Comparing the Incomparable

Michael Cysouw
MPI-EVA Leipzig

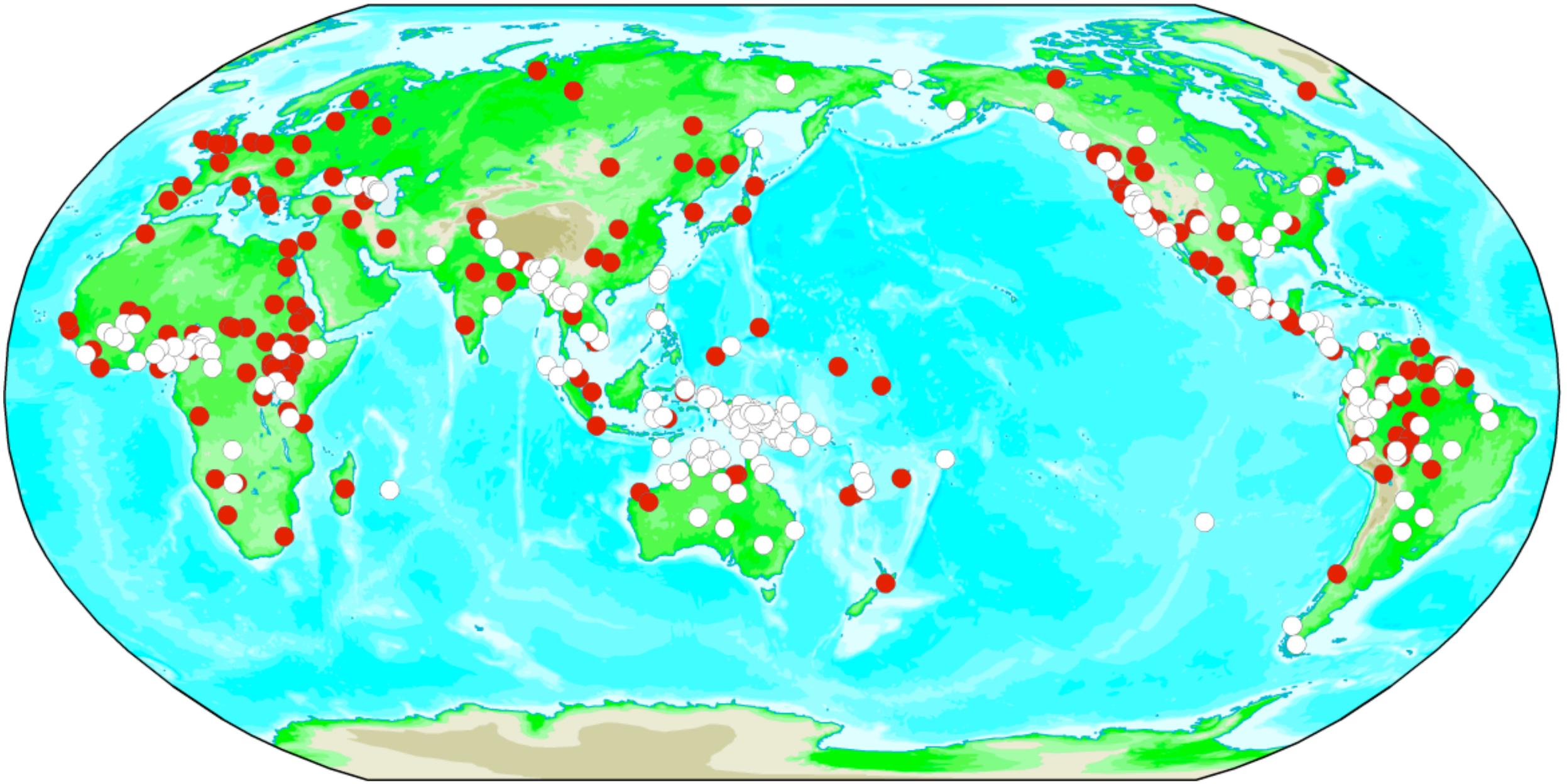
On the problem of comparing languages

Michael Cysouw
MPI-EVA Leipzig

goals

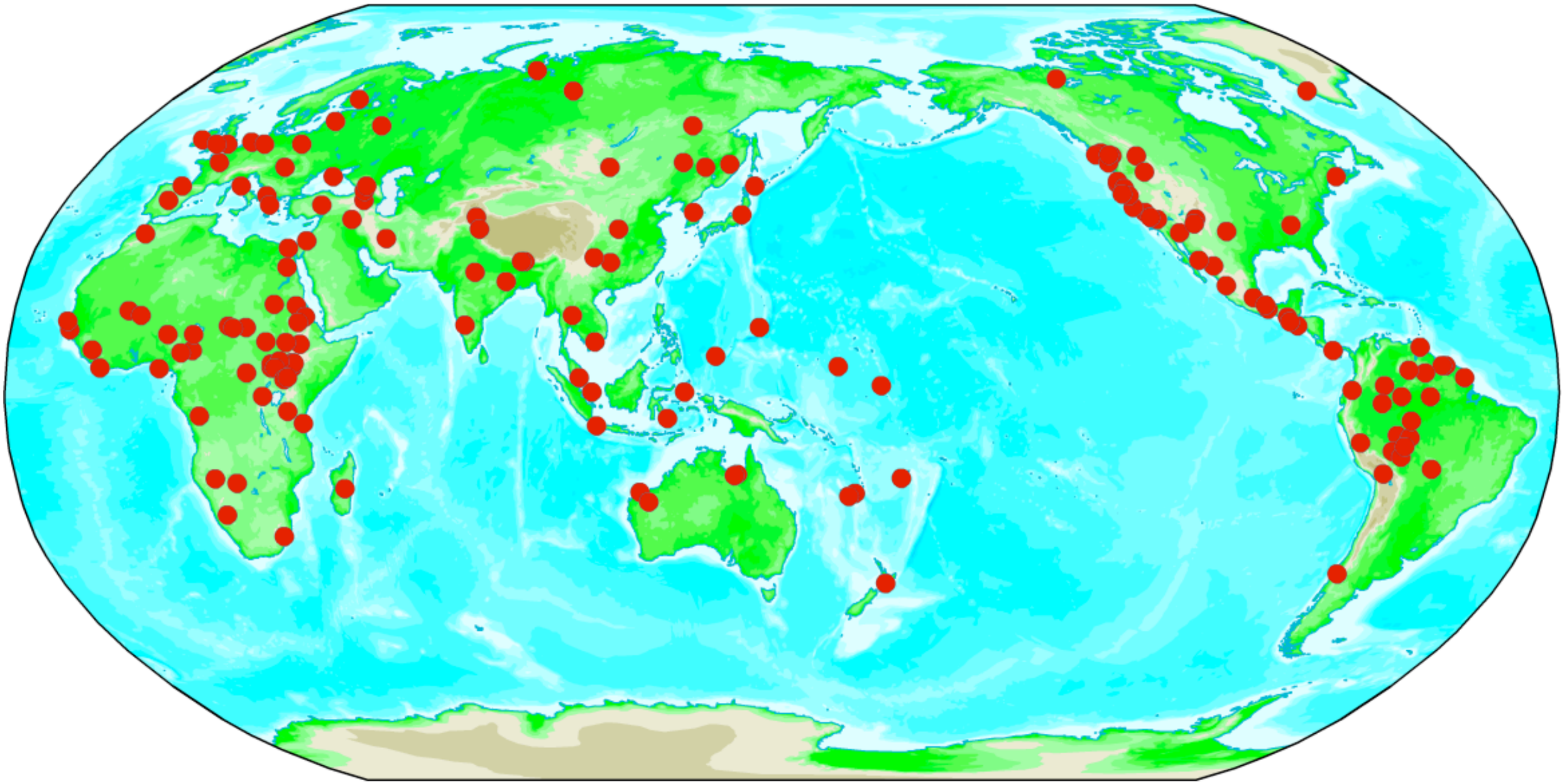
- Go beyond "simple" language typology
  - Not just Type A, B, C but full metric on languages
- Minimize and simplify comparative judgments
  - It is difficult to keep comparison constant
  - Relegate as much as possible to language-specific analysis
  - Speed up things, and allow for collaboration
- Allow for more data per language

Passive Constructions



Siewierska, Anna. "Passive Constructions." World Atlas of Language Structures. Eds. Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie. Oxford: Oxford University Press, 2005. 434-37.

Passive Constructions



Siewierska, Anna. "Passive Constructions." World Atlas of Language Structures. Eds. Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie. Oxford: Oxford University Press, 2005. 434-37.

"A construction has been classified as passive if it displays the following five properties:

- it contrasts with another construction, the active;
- the subject of the active corresponds to a non-obligatory oblique phrase of the passive or is not overtly expressed;
- the subject of the passive, if there is one, corresponds to the direct object of the active;
- the construction is pragmatically restricted relative to the active;
- the construction displays some special morphological marking of the verb."

Siewierska, Anna. "Passive Constructions." World Atlas of Language Structures. Eds. Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie. Oxford: Oxford University Press, 2005. 434-37.

Radical Relativism

- Constructions (including lexicon) are always language-specific
- In principle, each construction in each language should be uniquely named
- In practice, the same names are used again and again for reasons of readability
- This is currently confusing most readers (and authors!) of language comparisons

2.1.1. Emic independent clause classes

		Tr	Intr	Eq	Quot	
					Tr	Intr
		10	20	30	40	50
Decl	01	11	21	31	41	51
Ex	02	12	22	32	42	52
Q	03	13	23		43	53
Q	04	14	24	34	44	54
Impv	05	15	25		45	55
Impv	06	16	26		46	56

Chart II. Emic independent clause classes

2.1.2. Tagmemic independent clause formula

Cl = {±Mar: Cl mar +Nuc: Cl nuc 10-50}

The independent clause is subdivided into emic distribution classes 11-56 on the basis of the filler of the nucleus slot and of the distribution. Since the distribution classes do not otherwise differ in composition, they are not shown in separate formulas.

2.1.3. Independent clause citation

Ind cl = noy teč in neríyow--noropików to neč ka?ánoneb
'there that water they-drink-where--they-just-now-come the those animals' (those animals were just now going there where they drink that water).

kopi íkomórikon to mónči 'why-you killer the child' (why did you kill the child?).

ne soratíye--ónka to ka kí?inon 'here town-in--not the that-which caring-person' (here in the town there is no one who cares).

nihín--ónka ímokon 'my-daughter--not sleeper' (my daughter, don't sleep).

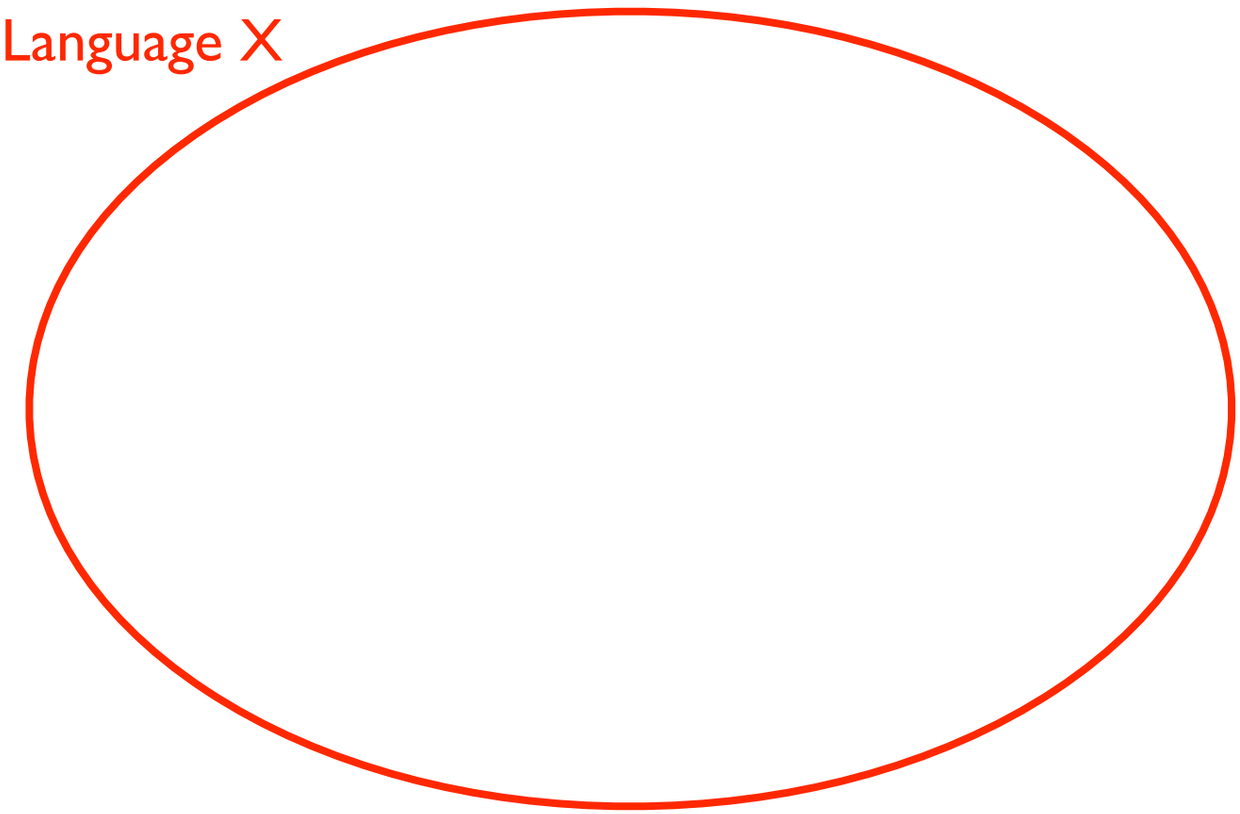
How to compare unique constructions across languages?

- Similarity Semantics: no identity, only similarity
- Meaning/function is a continuous space, without universal semantic meta-language
- The function-space can be sampled
- Similarity of constructions can be established based on this sample

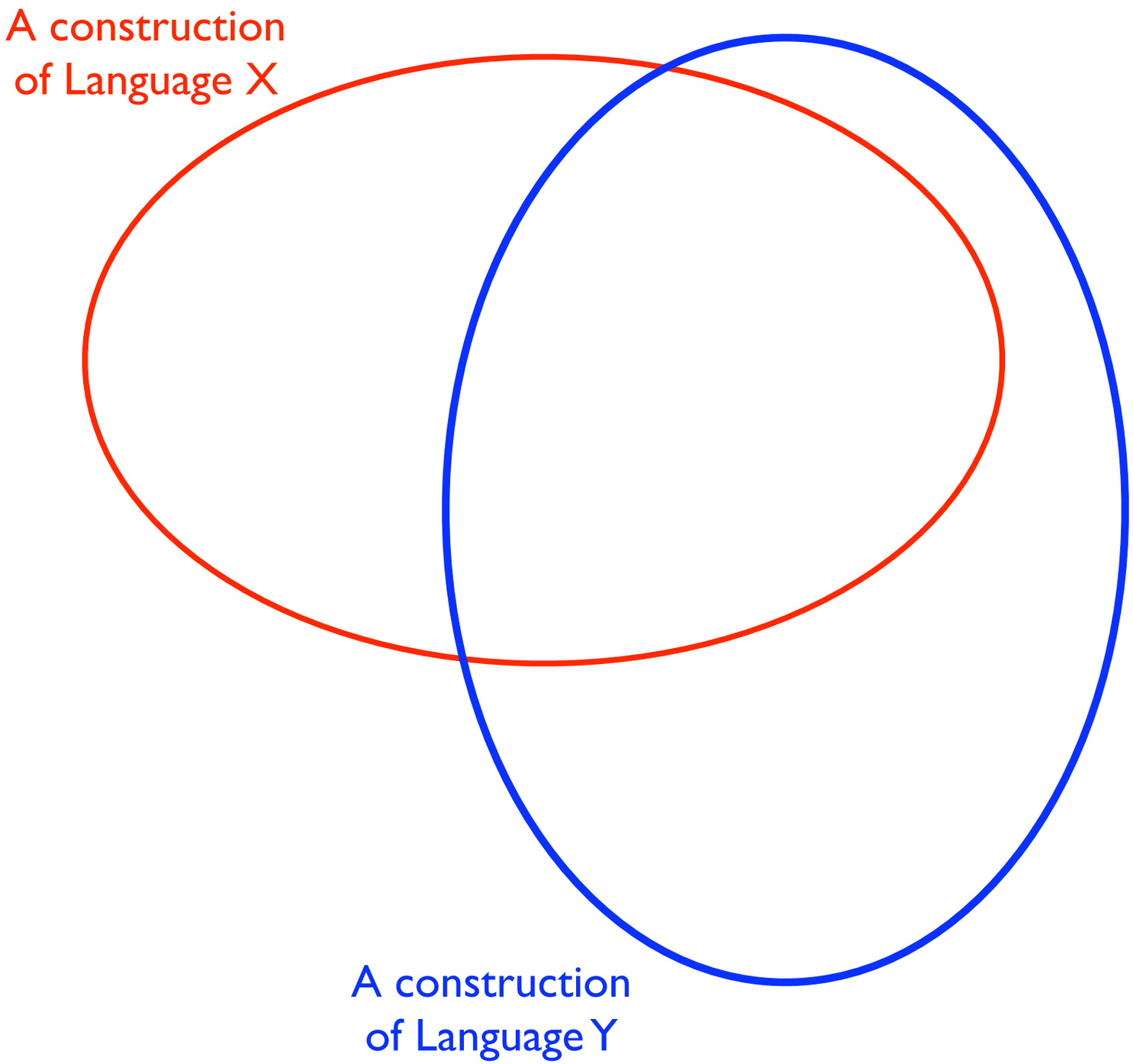
Meaning/Function-space

Meaning/Function-space

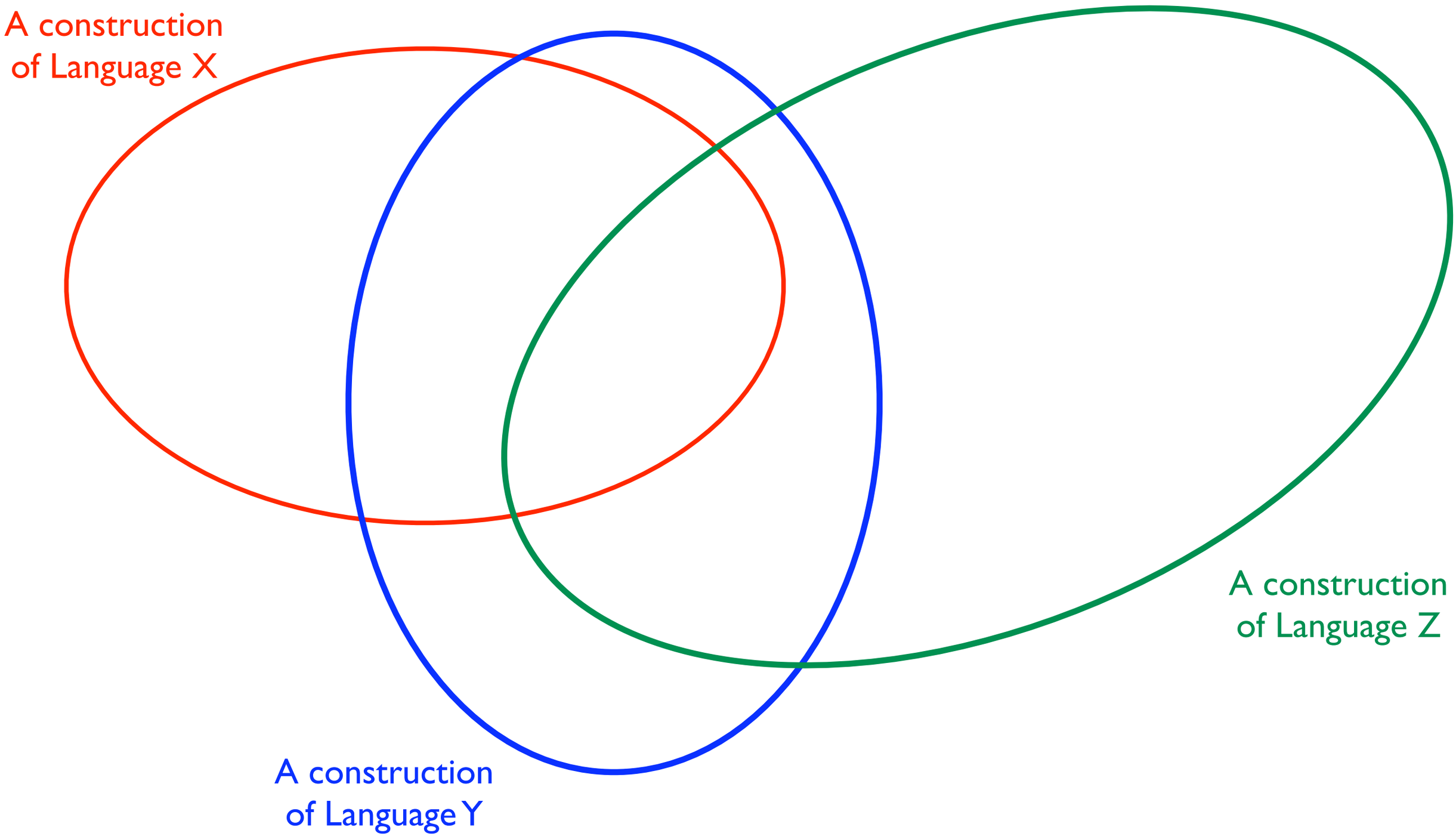
A construction
of Language X



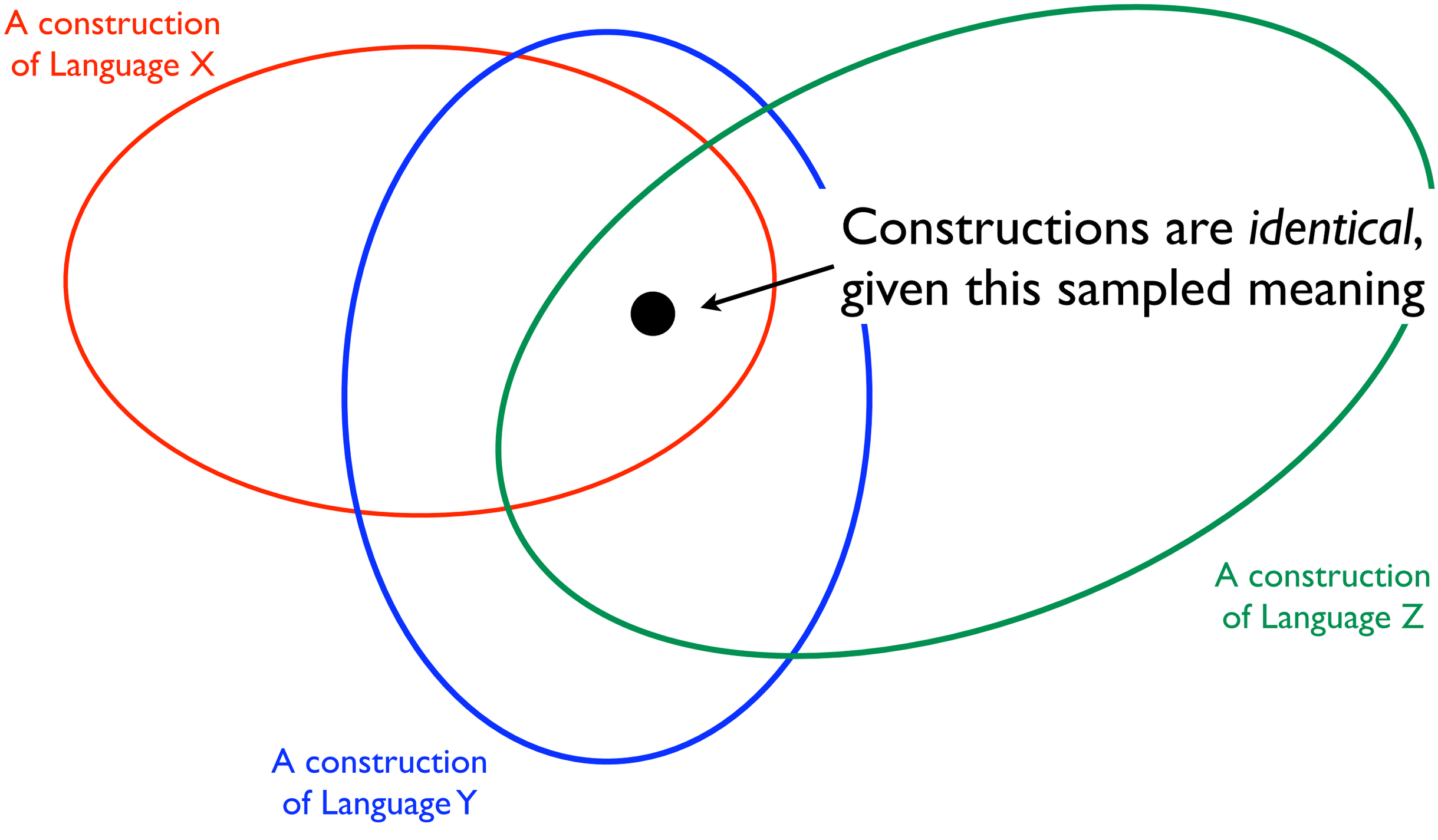
Meaning/Function-space



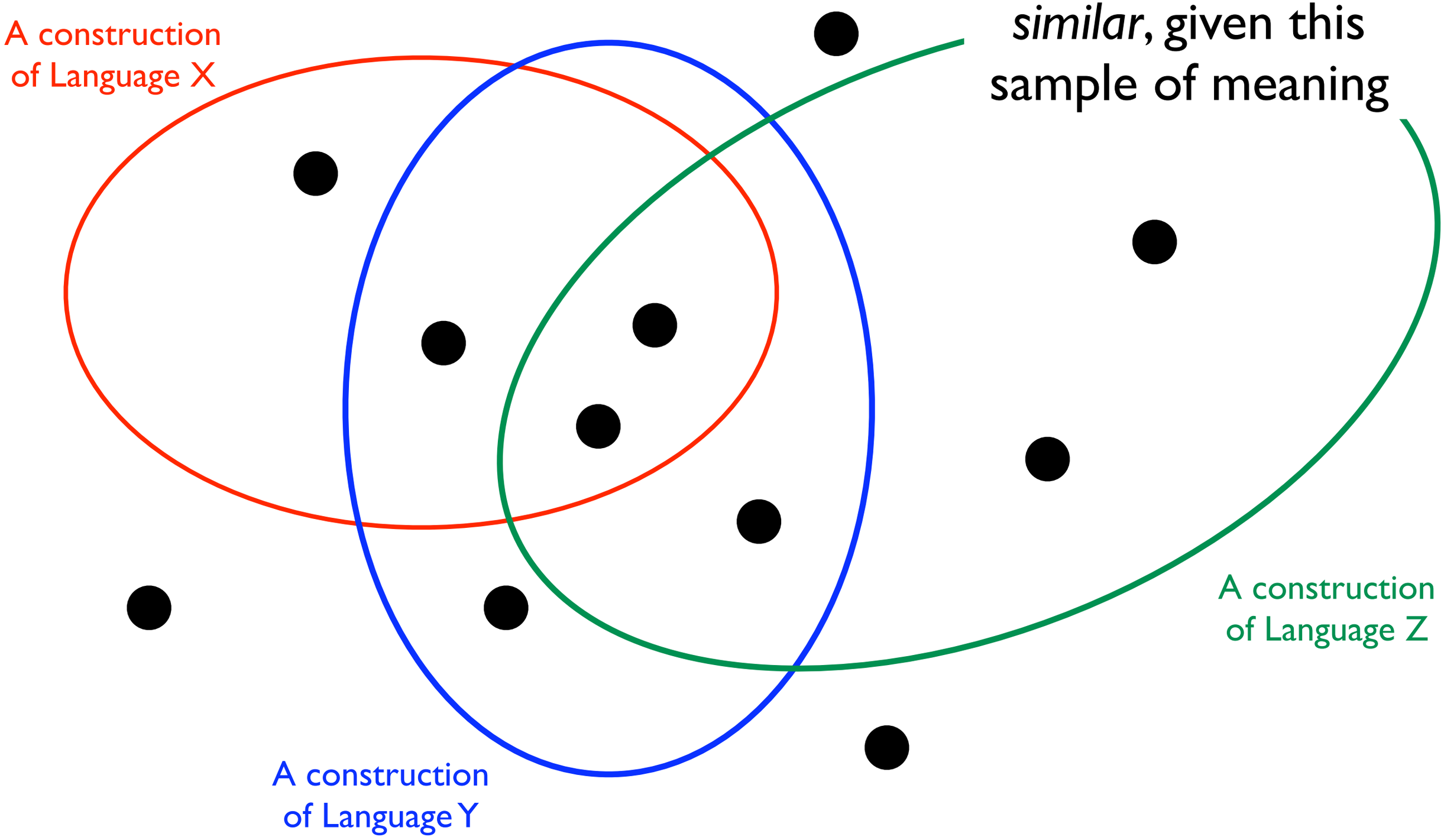
Meaning/Function-space



Meaning/Function-space

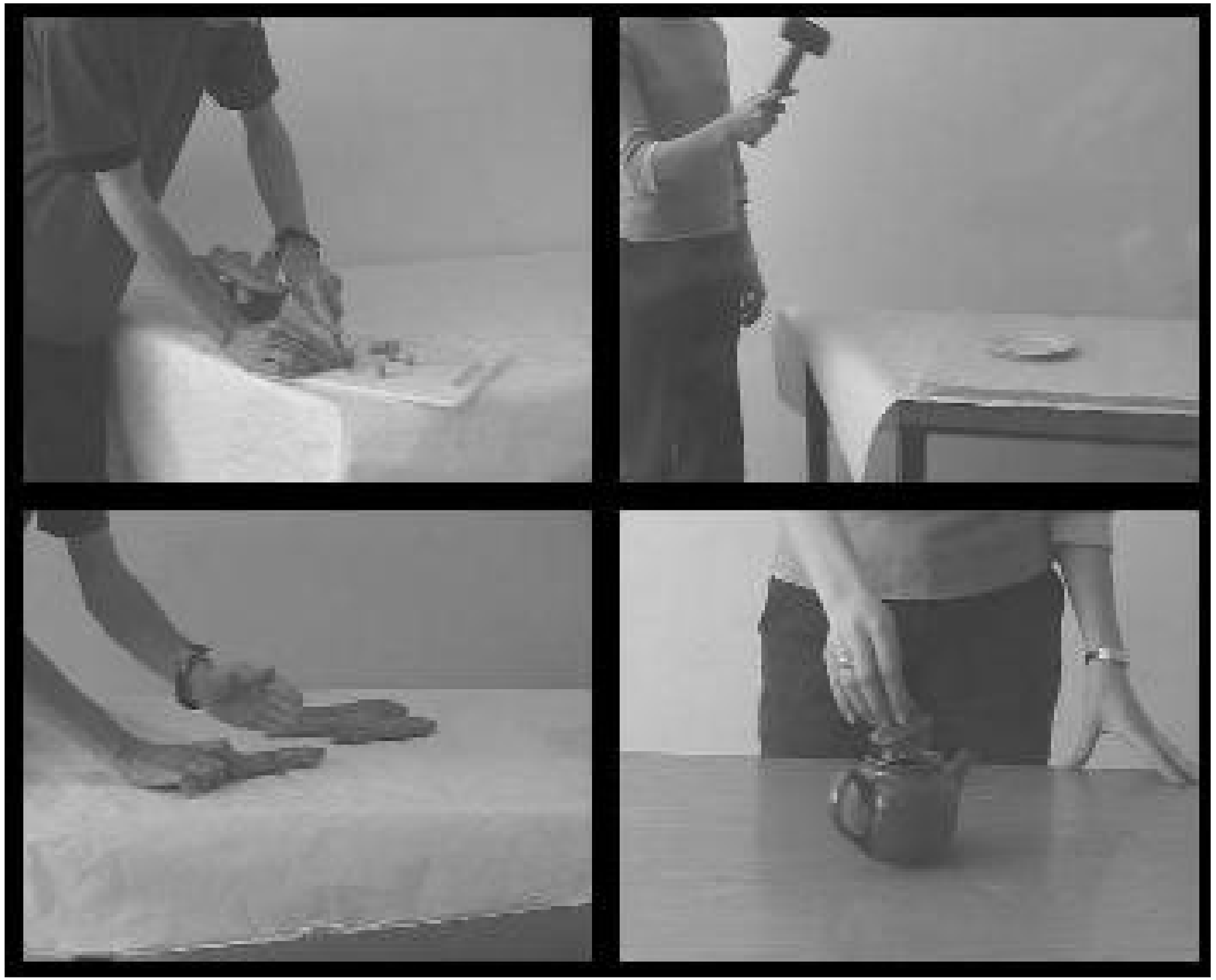


Meaning/Function-space



Sampling Meaning

- The meaning/function-space can be sampled by collecting *contextually situated expressions*
  - items in parallel texts
  - pictures, videos
  - translational questionnaires
  - (more abstract) "functions"
- Choice of contexts is not given, but depends on *theoretical question*



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

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.

“Constructions”

- Compare languages based on
  concrete expressions in context
- Compare similarity between expressions
  - within each language (“constructions”)
  - between languages (“strategies”)

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.'

Haspelmath, Martin (1997)
Indefinite Pronouns. Oxford: OUP.

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

somebody, someone

6. indirect negations
7. direct negation
8. standard of comparison
9. free choice

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
7. direct negation
8. standard of comparison
9. free choice

somebody, someone

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
7. direct negation
8. standard of comparison
9. free choice

somebody, someone

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

någon

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
7. direct negation
8. standard of comparison
9. free choice

somebody, someone

någon

jemand

aliquis

somebody, someone

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

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

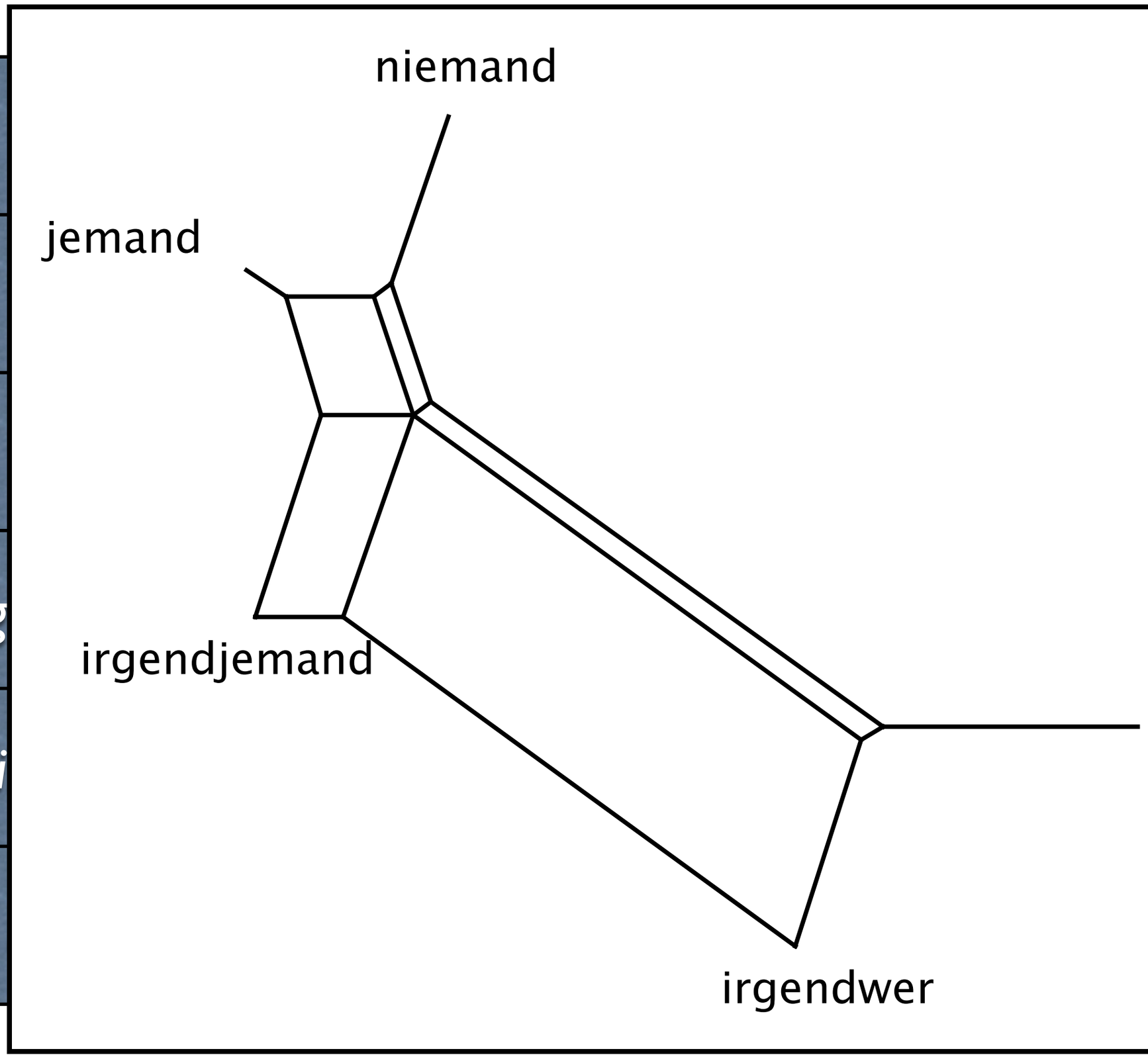
8. standard of comparison
9. free choice

irgentjemand, irgentwer

Language-specific dissimilarities

	wer	*jemand*	*irgend-jemand*	*irgent-wer*	*niemand*
wer	0	0.75	0.79	0.33	0.76
jemand	0.75	0	0.25	0.72	0.23
irgendjemand	0.79	0.25	0	0.42	0.43
irgendwer	0.33	0.72	0.42	0	0.75
niemand	0.76	0.23	0.43	0.75	0

Language-specific dissimilarities



				nand
				76
				23
	irg			43
	i			75
				0

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
jemand	X	X	X	X	X	X			
irgendjemand		X	X	X	X	X		X	X
irgendwer		X	X	X	X	X		X	X
niemand							X		

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
jemand	10	45	7	4	19	2			
irgendjemand		3	45	7	18	9		2	12
irgendwer		14	36	12	8	45		35	1
niemand							20		

token-perspective

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
jemand	X	X	X	X	X	X			
irgendjemand		X	X	X	X	X		X	X
irgendwer		X	X	X	X	X		X	X
niemand							X		

type-perspective

	jemand	*irgend-jemand*	*irgent-wer*	*niemand*
jemand	0	0.25	0.72	0.23
irgendjemand	0.25	0	0.42	0.43
irgendwer	0.72	0.42	0	0.75
niemand	0.23	0.43	0.75	0

	1	2	3	4	5	6	7	8	9
jemand	X	X	X	X	X	X			
irgendjemand		X	X	X	X	X		X	X
irgendwer		X	X	X	X	X		X	X
niemand							X		

	jemand	*irgend-jemand*	*irgent-wer*	*niemand*
jemand	0	0.25	0.72	0.23
irgendjemand	0.25	0	0.42	0.43
irgendwer	0.72	0.42	0	0.75
niemand	0.23	0.43	0.75	0

	1	2	3	4	5	6	7	8	9
jemand	X	X	X	X	X	X			
irgendjemand		X	X	X	X	X		X	X
irgendwer		X	X	X	X	X		X	X
niemand							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

	jemand	*irgend-jemand*	*irgent-wer*	*niemand*
jemand	0	0.25	0.72	0.23
irgendjemand	0.25	0	0.42	0.43
irgendwer	0.72	0.42	0	0.75
niemand	0.23	0.43	0.75	0

	1	2	3	4	5	6	7	8	9
jemand	X	X	X	X	X	X			
irgendjemand		X	X	X	X	X		X	X
irgendwer		X	X	X	X	X		X	X
niemand							X		

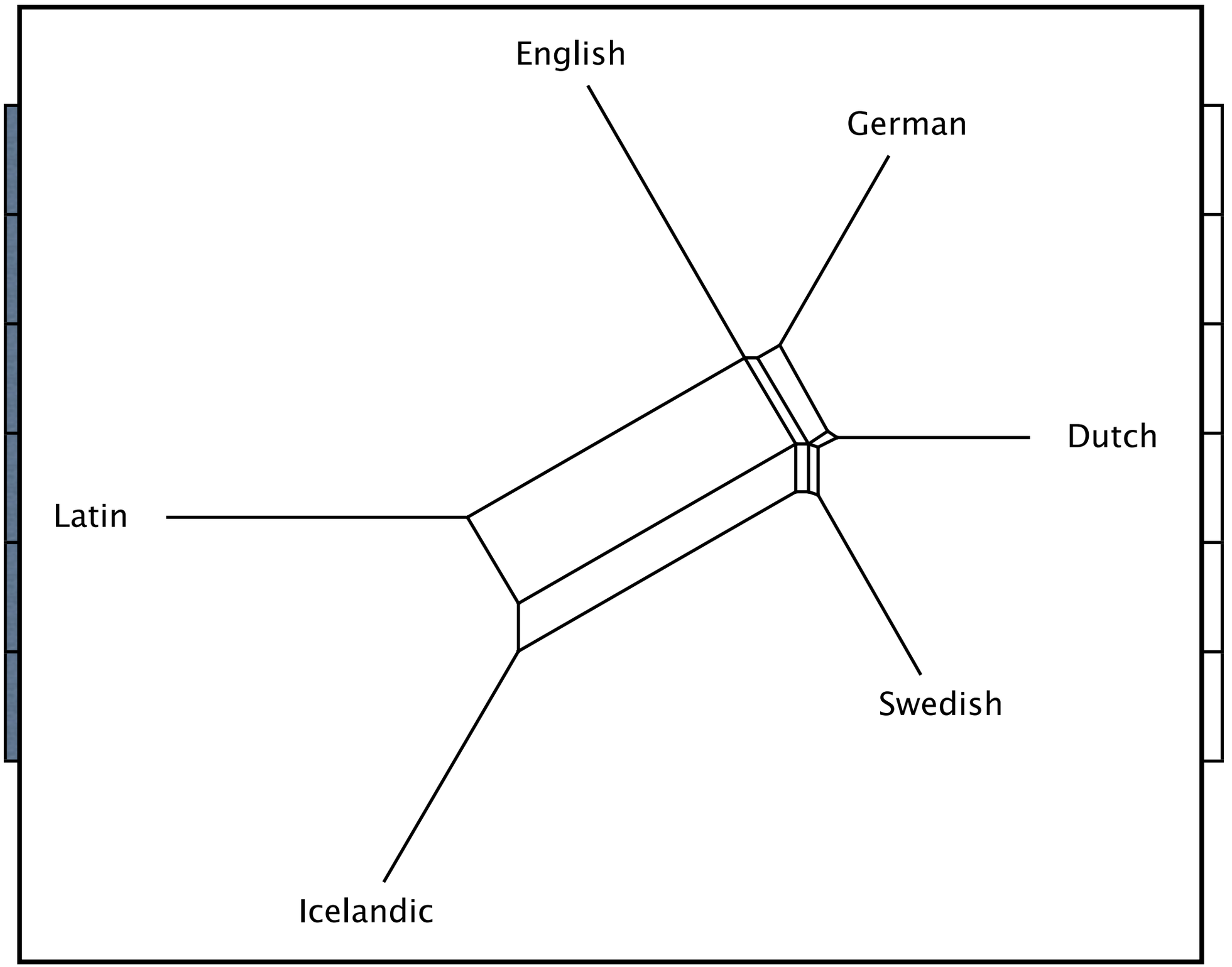
	1	2	3	4	5	6	7	8	9
1	0	0.36	0.36	0.36	0.36	0.36	0.37	0.57	0.57
2	0.36	0	0	0	0	0	0.55	0.09	0.09
3	0.36	0	0	0	0	0	0.55	0.09	0.09
4	0.								09
5	0.								09
6	0.36	0	0	0	0	0	0.55	0.09	0.09
7	0.37	0.55	0.55	0.55	0.55	0.55	0	0.69	0.69
8	0.57	0.09	0.09	0.09	0.09	0.09	0.69	0	0
9	0.57	0.09	0.09	0.09	0.09	0.09	0.69	0	0

Language-specific perspective on sampled functions

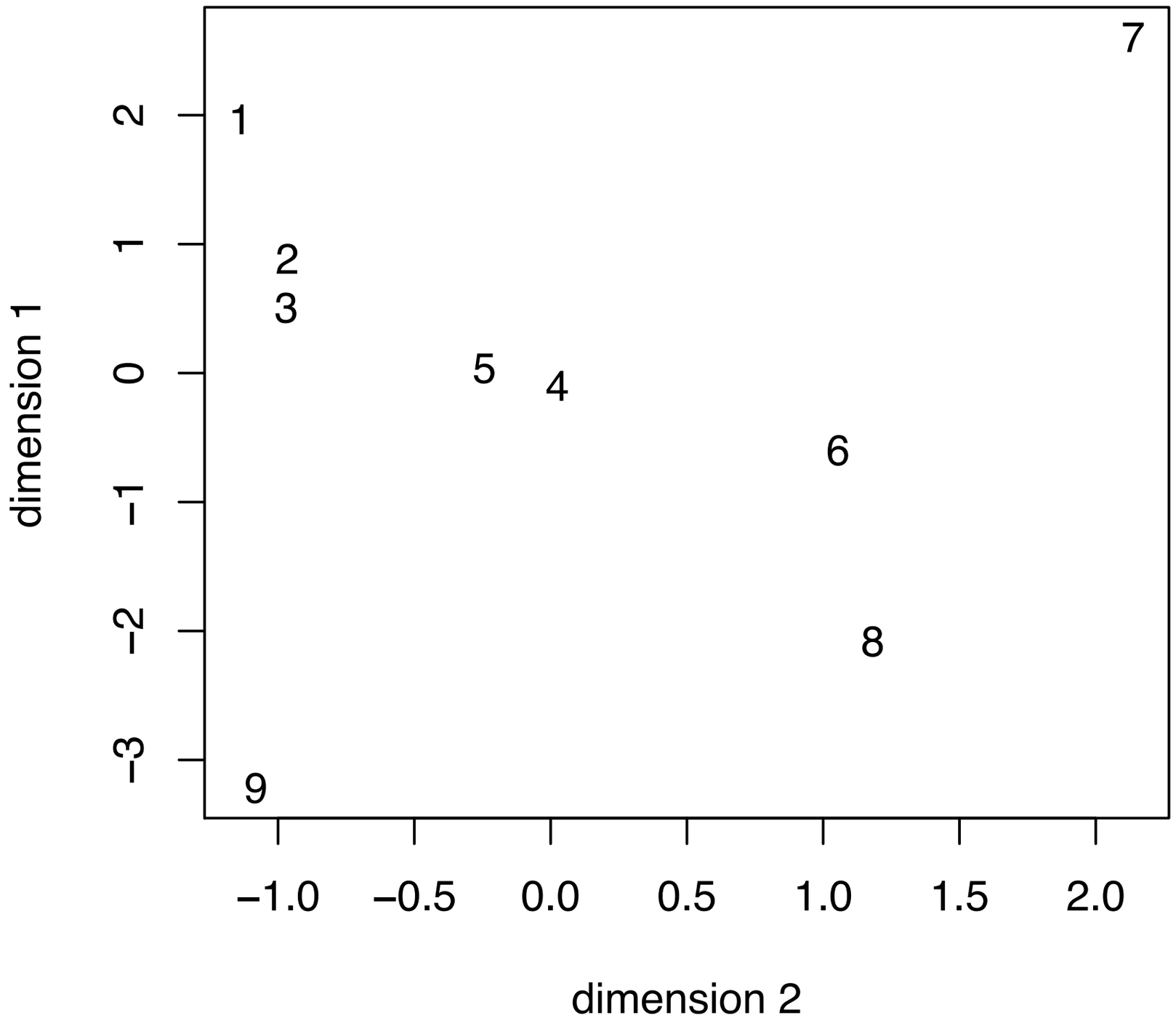
correlate language-specific perspectives with each other …

Dutch	0	0.53	0.40	0.67	0.71	0.36
English	0.53	0	0.44	0.70	0.82	0.50
German	0.40	0.44	0	0.86	0.57	0.47
Icelandic	0.67	0.70	0.86	0	0.55	0.63
Latin	0.71	0.82	0.57	0.55	0	0.80
Swedish	0.36	0.50	0.47	0.63	0.80	0

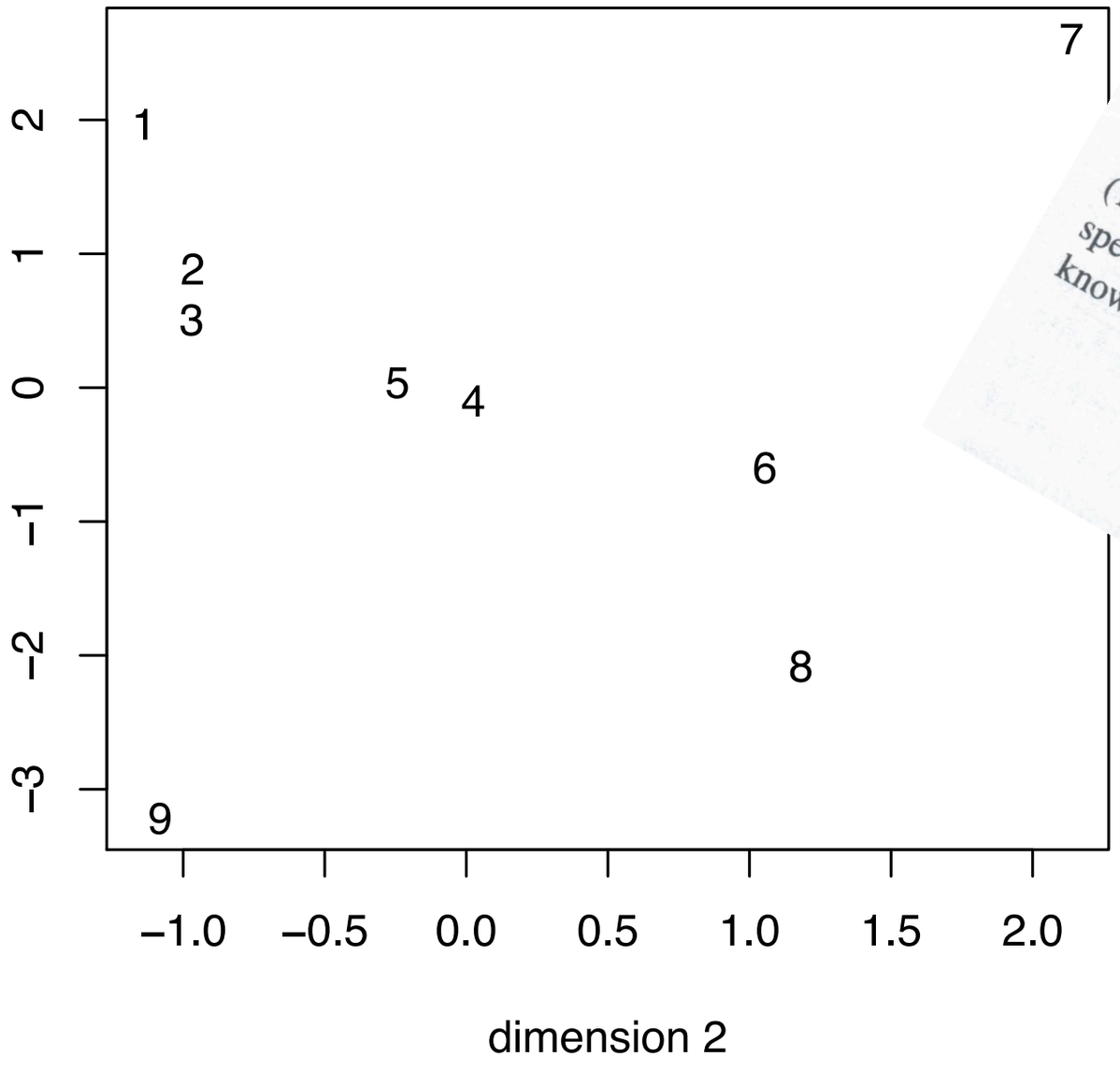
correlate language-specific perspectives with each other …

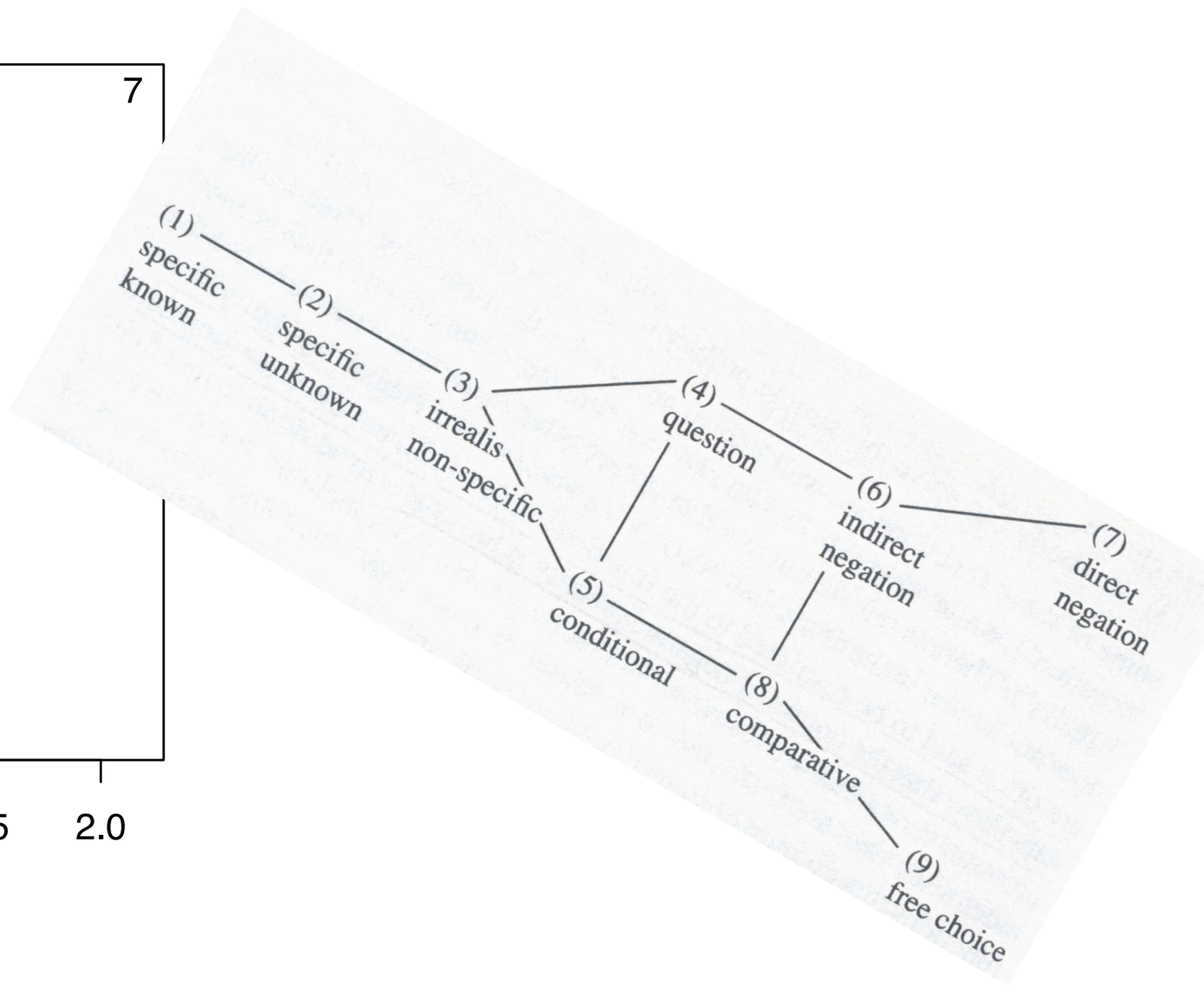


add language-specific perspectives together …



add language-specific perspectives together …





Interim Summary

- Start with language-specific analysis:
  - Establish language-specific similarity between expressions
  - Establish mapping of expressions to sample of functions
  - Combine these to obtain the language-specific perspective on the sample of functions
- Adding up language-specific perspectives results in metric on meaning
- Correlating language-specific perspectives results in a language typology