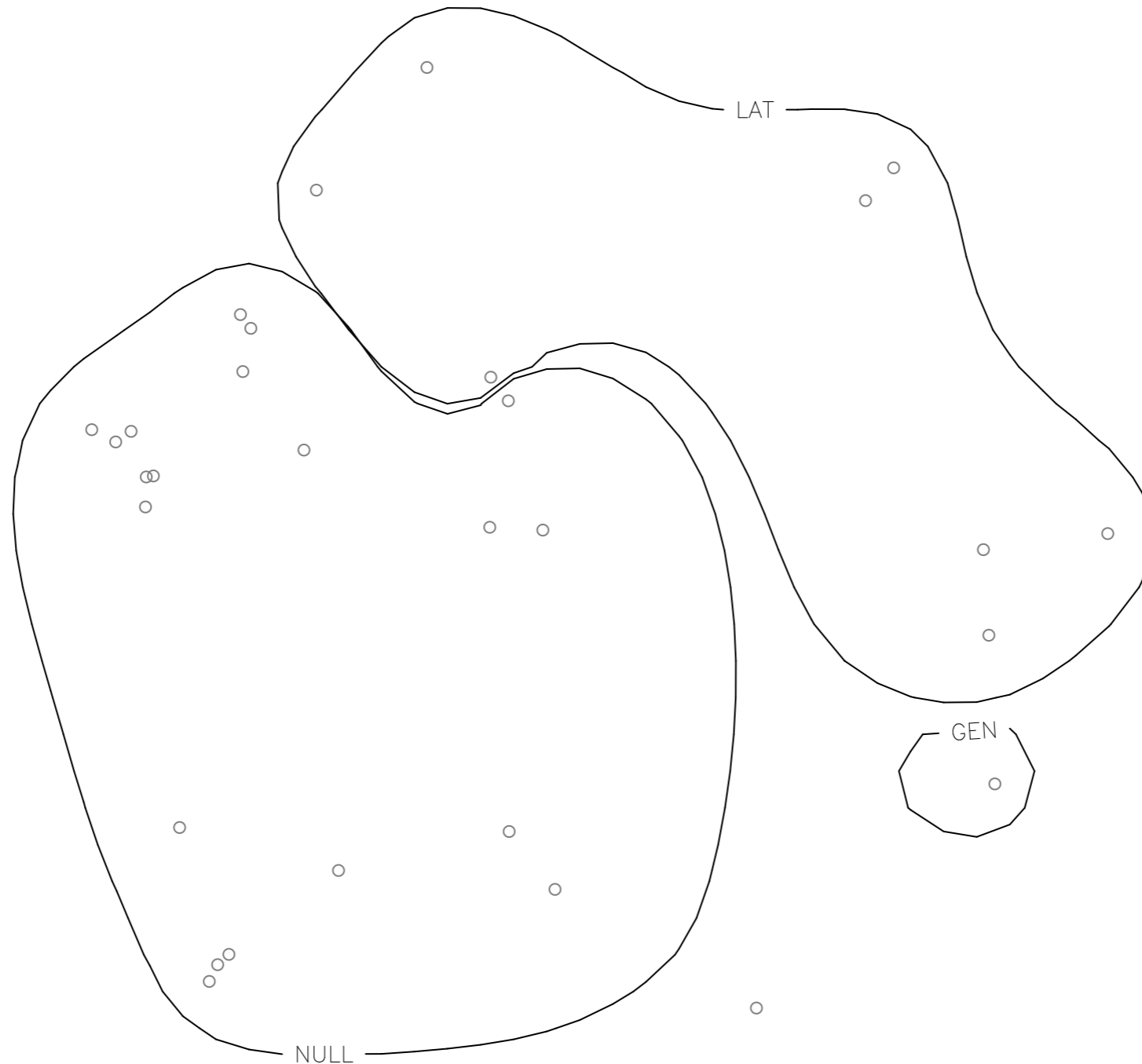
# Proto-West-Tsezic

“directional”



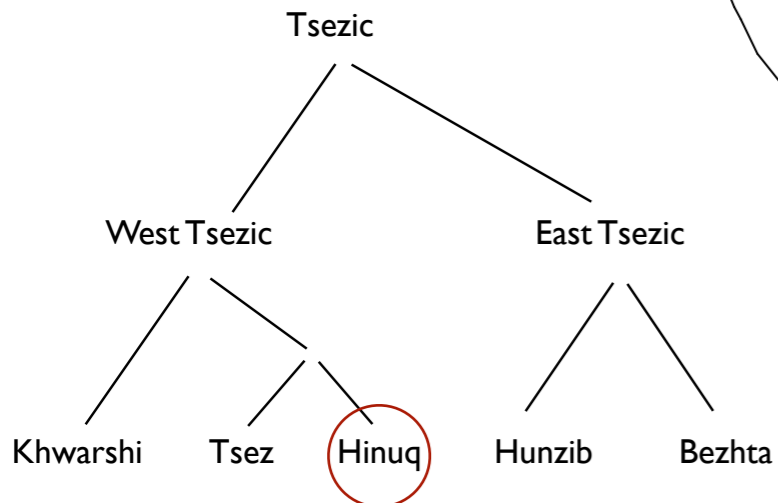
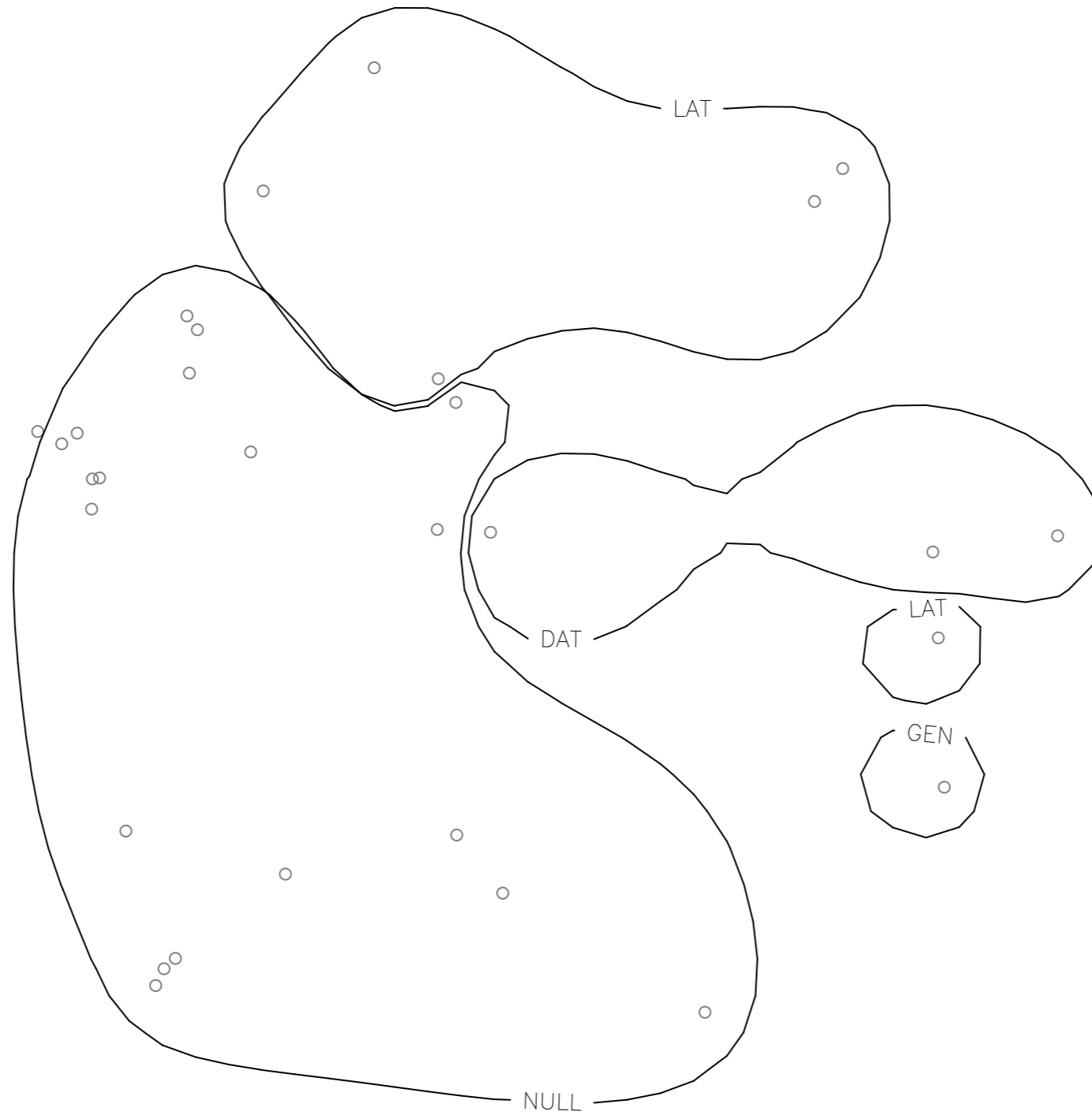
# Proto-Hinuq-Tsez

“directional”



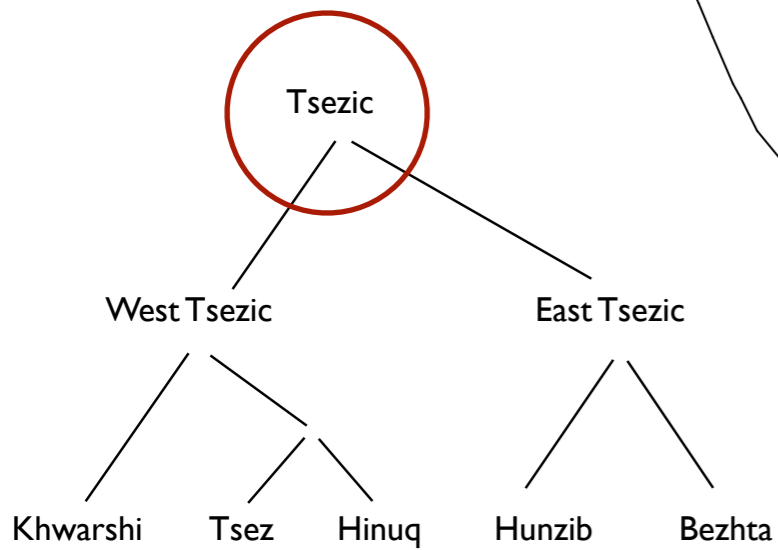
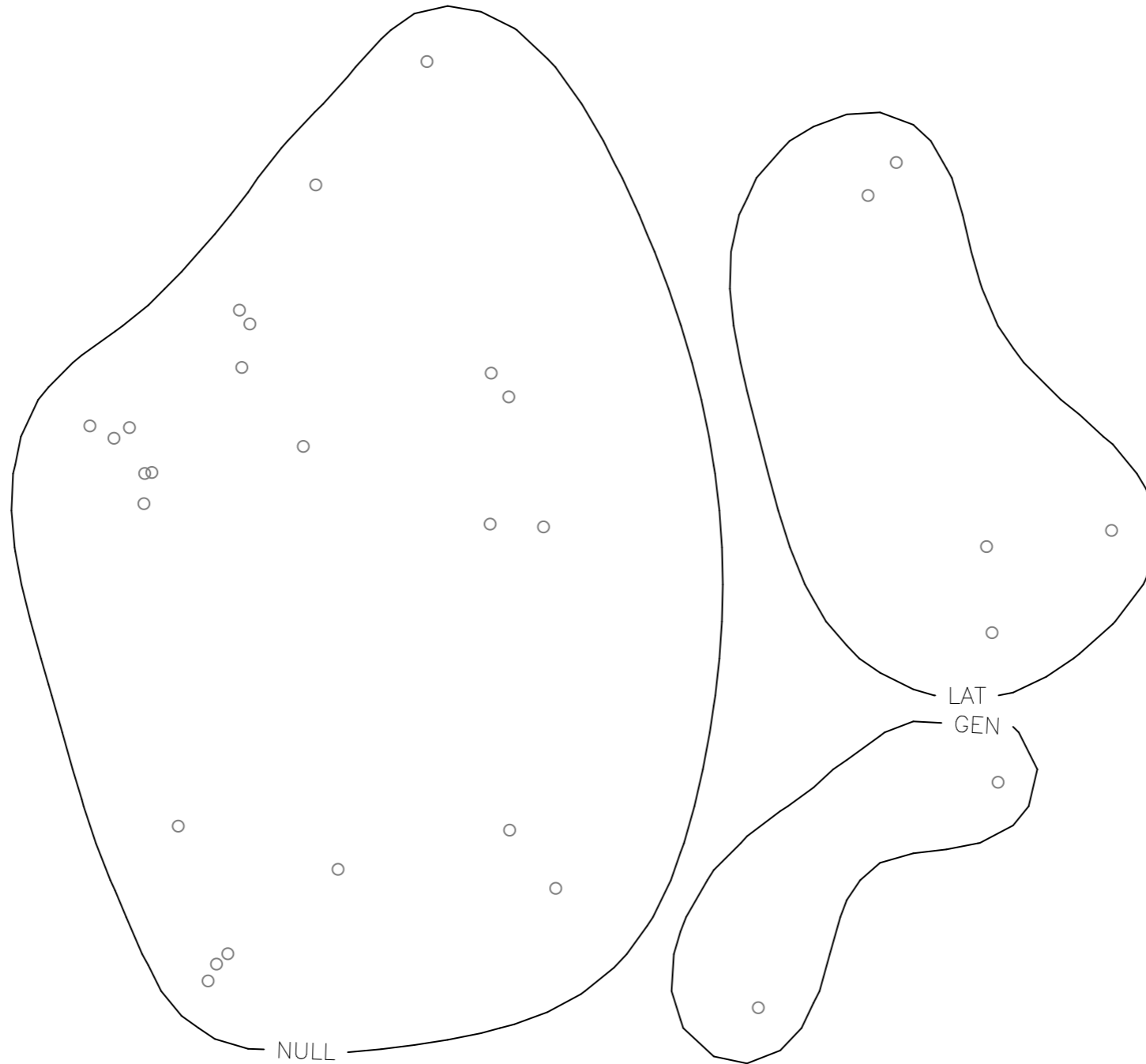
# Hinuq

“directional”



# Proto-Tsezic

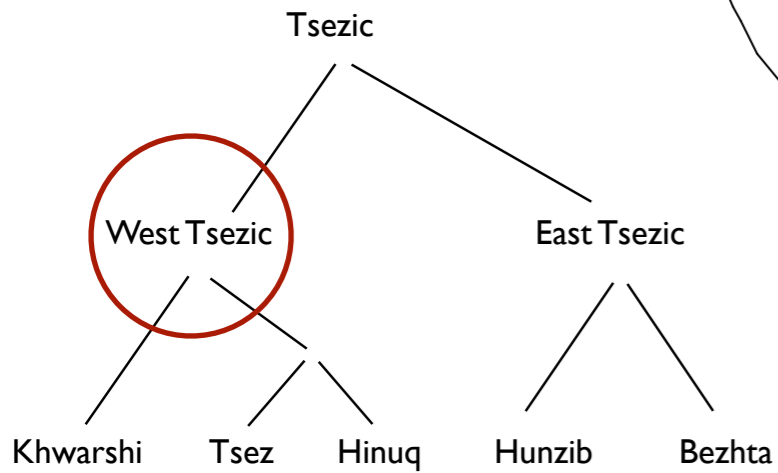
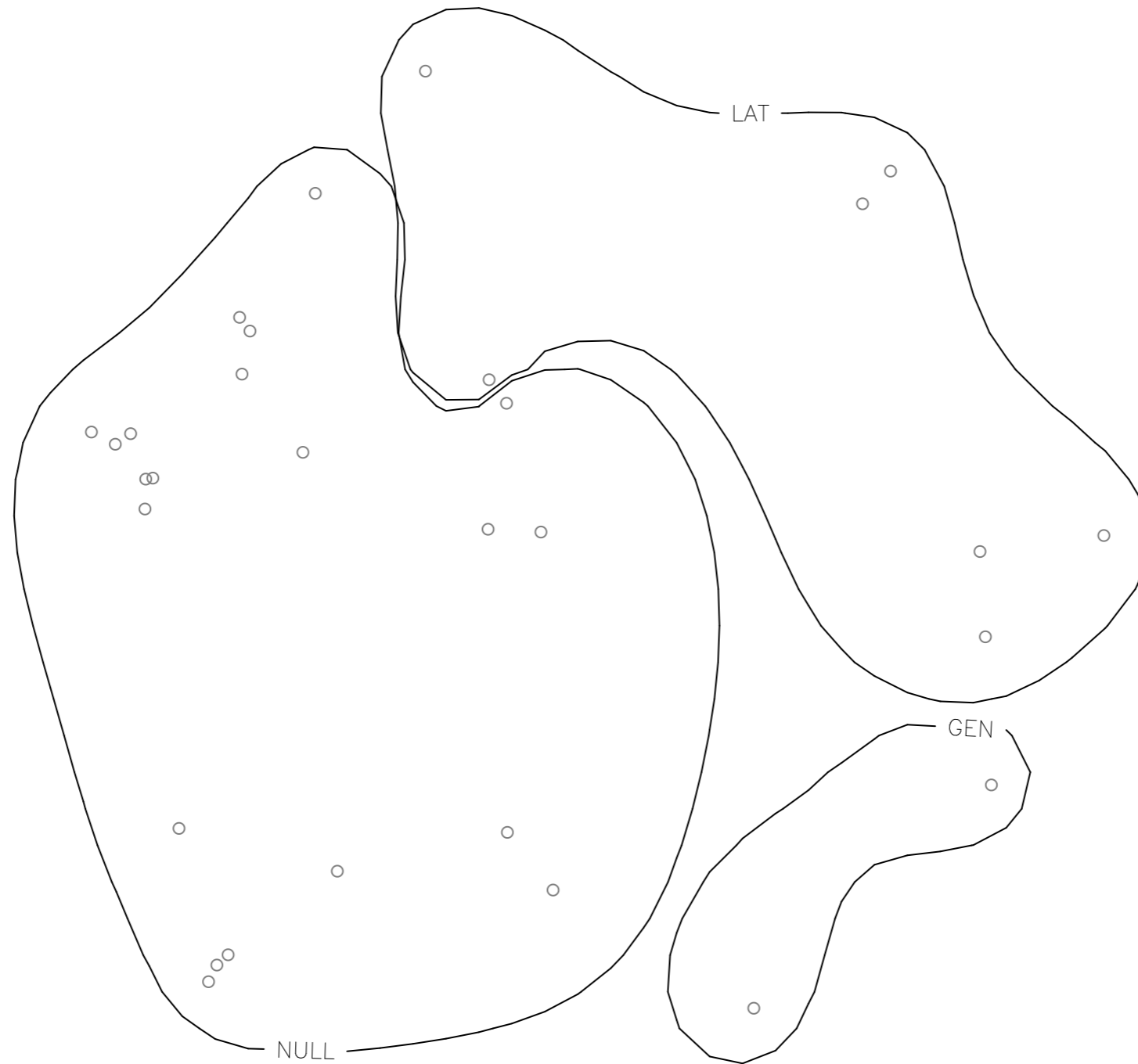
“directional”





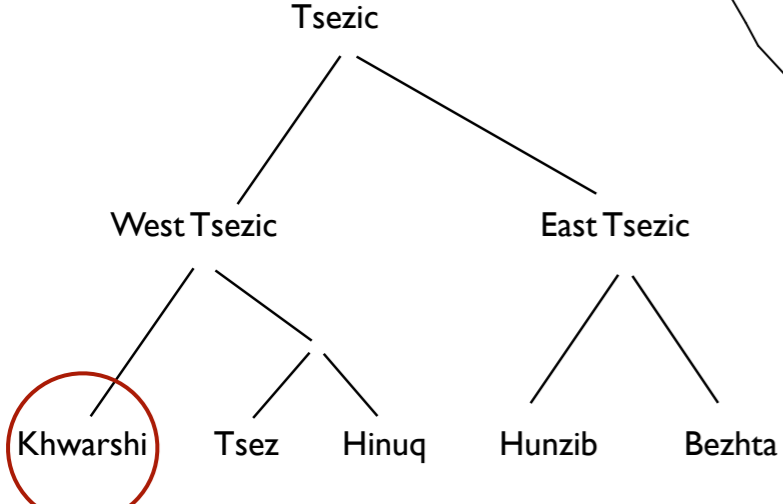
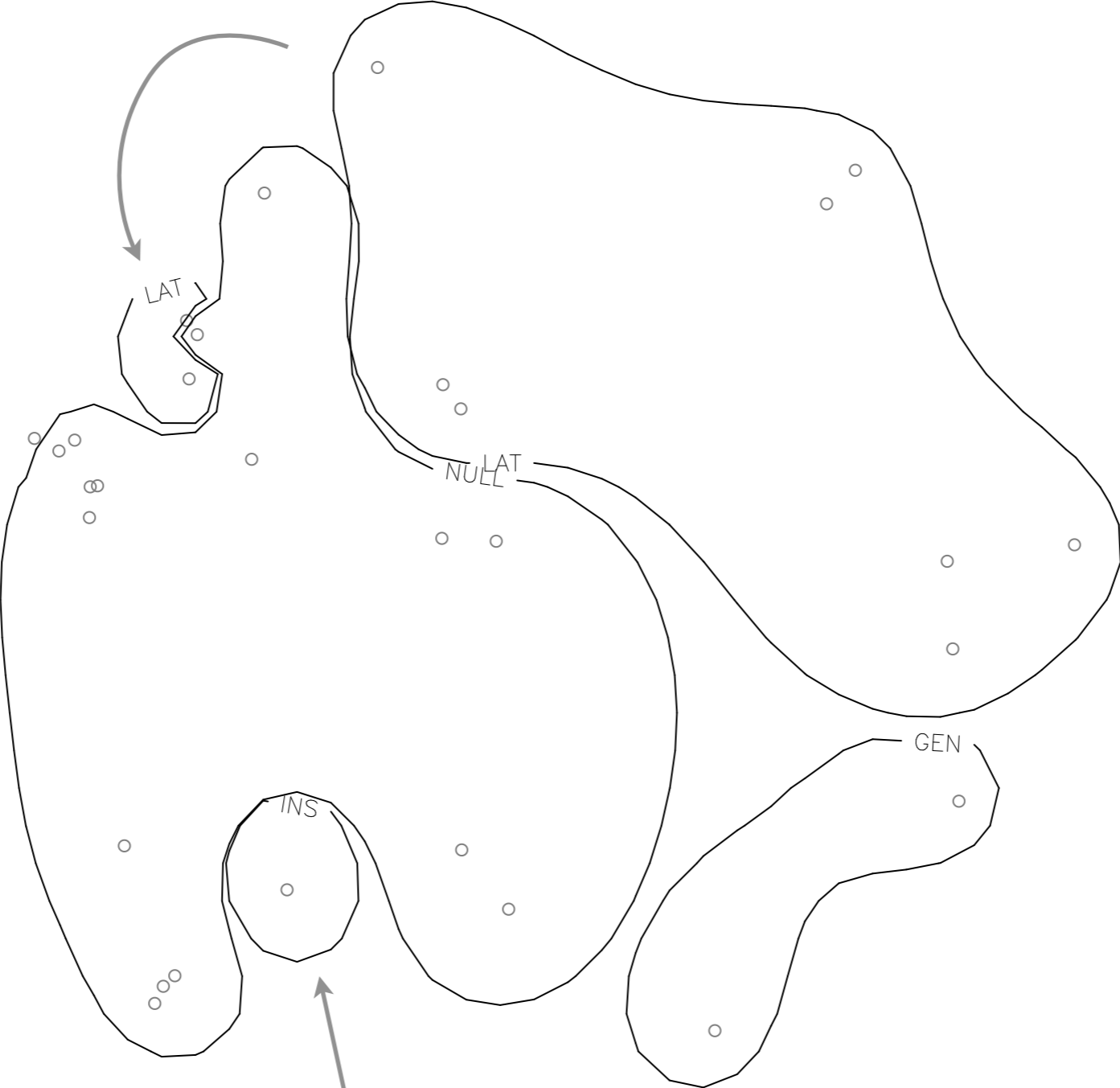
# Proto-West-Tsezic

“directional”



# Khwarshi

“directional”



1. **specific, known to the speaker**

‘*Somebody* called while you were away: guess who!’

2. **specific, unknown to the speaker**

‘I heard *something*, but I couldn’t tell what it was.’

3. **non-specific, irrealis**

‘Please try *somewhere* else.’

4. **polar question**

‘Did *anybody* tell you anything about it?’

5. **conditional protasis**

‘If you see *anything*, tell me immediately.’

6. **indirect negations**

‘I don’t think that *anybody* knows the answer.’

7. **direct negation**

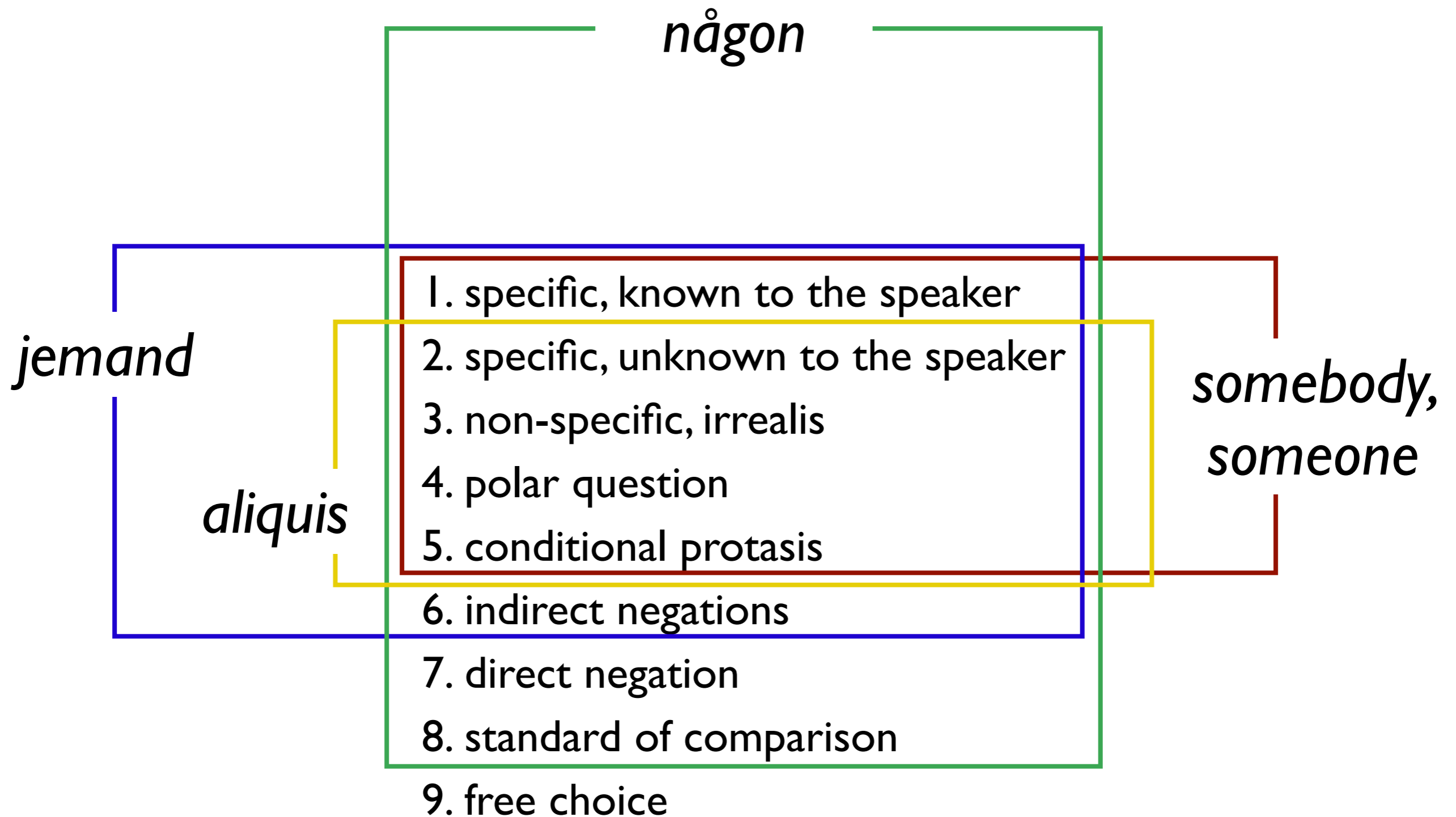
‘Nobody knows the answer.’

8. **standard of comparison**

‘In Freiburg, the weather is nicer than *anywhere* in Germany’

9. **free choice**

‘*Anybody* can solve this simple problem.’



*Ich glaube nicht, daß jemand die Antwort weiß*  
*I don't think that ~~somebody~~ anybody knows the answer*

# German indefinite pronouns (human only)

*jemand*

1. specific, known to the speaker
2. specific, unknown to the speaker
3. non-specific, irrealis
4. polar question
5. conditional protasis
6. indirect negations

*niemand*

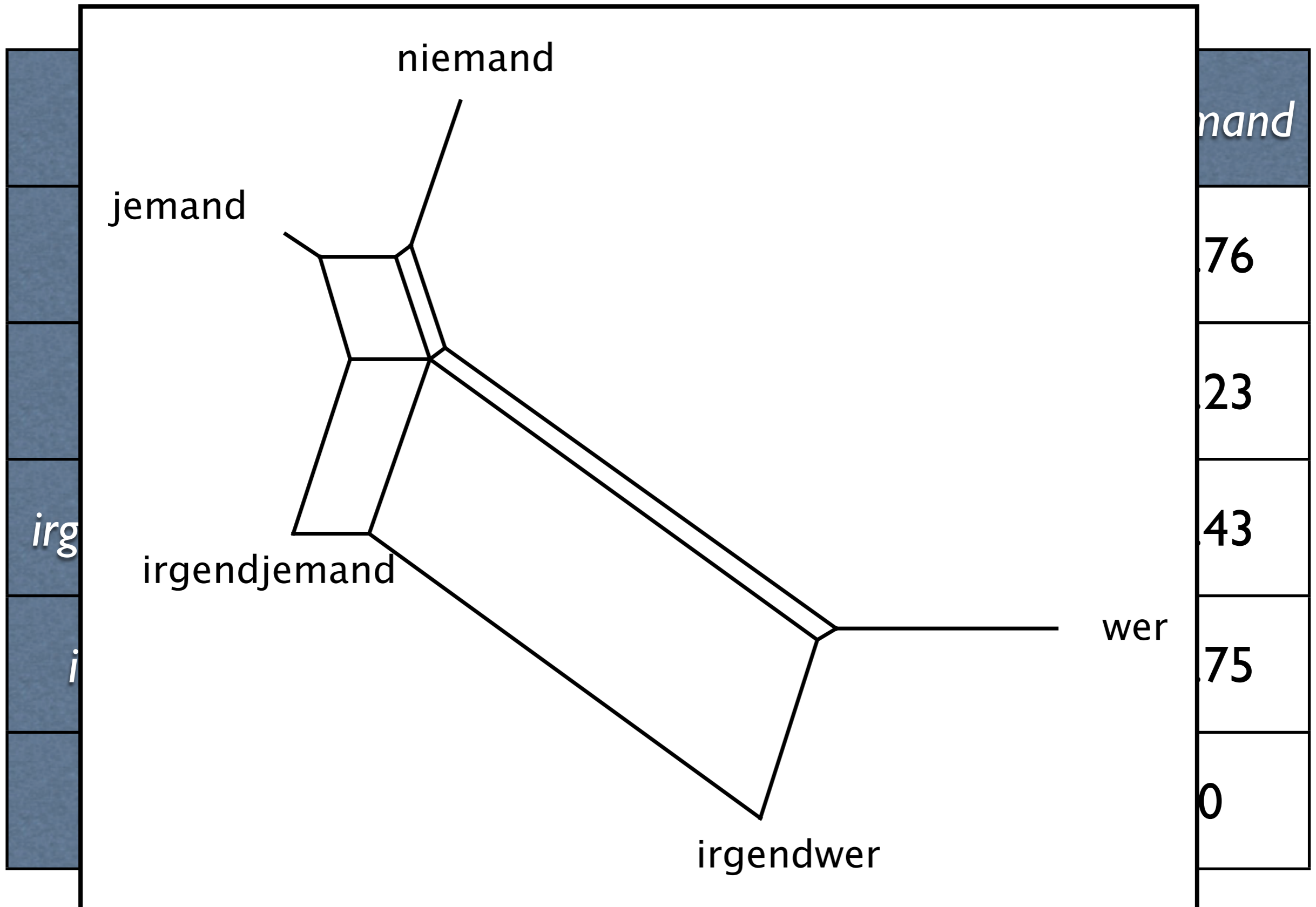
7. direct negation

8. standard of comparison

9. free choice

*irgentjemand,  
irgentwer*

# Language-specific dissimilarities



1. specific, known to the speaker
2. specific, unknown to the speaker
3. non-specific, irrealis
4. polar question
5. conditional protasis
6. indirect negations
7. direct negation
8. standard of comparison
9. free choice

	1	2	3	4	5	6	7	8	9
<i>jemand</i>	10	45	7	4	19	2			
<i>irgendjemand</i>		3	45	7	18	9		2	12
<i>irgendwer</i>		14	36	12	8	45		35	1
<i>niemand</i>							20		

types perspective

	<i>jemand</i>	<i>irgend-jemand</i>	<i>irgent-wer</i>	<i>niemand</i>
<i>jemand</i>	0	0.25	0.72	0.23
<i>irgendjemand</i>	0.25	0	0.42	0.43
<i>irgendwer</i>	0.72	0.42	0	0.75
<i>niemand</i>	0.23	0.43	0.75	0

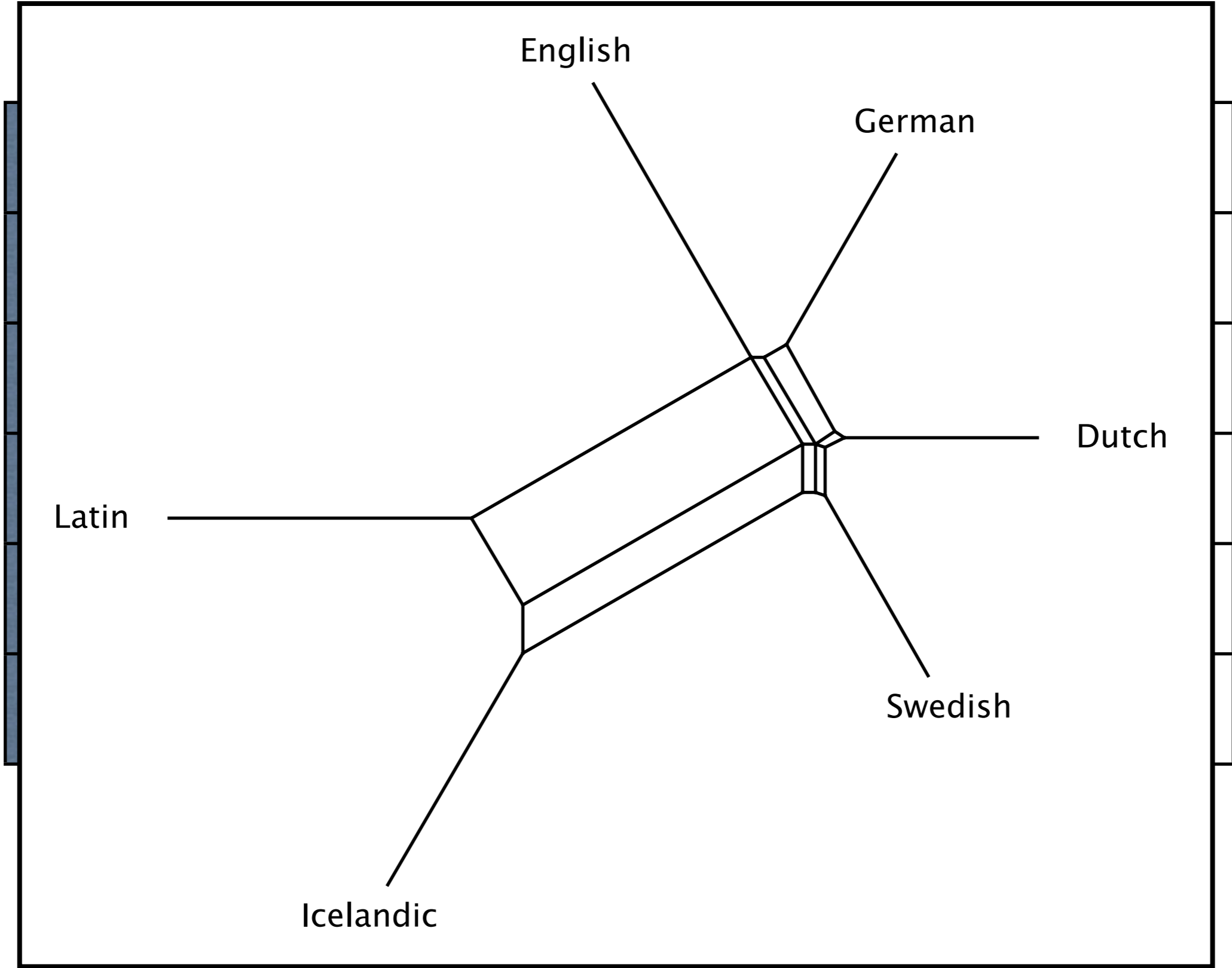
	1	2	3	4	5	6	7	8	9
<i>jemand</i>	X	X	X	X	X	X			
<i>irgendjemand</i>		X	X	X	X	X		X	X
<i>irgendwer</i>		X	X	X	X	X		X	X
<i>niemand</i>							X		

	1	2	3	4	5	6	7	8	9
1	0	0.36	0.36	0.36	0.36	0.36	0.37	0.57	0.57
2	0.36	0	0	0	0	0	0.55	0.09	0.09
3	0.36	0	0	0	0	0	0.55	0.09	0.09
4	0.36	0	0	0	0	0	0.55	0.09	0.09
5	0.36	0	0	0	0	0	0.55	0.09	0.09
6	0.36	0	0	0	0	0	0.55	0.09	0.09
7	0.37	0.55	0.55	0.55	0.55	0.55	0	0.69	0.69
8	0.57	0.09	0.09	0.09	0.09	0.09	0.69	0	0
9	0.57	0.09	0.09	0.09	0.09	0.09	0.69	0	0

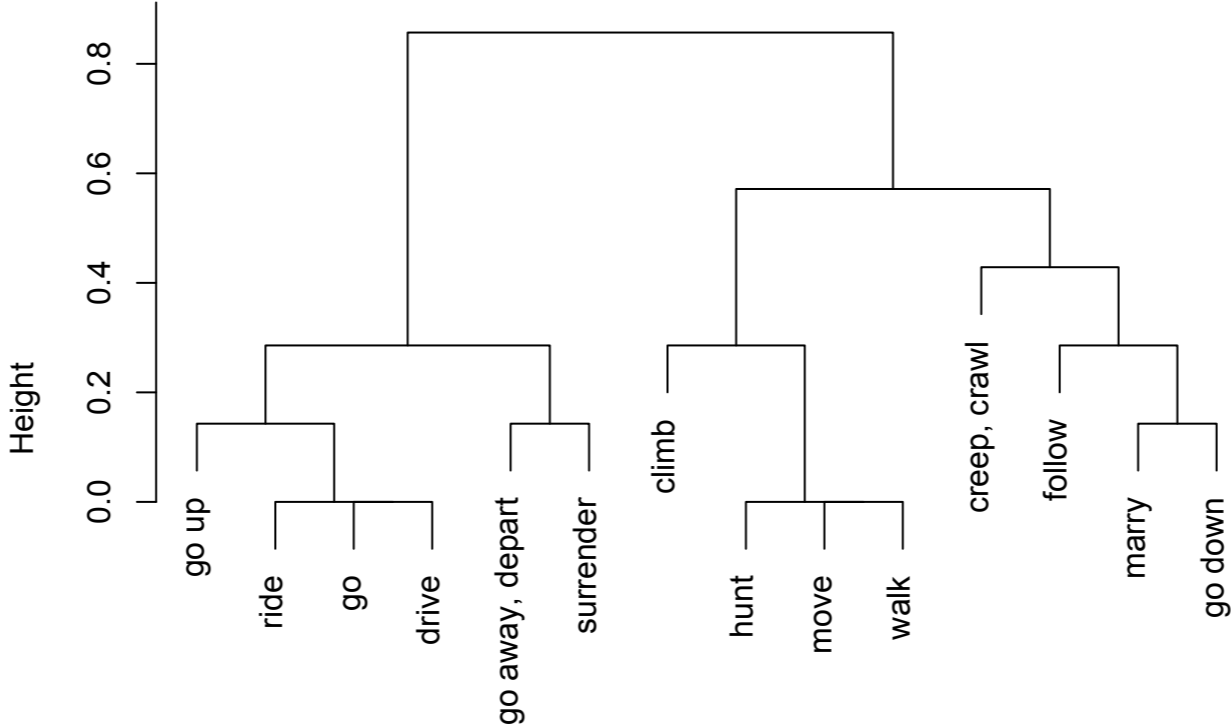
**Language-specific perspective  
on sampled functions**



correlate language-specific perspectives  
with each other ...



### Cluster Dendrogram



```
as.dist(d[freq > 1, freq > 1])  
hclust (*, "complete")
```

