Measuring Language Similarities

Michael Cysouw

Geneling, 17 Juli 2006

MAX-PLANCK-GESELLSCHAFT

Preamble

- Measuring language similarities is independent of their interpretation
  - Genealogical Relationship ?
  - Borrowing ?
  - Human Language Universal ?
- It is a separate research questions which similarities should have what interpretation

Method of comparison depends on datastructure

	Primary data Elements of languages	**Secondary data** Statements about languages
	Utterances	Lexemes	Descriptions
Monolingual data	Text	Dictionary	Grammar

Measurements on Monolingual Datasources

- Textcounts, e.g.:
  - Greenberg's morphological indices
  - Givón's anaphorical distance
  - Bickel's referential density
- Gradient Grammatical Indices, e.g.:
  - Maddieson's number of consonants
  - Bickel's index of inflectional synthesis
- Dictionary counts ?

Method of comparison depends on datastructure

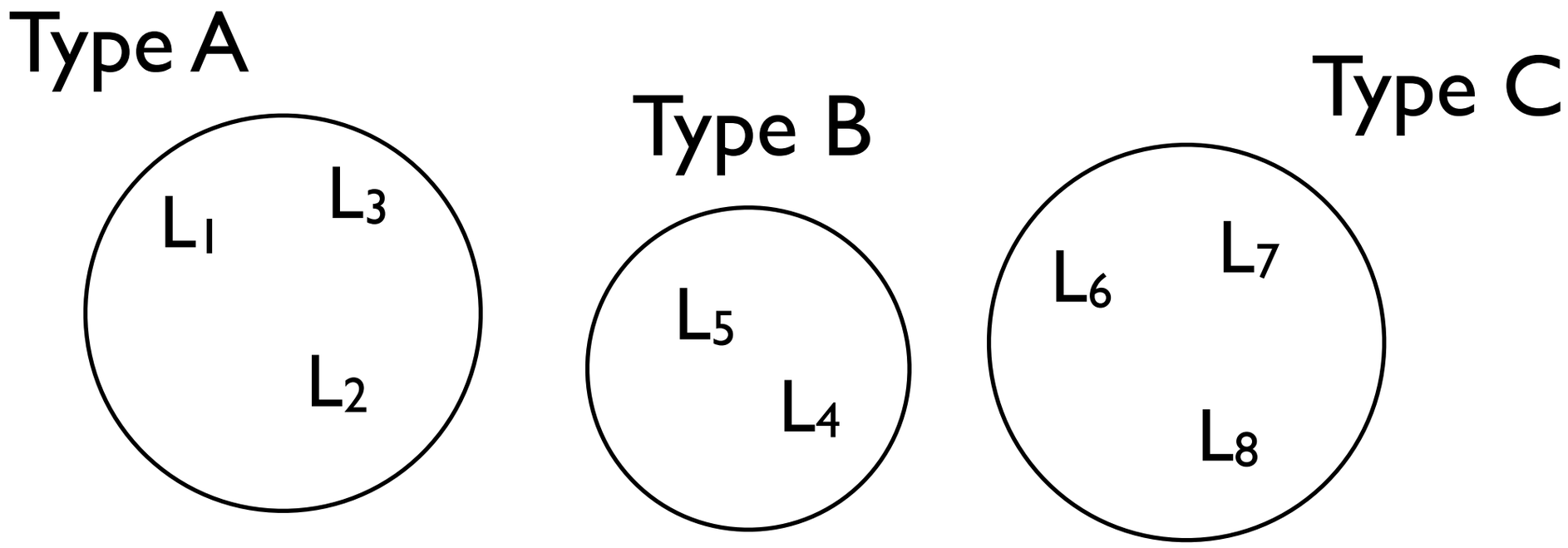
	Primary data Elements of languages	**Secondary data** Statements about languages
	Utterances	Lexemes	Descriptions
Monolingual data	Text	Dictionary	Grammar

Method of comparison depends on datastructure

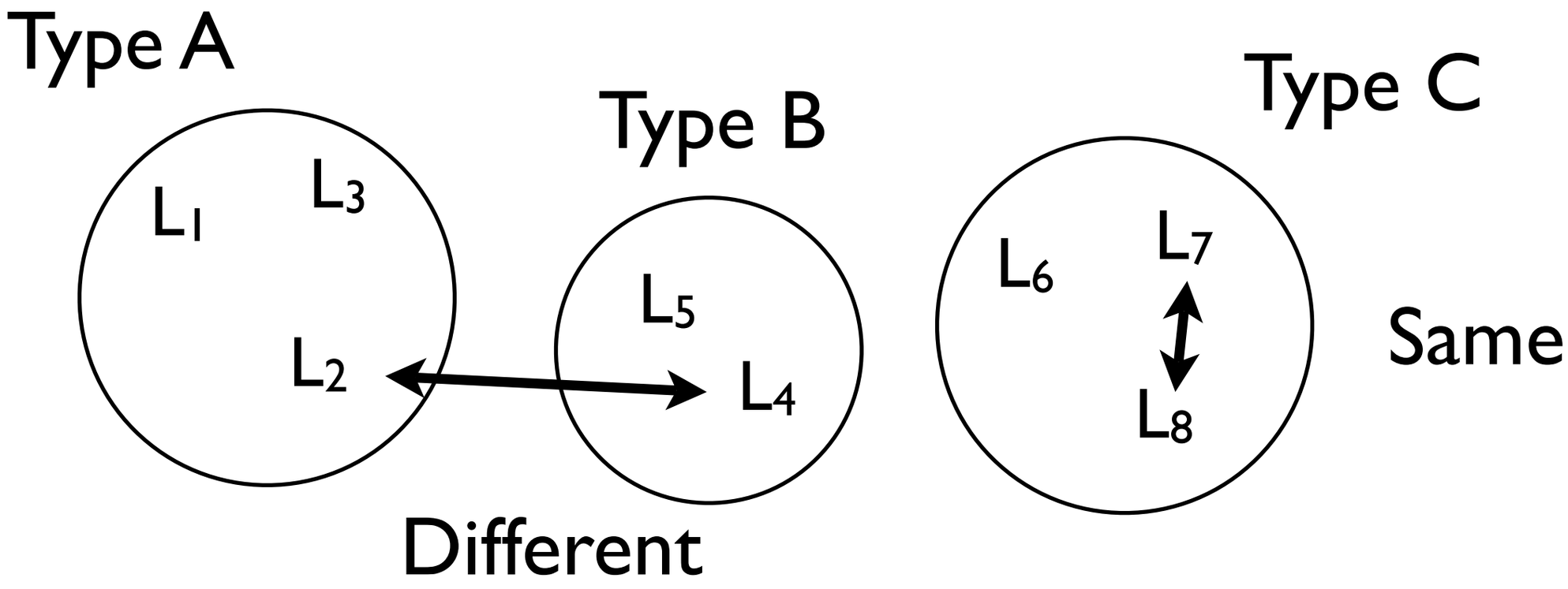
	Primary data Elements of languages	**Secondary data** Statements about languages
	Utterances	Lexemes	Descriptions
Monolingual data	Text	Dictionary	Grammar
Comparative data	Contextually situated exemplars	Wordlists	Typological databases

Typological Databases

Typology: Grouping of Languages into Types



Little Information …



We want more fine-grained measurements of similarity !

Solution: Use many Typologies (think: WALS)

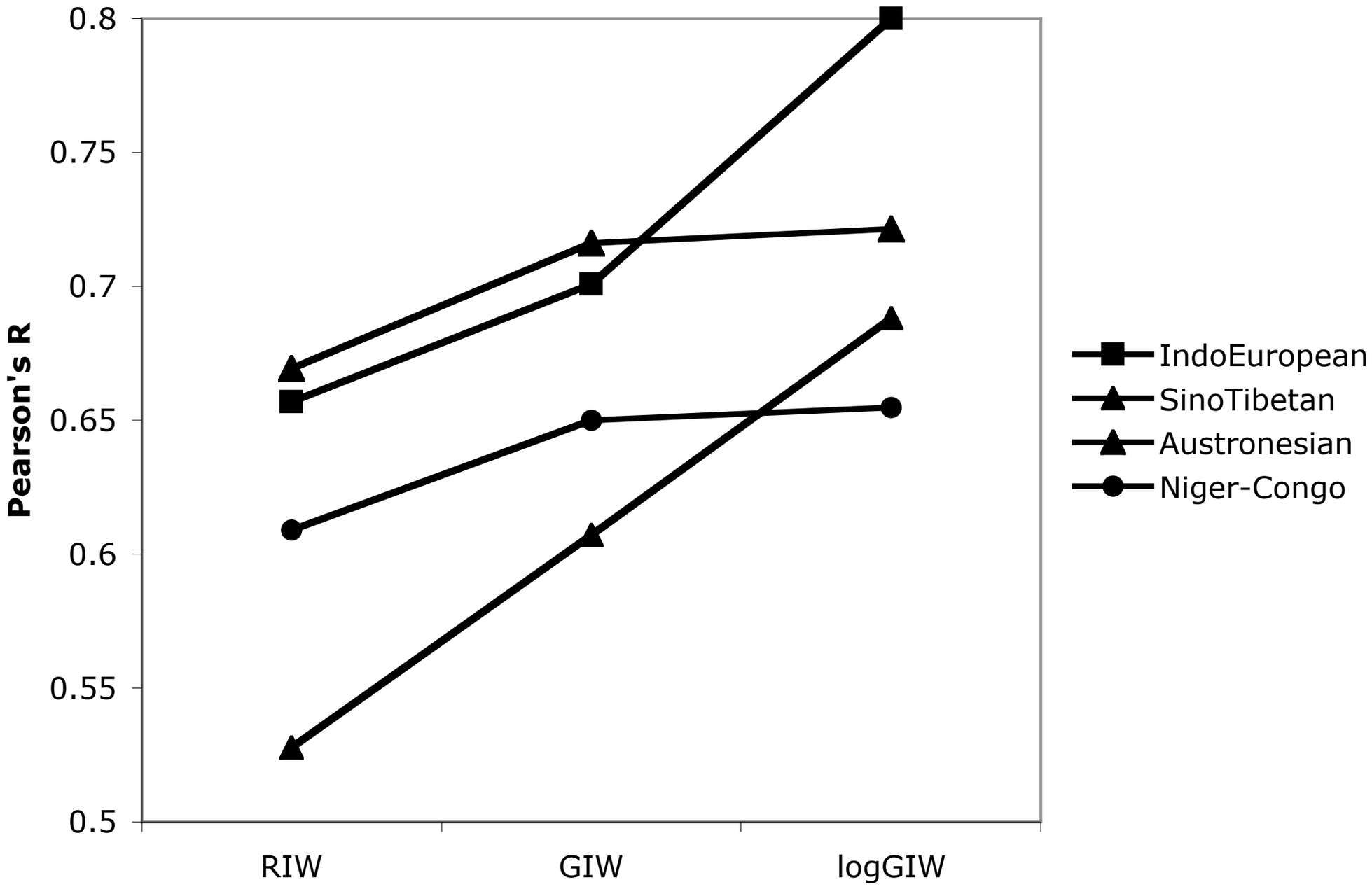
- Measures of similarity between L_1 and L_2 for a set of W typologies:
  - $S(L_1,L_2)$ = # Similarities
  - $S(L_1,L_2)$ = W – # Differences
  - $S(L_1,L_2)$ = $^{\text{\# Similarities}}/_{\text{\# Data available for both}}$
  - $S(L_1,L_2)$ = $^{\text{\# Similarities}}/_{\text{\# Similarities + \# Differences}}$

  - 'Relativer Identitätswert' (Goebl 1984)

Improvement for Similarities

- Unusual types are more indicative of similarity:
  - Instead of counting every similarity as '1'
  - use: 1-(fraction having this type)
  - 'Gewichteter Identitätswert' (Goebl 1984)
  - This idea is related to statistical information
  - use: -log(fraction having this type)

Specifying Similarities



Improving Differences

- $S(L_1,L_2) = {}^{\#\text{ Similarities}}/_{\#\text{ Similarities} + \#\text{ Differences}}$

- Instead of counting every differences as ‘1’, take into account that some types are more similar to each other than others

WALS – World Atlas of Language Structures

LANGUAGE VIEWER

COMPOSER

WALS the Feature Viewer

SHOW MAP

select a feature

- thematically
- alphabetically
- user-defined

SHRINK LIST

search for a feature

51

SEARCH

FEATURE PROFILE area: Nominal Categories

51. Position of Case Affixes

Author: **Matthew S. Dryer**

934 languages

DESCRIPTION

symbol	include	click to list languages below [no. of lgs : of genera : of families]	Merge: 1.	2.	3.	4.	5.	6.	7.	8.	9.
●	■	1. Case suffixes [431:174:90]	■	□	□	□	□	□	□	□	□
●	■	2. Case prefixes [35:19:14]		■	□	□	□	□	□	□	□
●	■	3. Case tone [4:2:1]			■	□	□	□	□	□	□
●	■	4. Case stem change [2:1:1]				■	□	□	□	□	□
●	■	5. Mixed morphological case [8:7:6]					■	□	□	□	□
◆	■	6. Postpositional clitics [95:59:36]						■	□	□	□
◆	■	7. Prepositional clitics [15:10:8]							■	□	□
◆	■	8. Inpositional clitics [6:3:1]								■	□
●	■	9. No case affixes or adpositional clitics [338:145:56]									■

arrange the languages by

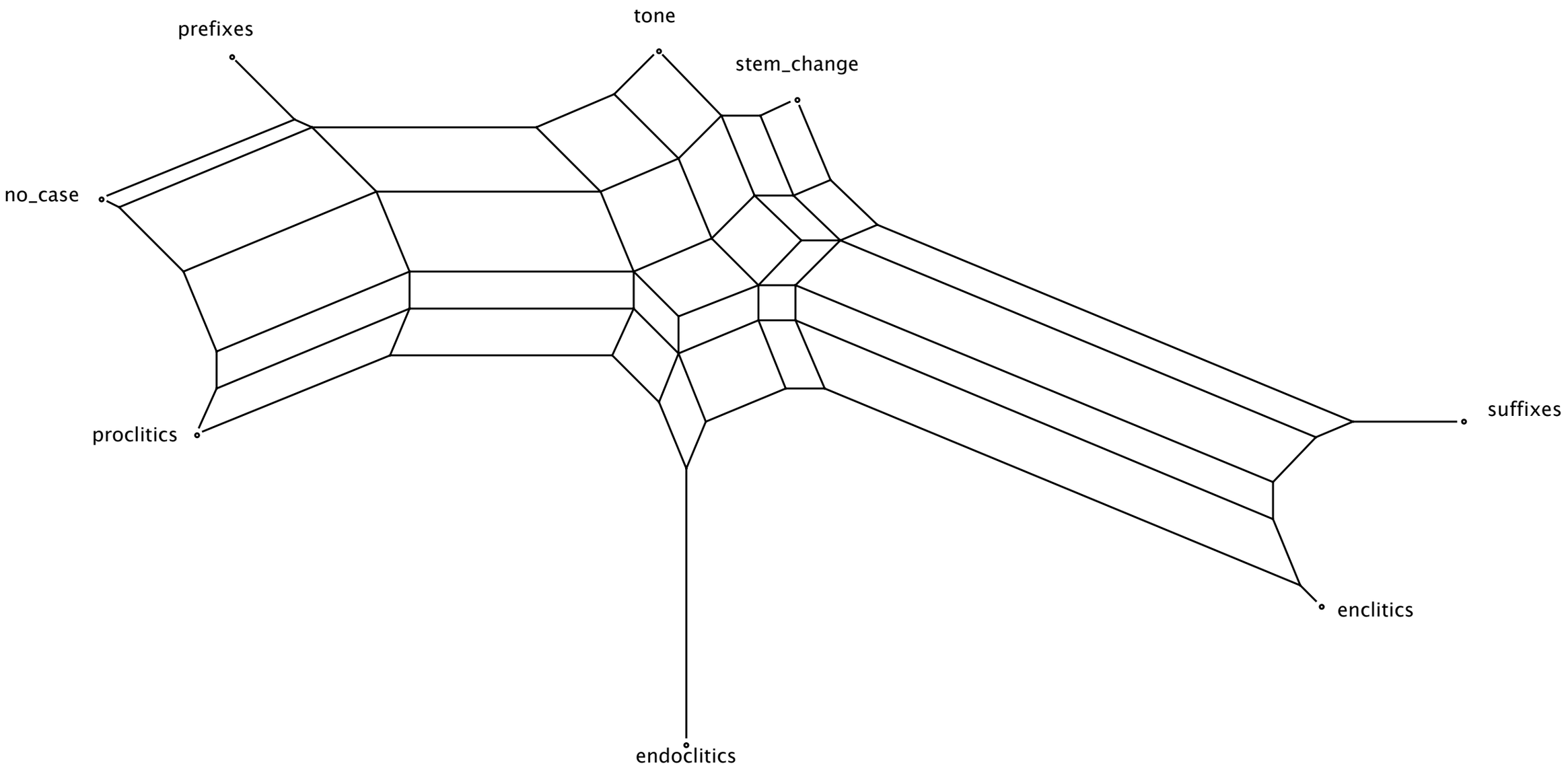
languages

COPY LIST

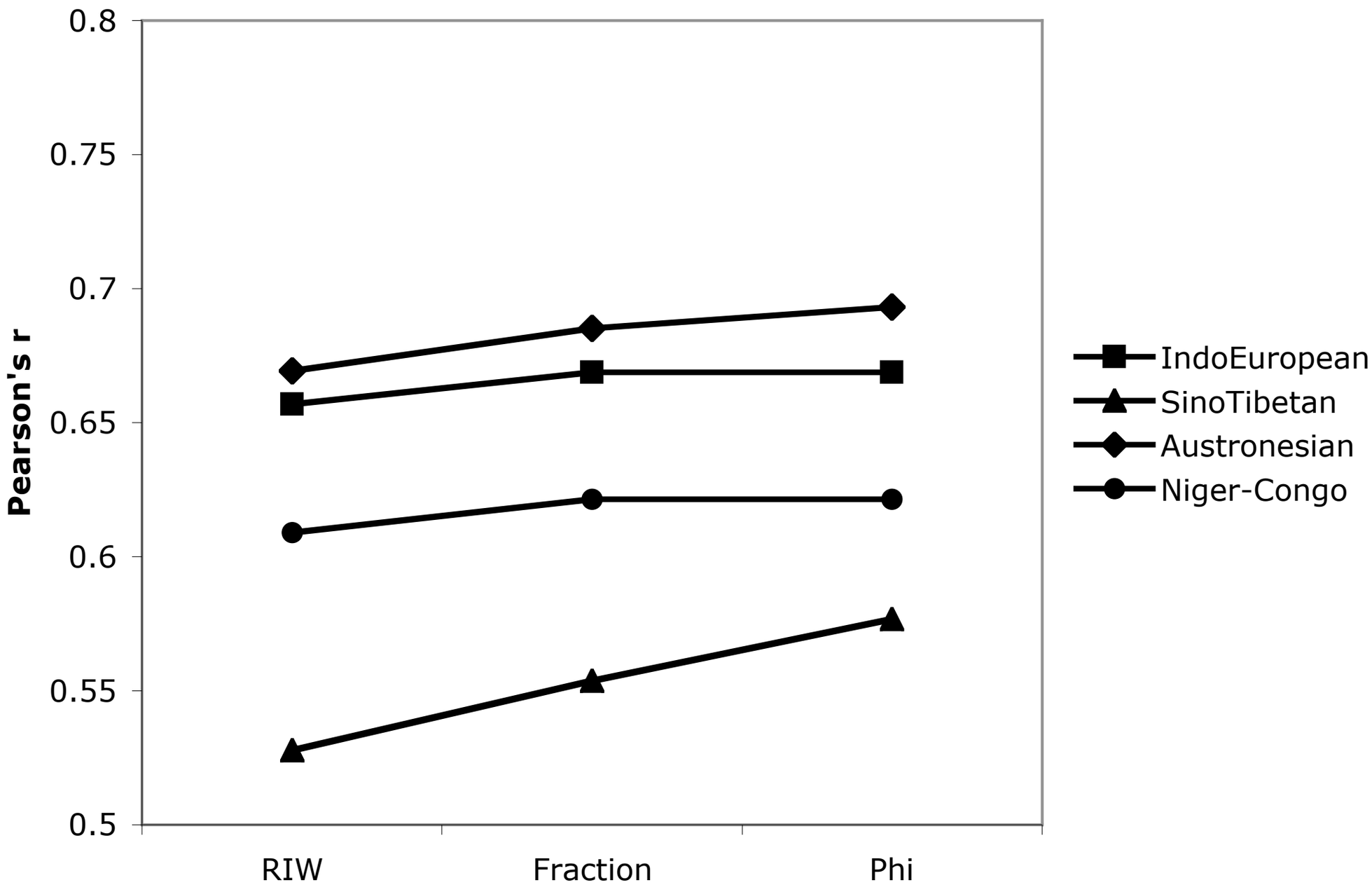
Type Similarities

	suffix	prefix	tone	stem	mixed	enclit	proclit	endo	no
suffixes	1	0.04	0.67	0.61	0.36	0.80	0.18	0.26	0.00
prefixes	0.04	1	0.72	0.74	0.88	0.21	0.62	0.48	0.88
tone	0.67	0.72	1	0.77	0.75	0.43	0.52	0.53	0.74
stem_change	0.61	0.74	0.77	1	0.76	0.48	0.53	0.53	0.64
mixed	0.36	0.88	0.75	0.76	1	0.29	0.96	0.52	0.86
enclitics	0.80	0.21	0.43	0.48	0.29	1	0.26	0.43	0.18
proclitics	0.18	0.62	0.52	0.53	0.96	0.26	1	0.52	0.89
endoclitics	0.26	0.48	0.53	0.53	0.52	0.43	0.52	1	0.29
no_case	0.00	0.88	0.74	0.64	0.86	0.18	0.89	0.29	1

Network of Types

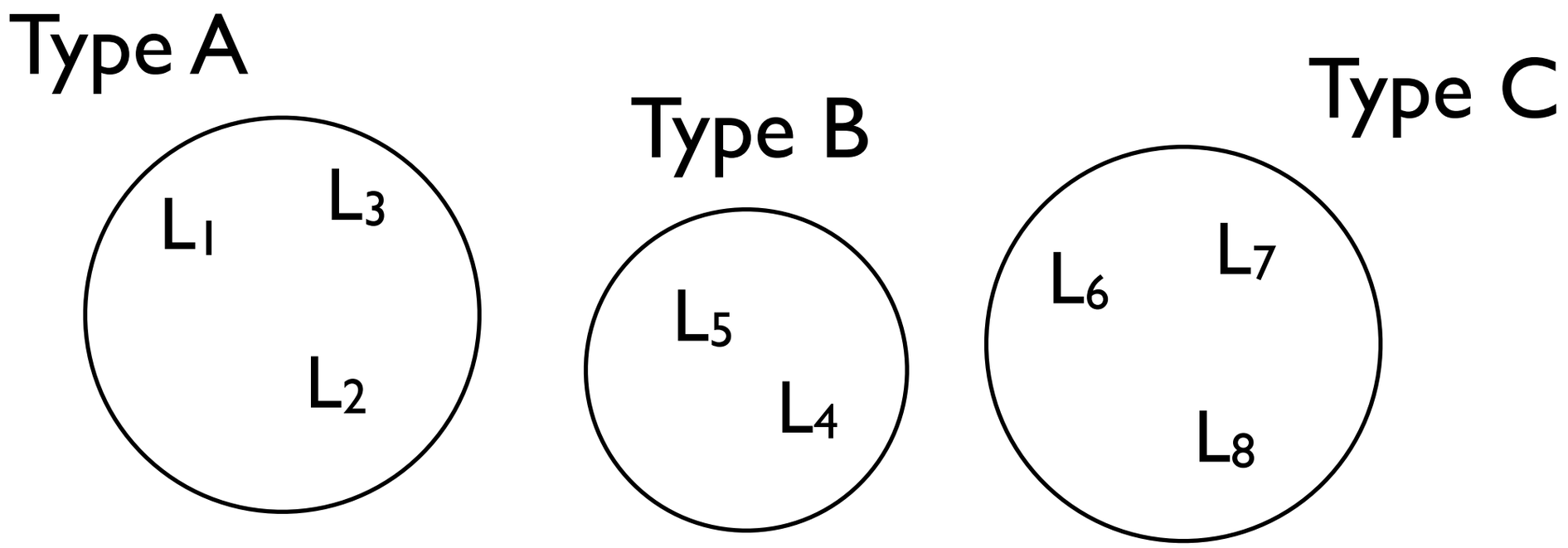


Specifying Differences



Further Improvements

- Combine both of these improvement (?)
- Investigate which typologies are most strongly linked to goal at hand:
  - genealogically stable features
  - less likely borrowable features
- Make finer grained typologies!


Type A
L1
L3
L2
Type B
L5
L4
Type C
L6
L7
L8

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1									
L_2									
L_3									
L_4									
L_5									
L_6									
L_7									
L_8									
…									

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	I								
L_2		I							
L_3			I						
L_4				I					
L_5					I				
L_6						I			
L_7							I		
L_8								I	
…									

Type A Type B Type C

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	I								
L_2		I							
L_3			I						
L_4				I					
L_5					I				
L_6						I			
L_7							I		
L_8								I	
…									

	Type A			Type B		Type C			
	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	I	I	I						
L_2	I	I	I						
L_3	I	I	I						
L_4				I	I				
L_5				I	I				
L_6						I	I	I	
L_7						I	I	I	
L_8						I	I	I	
…									

	Type A			Type B		Type C			
	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	1	1	1	0	0	0	0	0	
L_2	1	1	1	0	0	0	0	0	
L_3	1	1	1	0	0	0	0	0	
L_4	0	0	0	1	1	0	0	0	
L_5	0	0	0	1	1	0	0	0	
L_6	0	0	0	0	0	1	1	1	
L_7	0	0	0	0	0	1	1	1	
L_8	0	0	0	0	0	1	1	1	
…									

Undifferentiated Typology

	Type A			Type B		Type C			
	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	1	1	1	0.37	0.37	0.28	0.28	0.28	
L_2	1	1	1	0.37	0.37	0.28	0.28	0.28	
L_3	1	1	1	0.37	0.37	0.28	0.28	0.28	
L_4	0.37	0.37	0.37	1	1	0.58	0.58	0.58	
L_5	0.37	0.37	0.37	1	1	0.58	0.58	0.58	
L_6	0.28	0.28	0.28	0.58	0.58	1	1	1	
L_7	0.28	0.28	0.28	0.58	0.58	1	1	1	
L_8	0.28	0.28	0.28	0.58	0.58	1	1	1	
…									

Differentiated Typology

	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	…
L_1	1	0.55	0.72	0.31	0.70	0.61	0.50	0.58	
L_2	0.55	1	0.55	0.31	0.40	0.44	0.31	0.48	
L_3	0.72	0.55	1	0.29	0.53	0.51	0.48	0.60	
L_4	0.31	0.31	0.29	1	0.38	0.36	0.26	0.27	
L_5	0.70	0.40	0.53	0.38	1	0.64	0.51	0.46	
L_6	0.61	0.44	0.51	0.36	0.64	1	0.57	0.43	
L_7	0.50	0.31	0.48	0.26	0.51	0.57	1	0.47	
L_8	0.58	0.48	0.60	0.27	0.46	0.43	0.47	1	
…									

‘Deconstructed’ Typology, or a ‘Typology without Types’

Contextually Situated Exemplars

Contextually Situated Exemplars ?

- Choose a domain to make a typology
- Choose a (large) set of **situations in context** that fit into the domain
- Investigate these situations in all languages to compare
- Easy way: use existing translations ('parallel texts')

Appendix

The TMA questionnaire

Context indications are given within square brackets. Words within parentheses are not to be translated.

Part A – sentences

(1) [Standing in front of a house] The house BE BIG

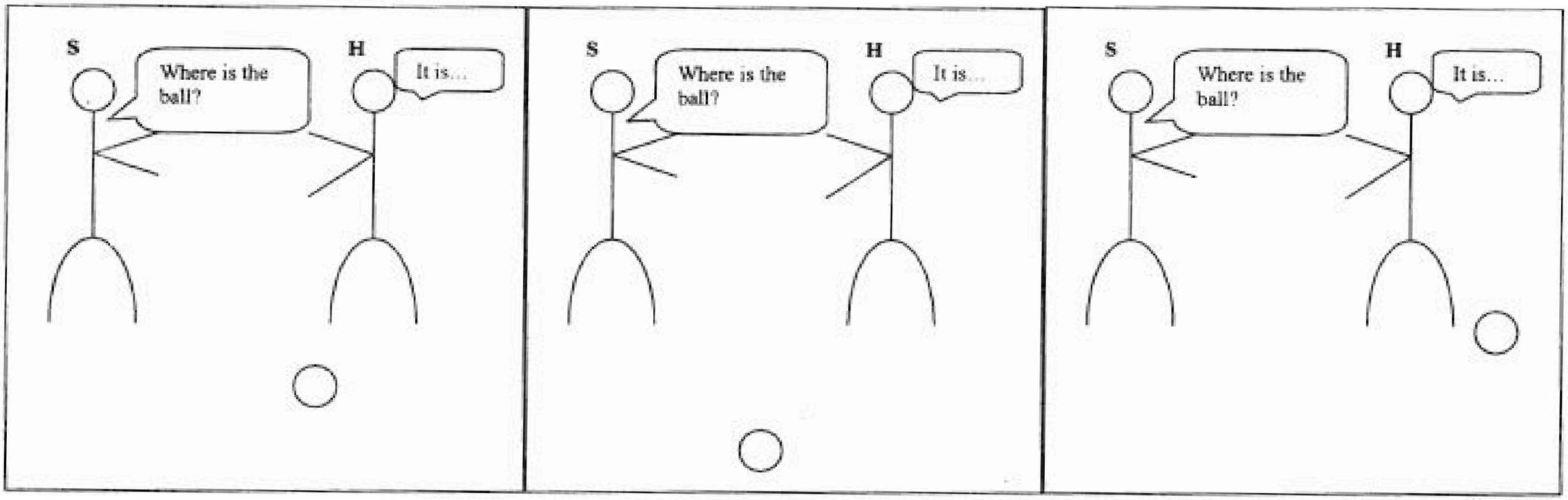
(2) [Talking about the house in which the speaker lives (the house is out of sight)] The house BE BIG

(3) [Talking about a house in which the speaker used to live but which has now been torn down] The house BE BIG

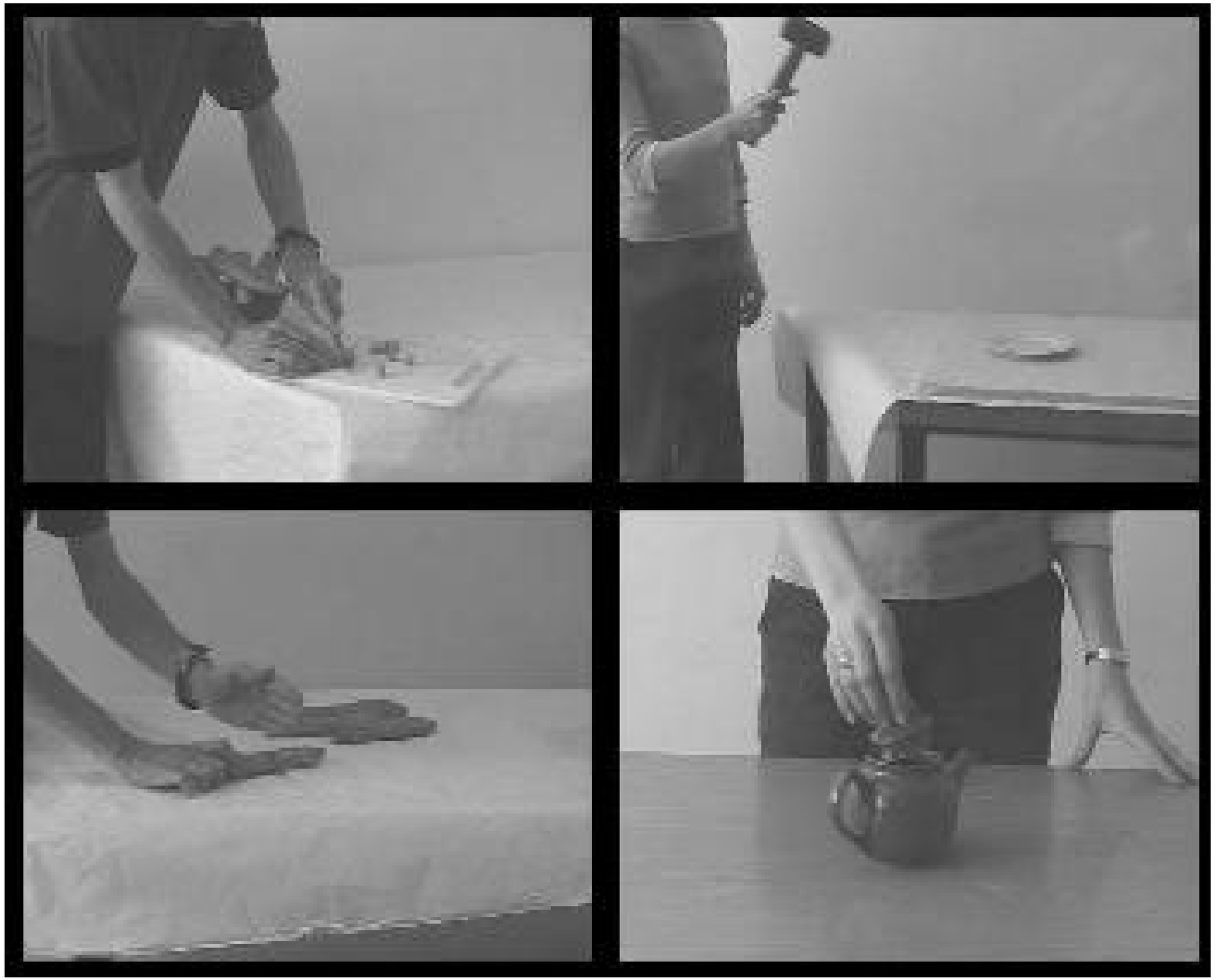
(4) [Talking about a house which the speaker saw for the first time yesterday and doesn't see now:] The house BE BIG

(5) [Q: What your brother DO right now? (=What activity is he engaged in?) A by someone who can see him] He WRITE letters

Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.



Da Milano, Federica (2005) *La Deissi Spaziale nelle Lingue d'Europe*. Pavia: Francoangeli.



Majid, Asifa *et al.* (2004) Event categorization: A crosslinguistic perspective. Proceedings of AMCSS, pp. 885-890.

	MRD	LIT	ENG	FRE
1050	sams	eiti	go	aller
1070	sams	eiti	come	venir
1090	sams	eiti	come	venir
1104	lisems	kopti	come	sortir
1105	valgoms	zengti	descend	descendre
1114	–	–	come	se faire entendre
1120	vetjams	varyti	drive	pousser
1140	sams	eiti	come	se rendre
1160	jutams	eiti	walk	marcher

	MRD	LIT	ENG	FRE
1050	sams	eiti	go	aller
1070	sams	eiti	come	venir
1090	sams	eiti	come	venir
1104	lisems	kopti	come	sortir
1105	valgoms	zengti	descend	descendre
1114	–	–	come	se faire entendre
1120	vetjams	varyti	drive	pousser
1140	sams	eiti	come	se rendre
1160	jutams	eiti	walk	marcher

Contextually Situated Exemplar

	MRD	LIT	ENG	FRE
1050	sams	eiti	go	aller
1070	sams	eiti	come	venir
1090	sams	eiti	come	venir
1104	lisems	kopti	come	sortir
1105	valgoms	zengti	descend	descendre
1114	–	–	come	se faire entendre
1120	vetjams	varyti	drive	pousser
1140	sams	eiti	come	se rendre
1160	jutams	eiti	walk	marcher

Languoid

	MRD	LIT	ENG	FRE
1050	sams	eiti	go	aller
1070	sams	eiti	come	venir
1090	sams	eiti	come	venir
1104	lisems	kopti	come	sortir
1105	valgoms	zengti	descend	descendre
1114	–	–	come	se faire entendre
1120	vetjams	varyti	drive	pousser
1140	sams	eiti	come	se rendre
1160	jutams	eiti	walk	marcher

Language Specific Category

Using Contextually Situated Exemplars

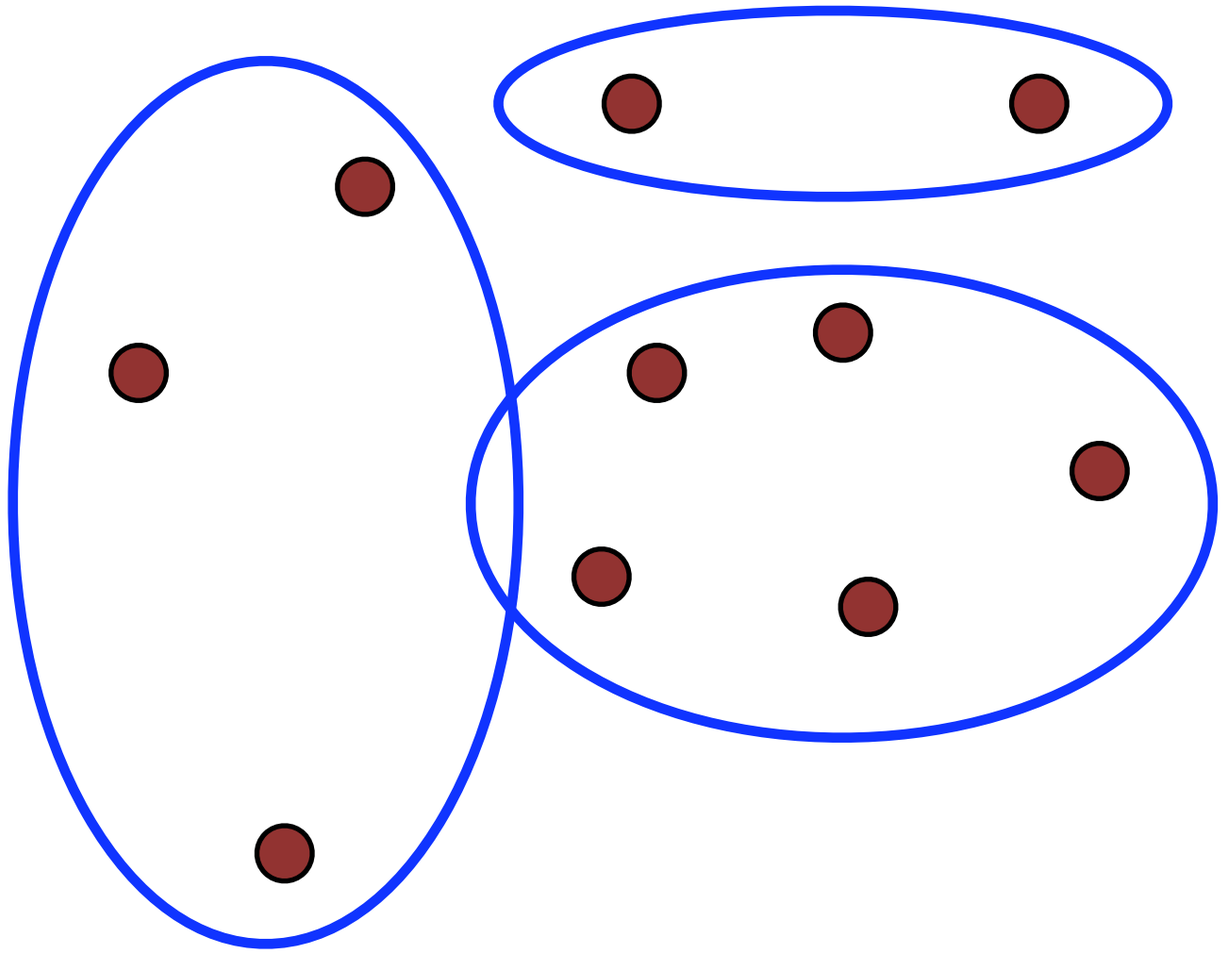
A. Compare Exemplars (cf. Semantic Map)

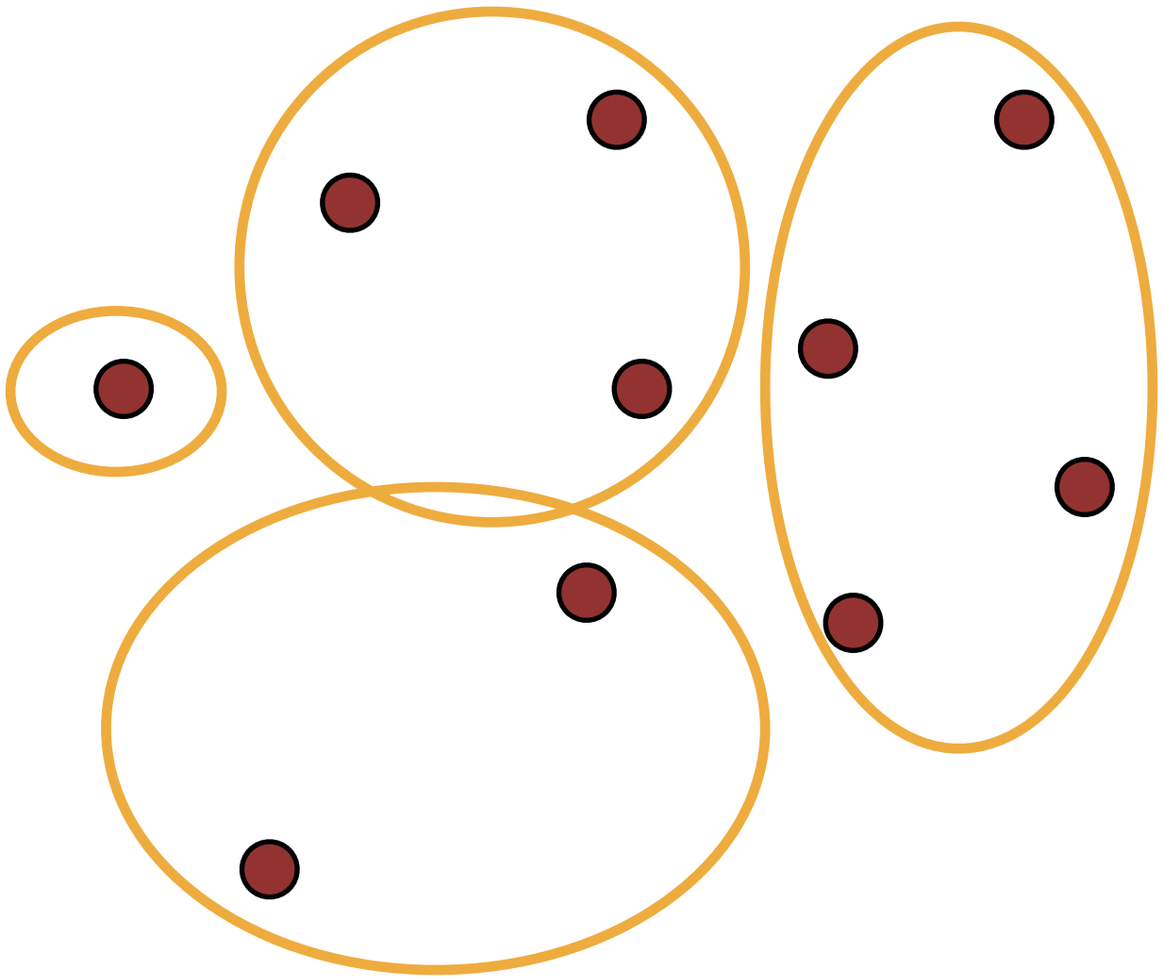
B. Compare Categories (The Real Thing!)

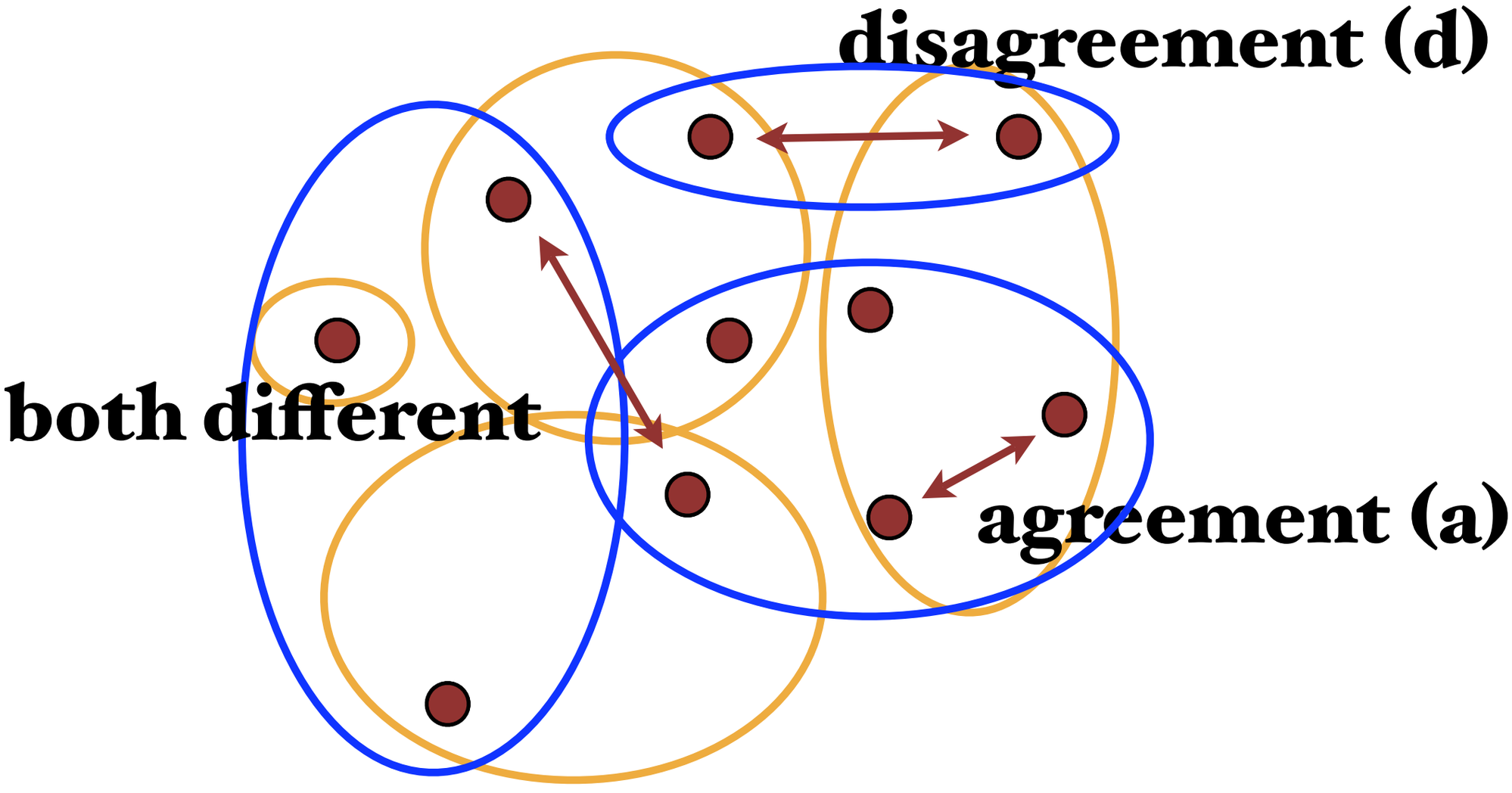
C. Compare Languoids (cf. Typology)

	MRD	LIT	ENG	FRE
1050	sams	eiti	go	aller
1070	sams	eiti	come	venir
1090	sams	eiti	come	venir
1104	lisems	kopti	come	sortir
1105	valgoms	zengti	descend	descendre
1114	–	–	come	se faire entendre
1120	vetjams	varyti	drive	pousser
1140	sams	eiti	come	se rendre
1160	jutams	eiti	walk	marcher

	MRD	LIT	ENG	FRE
1050	sams	eiti	go	aller
1070	sams	eiti	come	venir
1090	sams	eiti	come	venir
1104	lisems	kopti	come	sortir
1105	valgoms	zengti	descend	descendre
1114	–	–	come	se faire entendre
1120	vetjams	varyti	drive	pousser
1140	sams	eiti	come	se rendre
1160	jutams	eiti	walk	marcher



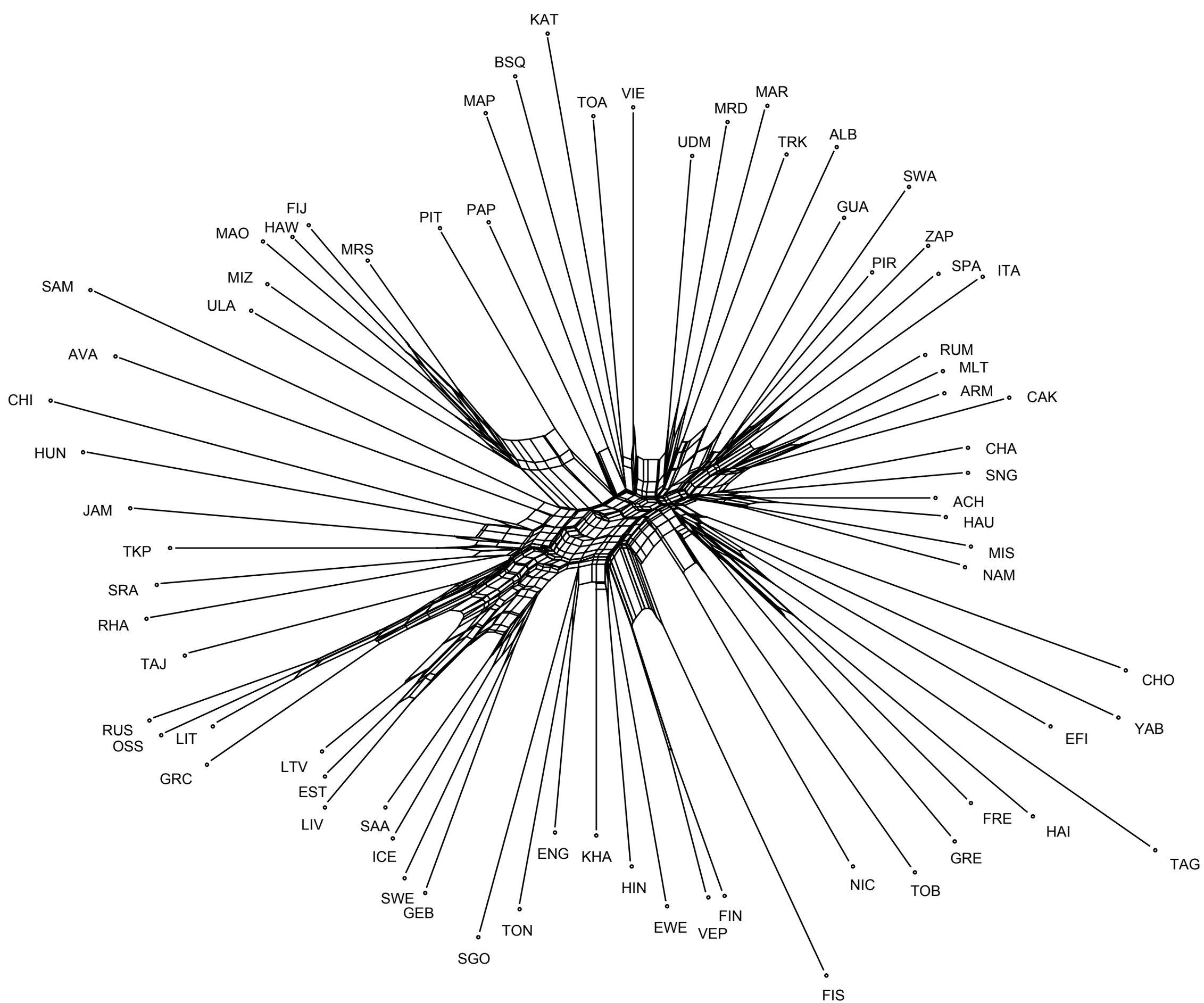




Jaccard similarity:

$a/(a+d)$

	ACH	ALB	ARM	AVA	BSQ	CAK	CHA	CHI	CHO	EFI	ENG	EST	EWE	FIJ	FIN	FIS	FRE	GEB	GRC	GRE	GUA	HAI	HAU	HAW	HIN	HUN	ICE	ITA	JAM	KAT	KHA	LIT	LIV	LTV	MAO	MAP	MA
ACH	1	0.43	0.62	0.31	0.41	0.57	0.6	0.22	0.41	0.51	0.55	0.46	0.54	0.35	0.46	0.35	0.45	0.44	0.4	0.39	0.52	0.41	0.72	0.41	0.55	0.23	0.48	0.47	0.32	0.36	0.55	0.37	0.47	0.45	0.36	0.43	0.4
ALB	0.43	1	0.53	0.26	0.41	0.45	0.41	0.2	0.32	0.34	0.38	0.3	0.33	0.26	0.31	0.29	0.4	0.33	0.25	0.41	0.53	0.35	0.49	0.29	0.34	0.28	0.3	0.5	0.3	0.28	0.37	0.26	0.28	0.33	0.28	0.4	0.4
ARM	0.62	0.53	1	0.27	0.4	0.63	0.61	0.26	0.37	0.4	0.54	0.46	0.47	0.34	0.42	0.4	0.41	0.44	0.4	0.39	0.57	0.33	0.67	0.36	0.53	0.24	0.46	0.49	0.32	0.31	0.49	0.34	0.46	0.49	0.34	0.37	0.3
AVA	0.31	0.26	0.27	1	0.27	0.24	0.27	0.25	0.18	0.22	0.32	0.36	0.27	0.31	0.33	0.21	0.26	0.32	0.24	0.25	0.31	0.23	0.26	0.28	0.32	0.25	0.34	0.26	0.28	0.25	0.32	0.29	0.33	0.34	0.3	0.27	0.3
BSQ	0.41	0.41	0.4	0.27	1	0.41	0.34	0.22	0.25	0.28	0.29	0.28	0.32	0.29	0.32	0.21	0.36	0.26	0.23	0.37	0.44	0.33	0.38	0.3	0.32	0.26	0.28	0.42	0.25	0.32	0.34	0.25	0.27	0.3	0.34	0.36	0.3
CAK	0.57	0.45	0.63	0.24	0.41	1	0.55	0.21	0.39	0.4	0.4	0.38	0.42	0.3	0.36	0.34	0.44	0.31	0.32	0.37	0.56	0.35	0.61	0.31	0.43	0.21	0.37	0.48	0.27	0.26	0.41	0.29	0.38	0.39	0.29	0.41	0.3
CHA	0.6	0.41	0.61	0.27	0.34	0.55	1	0.24	0.41	0.41	0.6	0.41	0.46	0.33	0.43	0.39	0.43	0.43	0.39	0.43	0.47	0.33	0.59	0.39	0.45	0.24	0.5	0.46	0.3	0.35	0.57	0.33	0.43	0.46	0.33	0.37	0.3
CHI	0.22	0.2	0.26	0.25	0.22	0.21	0.24	1	0.16	0.2	0.27	0.37	0.22	0.28	0.26	0.19	0.19	0.31	0.28	0.18	0.25	0.18	0.23	0.23	0.25	0.29	0.3	0.2	0.33	0.25	0.3	0.29	0.33	0.33	0.25	0.25	0.2
CHO	0.41	0.32	0.37	0.18	0.25	0.39	0.41	0.16	1	0.29	0.28	0.24	0.31	0.21	0.26	0.22	0.27	0.24	0.22	0.27	0.33	0.27	0.45	0.23	0.27	0.17	0.26	0.32	0.21	0.26	0.31	0.21	0.24	0.27	0.2	0.29	0.3
EFI	0.51	0.34	0.4	0.22	0.28	0.4	0.41	0.2	0.29	1	0.32	0.27	0.36	0.23	0.31	0.27	0.32	0.26	0.23	0.32	0.33	0.32	0.51	0.26	0.35	0.18	0.29	0.37	0.22	0.27	0.41	0.2	0.29	0.27	0.25	0.32	0.3
ENG	0.55	0.38	0.54	0.32	0.29	0.4	0.6	0.27	0.28	0.32	1	0.58	0.47	0.47	0.48	0.37	0.32	0.52	0.5	0.34	0.39	0.25	0.5	0.46	0.54	0.33	0.6	0.37	0.44	0.35	0.58	0.5	0.56	0.6	0.41	0.31	0.3
EST	0.46	0.3	0.46	0.36	0.28	0.38	0.41	0.37	0.24	0.27	0.58	1	0.46	0.49	0.48	0.3	0.28	0.58	0.49	0.25	0.39	0.21	0.41	0.38	0.48	0.39	0.63	0.29	0.46	0.31	0.58	0.6	0.8	0.79	0.42	0.31	0.3
EWE	0.54	0.33	0.47	0.27	0.32	0.42	0.46	0.22	0.31	0.36	0.47	0.46	1	0.29	0.42	0.29	0.35	0.4	0.33	0.3	0.38	0.31	0.52	0.31	0.52	0.25	0.46	0.32	0.32	0.31	0.49	0.35	0.43	0.49	0.32	0.33	0.3
FIJ	0.35	0.26	0.34	0.31	0.29	0.3	0.33	0.28	0.21	0.23	0.47	0.49	0.29	1	0.32	0.19	0.22	0.36	0.47	0.22	0.34	0.21	0.36	0.73	0.34	0.36	0.43	0.28	0.35	0.32	0.38	0.66	0.45	0.48	0.67	0.33	0.
FIN	0.46	0.31	0.42	0.33	0.32	0.36	0.43	0.26	0.26	0.31	0.48	0.48	0.42	0.32	1	0.42	0.31	0.45	0.36	0.33	0.38	0.24	0.45	0.3	0.5	0.28	0.48	0.36	0.37	0.28	0.48	0.38	0.46	0.48	0.3	0.27	0.3
FIS	0.35	0.29	0.4	0.21	0.21	0.34	0.39	0.19	0.22	0.27	0.37	0.3	0.29	0.19	0.42	1	0.32	0.36	0.25	0.29	0.35	0.21	0.39	0.22	0.41	0.21	0.34	0.32	0.26	0.19	0.34	0.25	0.32	0.33	0.2	0.19	0.2
FRE	0.45	0.4	0.41	0.26	0.36	0.44	0.43	0.19	0.27	0.32	0.32	0.28	0.35	0.22	0.31	0.32	1	0.3	0.22	0.43	0.43	0.52	0.39	0.24	0.36	0.21	0.35	0.51	0.25	0.26	0.34	0.22	0.26	0.28	0.23	0.35	0.3
GEB	0.44	0.33	0.44	0.32	0.26	0.31	0.43	0.31	0.24	0.26	0.52	0.58	0.4	0.36	0.45	0.36	0.3	1	0.41	0.26	0.36	0.2	0.41	0.36	0.49	0.31	0.57	0.3	0.39	0.25	0.46	0.48	0.56	0.59	0.33	0.22	0.
GRC	0.4	0.25	0.4	0.24	0.23	0.32	0.39	0.28	0.22	0.23	0.5	0.49	0.33	0.47	0.36	0.25	0.22	0.41	1	0.21	0.29	0.2	0.38	0.45	0.37	0.34	0.44	0.25	0.37	0.25	0.42	0.71	0.52	0.51	0.4	0.22	0.2
GRE	0.39	0.41	0.39	0.25	0.37	0.37	0.43	0.18	0.27	0.32	0.34	0.25	0.3	0.22	0.33	0.29	0.43	0.26	0.21	1	0.38	0.38	0.42	0.24	0.31	0.23	0.28	0.42	0.24	0.27	0.39	0.2	0.24	0.28	0.24	0.35	0.3
GUA	0.52	0.53	0.57	0.31	0.44	0.56	0.47	0.25	0.33	0.33	0.39	0.39	0.38	0.34	0.38	0.35	0.43	0.36	0.29	0.38	1	0.38	0.51	0.33	0.48	0.28	0.38	0.47	0.31	0.3	0.46	0.3	0.37	0.42	0.35	0.49	0.4
HAI	0.41	0.35	0.33	0.23	0.33	0.35	0.33	0.18	0.27	0.32	0.25	0.21	0.31	0.21	0.24	0.21	0.52	0.2	0.2	0.38	0.38	1	0.39	0.23	0.25	0.21	0.22	0.44	0.21	0.25	0.28	0.18	0.19	0.23	0.21	0.42	0.3
HAU	0.72	0.49	0.67	0.26	0.38	0.61	0.59	0.23	0.45	0.51	0.5	0.41	0.52	0.36	0.45	0.39	0.39	0.41	0.38	0.42	0.51	0.39	1	0.38	0.51	0.24	0.43	0.47	0.33	0.31	0.51	0.33	0.43	0.44	0.34	0.44	0.4
HAW	0.41	0.29	0.36	0.28	0.3	0.31	0.39	0.23	0.23	0.26	0.46	0.38	0.31	0.73	0.3	0.22	0.24	0.36	0.45	0.24	0.33	0.23	0.38	1	0.37	0.35	0.37	0.31	0.35	0.34	0.38	0.59	0.38	0.41	0.67	0.32	0.3
HIN	0.55	0.34	0.53	0.32	0.32	0.43	0.45	0.25	0.27	0.35	0.54	0.48	0.52	0.34	0.5	0.41	0.36	0.49	0.37	0.31	0.48	0.25	0.51	0.37	1	0.31	0.51	0.35	0.35	0.33	0.57	0.41	0.53	0.53	0.37	0.29	0.3
HUN	0.23	0.28	0.24	0.25	0.26	0.21	0.24	0.29	0.17	0.18	0.33	0.39	0.25	0.36	0.28	0.21	0.21	0.31	0.34	0.23	0.28	0.21	0.24	0.35	0.31	1	0.3	0.26	0.39	0.28	0.3	0.39	0.35	0.42	0.35	0.28	0.2
ICE	0.48	0.3	0.46	0.34	0.28	0.37	0.5	0.3	0.26	0.29	0.6	0.63	0.46	0.43	0.48	0.34	0.35	0.57	0.44	0.28	0.38	0.22	0.43	0.37	0.51	0.3	1	0.34	0.39	0.28	0.55	0.48	0.58	0.61	0.35	0.26	0.3
ITA	0.47	0.5	0.49	0.26	0.42	0.48	0.46	0.2	0.32	0.37	0.37	0.29	0.32	0.28	0.36	0.32	0.51	0.3	0.25	0.42	0.47	0.44	0.47	0.31	0.35	0.26	0.34	1	0.24	0.27	0.36	0.26	0.27	0.33	0.28	0.44	0.3
JAM	0.32	0.3	0.32	0.28	0.25	0.27	0.3	0.33	0.21	0.22	0.44	0.46	0.32	0.35	0.37	0.26	0.25	0.39	0.37	0.24	0.31	0.21	0.33	0.35	0.35	0.39	0.39	0.24	1	0.32	0.41	0.42	0.4	0.38	0.34	0.32	0.2
KAT	0.36	0.28	0.31	0.25	0.32	0.26	0.35	0.25	0.26	0.27	0.35	0.31	0.31	0.32	0.28	0.19	0.26	0.25	0.25	0.27	0.3	0.25	0.31	0.34	0.33	0.28	0.28	0.27	0.32	1	0.38	0.27	0.3	0.33	0.35	0.34	0.3
KHA	0.55	0.37	0.49	0.32	0.34	0.41	0.57	0.3	0.31	0.41	0.58	0.58	0.49	0.38	0.48	0.34	0.34	0.46	0.42	0.39	0.46	0.28	0.51	0.38	0.57	0.3	0.55	0.36	0.41	0.38	1	0.37	0.56	0.53	0.41	0.37	0.3
LIT	0.37	0.26	0.34	0.29	0.25	0.29	0.33	0.29	0.21	0.2	0.5	0.6	0.35	0.66	0.38	0.25	0.22	0.48	0.71	0.2	0.3	0.18	0.33	0.59	0.41	0.39	0.48	0.26	0.42	0.27	0.37	1	0.56	0.68	0.55	0.25	0.2
LIV	0.47	0.28	0.46	0.33	0.27	0.38	0.43	0.33	0.24	0.29	0.56	0.8	0.43	0.45	0.46	0.32	0.26	0.56	0.52	0.24	0.37	0.19	0.43	0.38	0.53	0.35	0.58	0.27	0.4	0.3	0.56	0.56	1	0.75	0.39	0.26	0.3
LTV	0.45	0.33	0.49	0.34	0.3	0.39	0.46	0.33	0.27	0.27	0.6	0.79	0.49	0.48	0.48	0.33	0.28	0.59	0.51	0.28	0.42	0.23	0.44	0.41	0.53	0.42	0.61	0.33	0.38	0.33	0.53	0.68	0.75	1	0.43	0.31	0.3
MAO	0.36	0.28	0.34	0.3	0.34	0.29	0.33	0.25	0.2	0.25	0.41	0.42	0.32	0.67	0.3	0.2	0.23	0.33	0.4	0.24	0.35	0.21	0.34	0.67	0.37	0.35	0.35	0.28	0.34	0.35	0.41	0.55	0.39	0.43	1	0.3	0.3
MAP	0.43	0.4	0.37	0.27	0.36	0.41	0.37	0.25	0.29	0.32	0.31	0.31	0.33	0.33	0.27	0.19	0.35	0.22	0.22	0.35	0.49	0.42	0.44	0.32	0.29	0.28	0.26	0.44	0.32	0.34	0.37	0.25	0.26	0.31	0.3	1	0.4
MAR	0.42	0.44	0.39	0.32	0.36	0.39	0.38	0.24	0.32	0.32	0.35	0.33	0.34	0.3	0.35	0.26	0.35	0.3	0.25	0.31	0.46	0.33	0.44	0.31	0.38	0.28	0.32	0.38	0.29	0.31	0.37	0.26	0.31	0.35	0.32	0.46	
MIS	0.64	0.39	0.56	0.29	0.36	0.5	0.6	0.23	0.36	0.43	0.6	0.41	0.5	0.36	0.44	0.34	0.39	0.41	0.38	0.39	0.48	0.36	0.67	0.38	0.5	0.25	0.46	0.45	0.3	0.32	0.54	0.33	0.44	0.47	0.34	0.43	0.4
MIZ	0.48	0.34	0.42	0.27	0.34	0.4	0.39	0.22	0.29	0.31	0.37	0.35	0.33	0.57	0.28	0.2	0.3	0.3	0.37	0.3	0.4	0.29	0.47	0.6	0.35	0.3	0.32	0.34	0.3	0.32	0.46	0.43	0.33	0.34	0.59	0.36	0.3
MLT	0.64	0.58	0.72	0.29	0.42	0.57	0.54	0.24	0.33	0.43	0.52	0.41	0.46	0.32	0.42	0.35	0.43	0.41	0.34	0.37	0.62	0.36	0.61	0.34	0.48	0.25	0.45	0.52	0.31	0.28	0.52	0.31	0.42	0.48	0.33	0.39	0.4
MRD	0.46	0.4	0.49	0.35	0.36	0.46	0.5	0.23	0.33	0.33	0.38	0.32	0.37	0.28	0.39	0.35	0.4	0.36	0.31	0.32	0.45	0.35	0.45	0.31	0.42	0.22	0.39	0.41	0.27	0.28	0.4	0.28	0.36	0.37	0.29	0.33	0.
MRS	0.51	0.32	0.45	0.3	0.33	0.38	0.5	0.27	0.29	0.34	0.52	0.44	0.36	0.67	0.37	0.27	0.29	0.41	0.55	0.27	0.38	0.26	0.5	0.64	0.41	0.31	0.44	0.36	0.4	0.36	0.46	0.61	0.44	0.46	0.58	0.37	0.3
NAM	0.62	0.46	0.59	0.29	0.37	0.52	0.56	0.21	0.39	0.46	0.47	0.38	0.49	0.36	0.41	0.29	0.42	0.33	0.33	0.44	0.5	0.37	0.68	0.37	0.49	0.26	0.39	0.47	0.29	0.35	0.54	0.3	0.37	0.42	0.34	0.45	0.4
NIC	0.48	0.33	0.41	0.24	0.28	0.39	0.52	0.21	0.29	0.33	0.41	0.35	0.36	0.28	0.32	0.3	0.37	0.31	0.33	0.33	0.41	0.26	0.44	0.34	0.4	0.24	0.34	0.35	0.27	0.3	0.42	0.29	0.34	0.36	0.3	0.3	0.3
OSS	0.29	0.23	0.34	0.3	0.22	0.25	0.28	0.33	0.17	0.19	0.46	0.58	0.31	0.64	0.31	0.21	0.17	0.46	0.62	0.17	0.27	0.16	0.3	0.5	0.34	0.31	0.42	0.2	0.4	0.25	0.33	0.8	0.5	0.57	0.5	0.23	0.2
PAP	0.49	0.41	0.45	0.35	0.4	0.43	0.53	0.28	0.33	0.33	0.51	0.49	0.4	0.41	0.43	0.29	0.38	0.42	0.35	0.42	0.51	0.3	0.5	0.41	0.46	0.33	0.48	0.43	0.4	0.35	0.57	0.37	0.44	0.47	0.39	0.48	0.4
PIR	0.63	0.47	0.59	0.31	0.44	0.58	0.52	0.22	0.36	0.46	0.41	0.4	0.45	0.34	0.41	0.31	0.45	0.34	0.33	0.42	0.59	0.41	0.61	0.35	0.43	0.26	0.39	0.54	0.32	0.32	0.53	0.31	0.37	0.42	0.34	0.52	0.4
PIT	0.55	0.4	0.48	0.29	0.38	0.48	0.53	0.25	0.31	0.39	0.47	0.41	0.4	0.52	0.35	0.29	0.4	0.36	0.4	0.37	0.47	0.3	0.58	0.49	0.45	0.28	0.41	0.42	0.35	0.32	0.48	0.44	0.39	0.41	0.49	0.36	0.3
RHA	0.33	0.3	0.35	0.37	0.3	0.27	0.3	0.31	0.16	0.21	0.44	0.48	0.32	0.41	0.35	0.24	0.29	0.45	0.3	0.25	0.35	0.21	0.3	0.38	0.41	0.37	0.46	0.3	0.43	0.29	0.41	0.39	0.43	0.46	0.38	0.26	0.2
RUM	0.64	0.52	0.63	0.3	0.43	0.58	0.56	0.22	0.39	0.44	0.53	0.4	0.5	0.32	0.44	0.37	0.48	0.38	0.34	0.46	0.55	0.41	0.62	0.37	0.5	0.23	0.42	0.58	0.32	0.31	0.53	0.3	0.38	0.43	0.33	0.42	0.4
RUS	0.29	0.23	0.33	0.25	0.21	0.27	0.3	0.3	0.18	0.18	0.46	0.51	0.3	0.54	0.31	0.2	0.18	0.4	0.65	0.18	0.26	0.18	0.31	0.51	0.34	0.36	0.39	0.22	0.39	0.25	0.35	0.77	0.43	0.54	0.45	0.25	0.2
SAA	0.52	0.32	0.51	0.31	0.28	0.42	0.54	0.31	0.26	0.32	0.59	0.65	0.47	0.4	0.55	0.4	0.33	0.55	0.48	0.27	0.4	0.23	0.48	0.37	0.55	0.32	0.65	0.36	0.41	0.29	0.56	0.5	0.62	0.63	0.34	0.28	0.3
SAM	0.26	0.22	0.24	0.22	0.2	0.22	0.29	0.24	0.19	0.23	0.28	0.28	0.23	0.31	0.24	0.18	0.23	0.21	0.24	0.2	0.24	0.2	0.27	0.28	0.25	0.25	0.27	0.24	0.24	0.24	0.28	0.24	0.24	0.25	0.26	0.27	0.2
SGO	0.42	0.37	0.4	0.31	0.31	0.3	0.33	0.27	0.29	0.28	0.45	0.45	0.39	0.31	0.34	0.22	0.26	0.38	0.36	0.27	0.39	0.26	0.4	0.3	0.43	0.29	0.39	0.28	0.41	0.25	0.47	0.4	0.41	0.44	0.32	0.3	0.3
SNG	0.66	0.48	0.62	0.27	0.36	0.55	0.61	0.22	0.44	0.45	0.51	0.41	0.49	0.32	0.39	0.32	0.4	0.36	0.36	0.45	0.5	0.36	0.64	0.36	0.48	0.25	0.46	0.46	0.33	0.33	0.53	0.33	0.4	0.43	0.32	0.43	0.4
SPA	0.53	0.52	0.61	0.24	0.45	0.62	0.48	0.21	0.38	0.37	0.38	0.34	0.37	0.29	0.35	0.28	0.47	0.27	0.28	0.43	0.56	0.42	0.62	0.32	0.37	0.27	0.31	0.6	0.28	0.26	0.4	0.28	0.32	0.35	0.3	0.51	0.3
SRA	0.42	0.31	0.41	0.29	0.3	0.3	0.39	0.34	0.25	0.28	0.44	0.49	0.38	0.3	0.39	0.28	0.29	0.52	0.42	0.3	0.34	0.23	0.37	0.32	0.42	0.29	0.48	0.29	0.38	0.28	0.46	0.4	0.46	0.43	0.32	0.25	0.2
SWA	0.57	0.44	0.49	0.22	0.36	0.54	0.43	0.17	0.37	0.44	0.31	0.28	0.36	0.27	0.34	0.28	0.39	0.24	0.26	0.38	0.5	0.47	0.59	0.27	0.33	0.22	0.29	0.48	0.24	0.25	0.38	0.23	0.28	0.31	0.25	0.48	0.3
SWE	0.45	0.25	0.4	0.32	0.27	0.31	0.43	0.28	0.26	0.25	0.52	0.51	0.41	0.41	0.5	0.3	0.29	0.6	0.43	0.24	0.33	0.19	0.39	0.39	0.52	0.33	0.62	0.29	0.4	0.3	0.47	0.53	0.49	0.54	0.36	0.21	0.2
TAG	0.27	0.23	0.22	0.15	0.18	0.22	0.29	0.14	0.22	0.3	0.19	0.14	0.21	0.12	0.2	0.19	0.29	0.17	0.13	0.28	0.21	0.3	0.26	0.15	0.18	0.13	0.18	0.28	0.12	0.21	0.22	0.11	0.16	0.17	0.13	0.26	0.2
TAJ	0.39	0.29	0.44	0.29	0.25	0.3	0.41	0.3	0.23	0.24	0.58	0.52	0.34	0.4	0.4	0.28	0.24	0.47	0.51	0.23	0.38	0.19	0.38	0.38	0.44	0.3	0.49	0.28	0.39	0.28	0.44	0.53	0.49	0.49	0.37	0.24	0.3
TKP	0.36	0.32	0.34	0.34	0.35	0.3	0.33	0.33	0.23	0.25	0.43	0.53	0.4	0.37	0.41	0.22	0.28	0.41	0.4	0.29	0.39	0.26	0.33	0.35	0.39	0.38	0.41	0.3	0.5	0.31	0.46	0.44	0.47	0.49	0.35	0.32	0.3
TOA	0.45	0.37	0.37	0.34	0.38	0.36	0.42	0.26	0.26	0.31	0.45	0.41	0.35	0.37	0.32	0.24	0.35	0.35	0.28	0.34	0.42	0.35	0.4	0.39	0.36	0.33	0.38	0.39	0.37	0.38	0.41	0.32	0.36	0.44	0.4	0.4	0.3
TOB	0.41	0.36	0.37	0.26	0.29	0.36	0.39	0.16	0.27	0.32	0.32	0.28	0.34	0.24	0.36	0.29	0.41	0.28	0.23	0.35	0.39	0.27	0.38	0.24	0.36	0.22	0.34	0.39	0.26	0.23	0.38	0.23	0.24	0.28	0.24	0.27	0.2
TON	0.48	0.29	0.4	0.3	0.27	0.35	0.47	0.24	0.29	0.3	0.6	0.48	0.38	0.46	0.38	0.32	0.31	0.36	0.37	0.31	0.37	0.23	0.48	0.4	0.43	0.28	0.44	0.33	0.35	0.33	0.5	0.37	0.42	0.45	0.35	0.34	0.3
TRK	0.48	0.5	0.46	0.34	0.38	0.42	0.43	0.24	0.28	0.34	0.42	0.37	0.41	0.29	0.4	0.32	0.41	0.4	0.28	0.37	0.5	0.35	0.48	0.31	0.42	0.29	0.38	0.43	0.31	0.28	0.39	0.29	0.35	0.4	0.31	0.37	0.4
UDM	0.57	0.46	0.54	0.3	0.34	0.49	0.5	0.24	0.35	0.39	0.42	0.35	0.41	0.31	0.37	0.33	0.4	0.33	0.32	0.38	0.5	0.37	0.57	0.34	0.46	0.23	0.39	0.47	0.27	0.28	0.43	0.31	0.35	0.38	0.3	0.38	0.4
ULA	0.46	0.33	0.4	0.27	0.32	0.37	0.37	0.24	0.28	0.34	0.36	0.38	0.34	0.49	0.3	0.22	0.27	0.31	0.42	0.29	0.41	0.3	0.46	0.5	0.35	0.3	0.34	0.37	0.29	0.33	0.4	0.47	0.39	0.44	0.5	0.38	0.3
VEP	0.44	0.32	0.45	0.35	0.32	0.38	0.47	0.29	0.27	0.3	0.49	0.49	0.42	0.34	0.76	0.42	0.31	0.47	0.39	0.32	0.38	0.25	0.43	0.34	0.48	0.27	0.46	0.32	0.35	0.26	0.48	0.41	0.48	0.51	0.3	0.24	0.
VIE	0.46	0.36	0.42	0.26	0.34	0.37	0.45	0.22	0.32	0.29	0.41	0.34	0.31	0.36	0.33	0.23	0.32	0.32	0.28	0.33	0.41	0.3	0.42	0.4	0.38	0.28	0.36	0.41	0.28	0.36	0.4	0.31	0.31	0.37	0.36	0.36	0.3
YAB	0.33	0.32	0.29	0.19	0.27	0.31	0.29	0.17	0.27	0.32	0.25	0.25	0.25	0.28	0.26	0.19	0.33	0.2	0.22	0.29	0.32	0.33	0.39	0.25	0.26	0.24	0.25	0.33	0.21	0.24	0.3	0.21	0.24	0.26	0.22	0.32	0.2
ZAP	0.5	0.47	0.52	0.26	0.36	0.51	0.48	0.22	0.36	0.35	0.38	0.37	0.37	0.34	0.39	0.28	0.36	0.31	0.3	0.41	0.51	0.37	0.58	0.33	0.36	0.28	0.35	0.45	0.34	0.3	0.44	0.29	0.35	0.38	0.33	0.47	0.3


KAT
BSQ
MAP
TOA
VIE
MAR
MRD
UDM
TRK
ALB
SWA
GUA
ZAP
PIR
SPA
ITA
RUM
MLT
ARM
CAK
CHA
SNG
ACH
HAU
MIS
NAM
CHO
YAB
EFI
FRE
HAI
TAG
GRE
NIC
TOB
FIS
FIN
VEP
EWE
HIN
KHA
ENG
TON
SGO
GEB
SWE
ICE
SAA
LIV
EST
LTV
GRC
LIT
OSS
RUS
TAJ
RHA
SRA
TKP
JAM
AVA
CHI
HUN
SAM
ULA
MIZ
MRS
MAO
HAW
FIJ
PAP
PIT

Wordlists

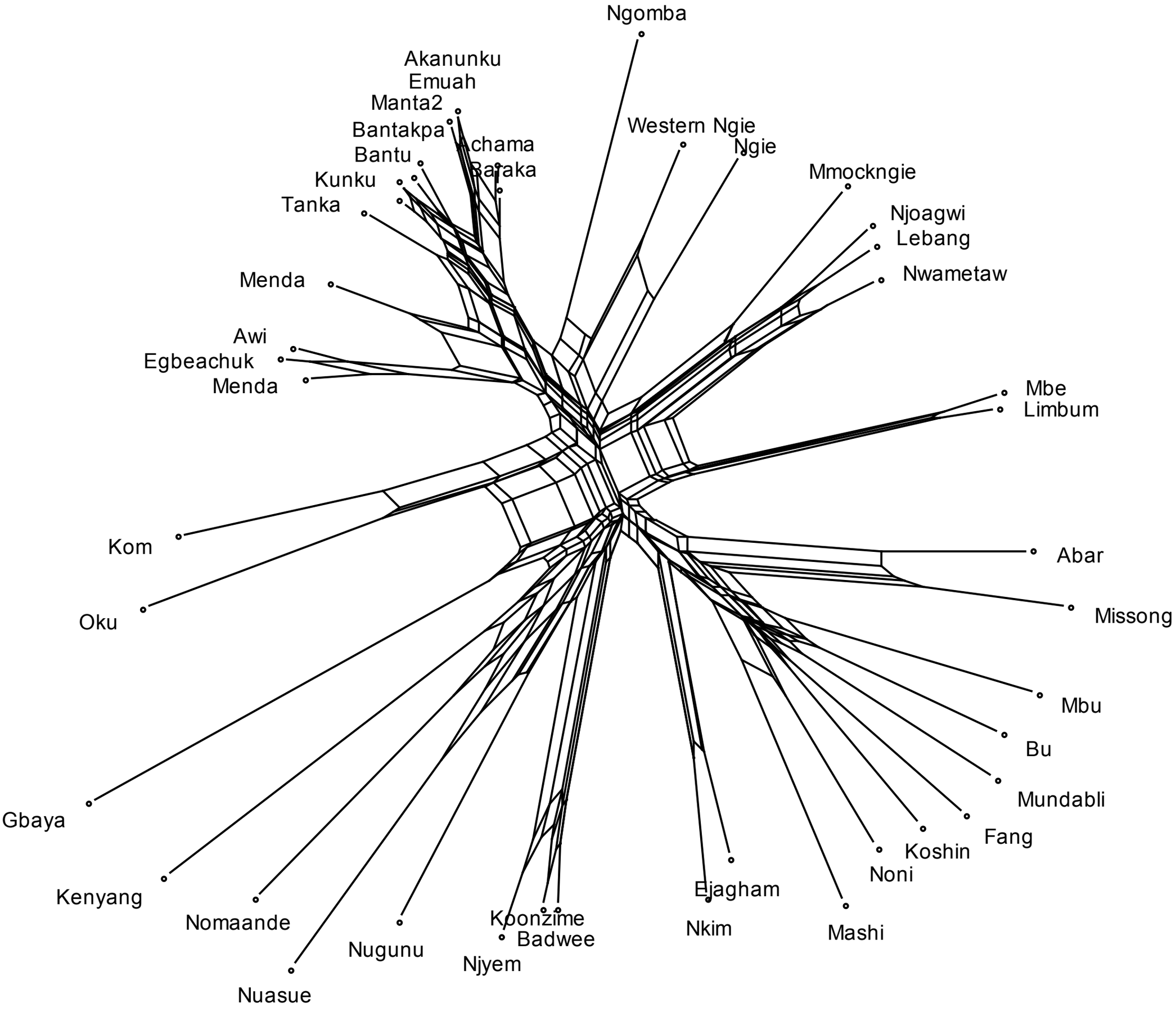
Reminder!

- Wordlists are **not** only useful for investigating genealogical relations
- Wordlists are just ‘parallel dictionaries’ that allow for quantification of lexical similarity
- The meaning of such similarities is a research question

Traditional Approach to Wordlist Analysis

- Presence vs. absence of cognates is used for computation of similarity
- Problems:
  - Judgement of cognacy presumes hypothesis about relationship
  - Very much of the available information is thrown away

Language	Group	‘water’					
Abar	Beboid	a		nj	a		
Missong	Beboid	a		nj	ɛ		
Bu	Beboid			ŋg	ɨ	n	
Mundabli	Beboid			ŋg	i		
Koshin	Beboid			nd	i		
Fang	Beboid			ndz	ia	m	
Mbu'	Beboid			mg	iə	ŋ	
Mashi	Beboid			ngw	ɔ		
Noni	Beboid			j	oo		
Ejagham	Ekoid				á	b	
Nkim	Ekoid			l	ɨ́	b	
Kendem	Mamfe		a	n	á	ʔ	
Kenyang	Mamfe	m	a	ny	ɛ	p	
Yambetta	Mbam A40	m	ə	n	í		
Nomaande	Mbam A40	m	e	ny	ïù	f	eà
Nuasue	Mbam A60	m	i	mb	i		
Badwee	Narrow Bantu, A80	m	o	d	ïù	b	eà
Koonzime	Narrow Bantu, A80	m	e	d	ïù	b	eà
Njyem	Narrow Bantu, A80	m	eè	d	ïù	b	oà


Ngomba
Akanunku
Emuah
Manta2
Bantakpa
Achama
Bantu
Baraka
Kunku
Tanka
Western Ngie
Ngie
Mmockngie
Njoagwi
Lebang
Nwametaw
Menda
Awi
Egbeachuk
Menda
Mbe
Limbum
Kom
Oku
Abar
Missong
Mbu
Bu
Mundabli
Fang
Koshin
Noni
Gbaya
Kenyang
Nomaande
Nugunu
Nuasue
Njyem
Koonzime
Badwee
Ejagham
Nkim
Mashi

Max Planck Institute
for Evolutionary Anthropology