Michael Cysouw
MPI-EVA Leipzig

Language Comparison as Corpus Linguistics

Michael Cysouw
MPI-EVA Leipzig

goals

goals

- Go beyond “simple” language typology
  - Not just Type A, B, C but full metric on languages

goals

- Go beyond "simple" language typology
  - Not just Type A, B, C but full metric on languages
- Minimize and simplify comparative judgments

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

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

goals

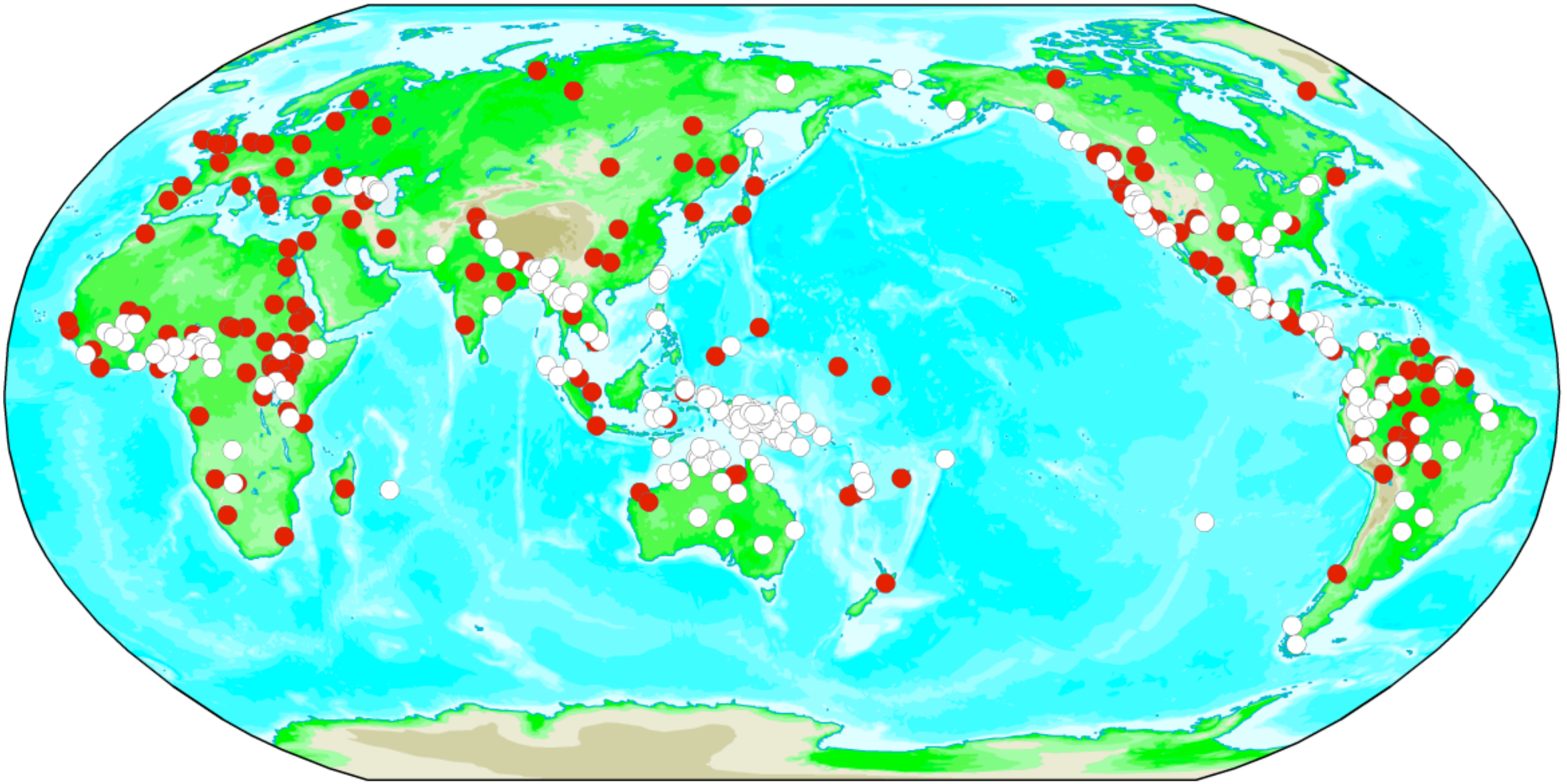
- Go beyond "simple" language typology
  - Not just Type A, B, C but full metric on languages
- Minimize and simplify comparative judgments
  - It is difficult to keep comparison constant
  - Relegate as much as possible to language-specific analysis
  - Speed up things, and allow for collaboration

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

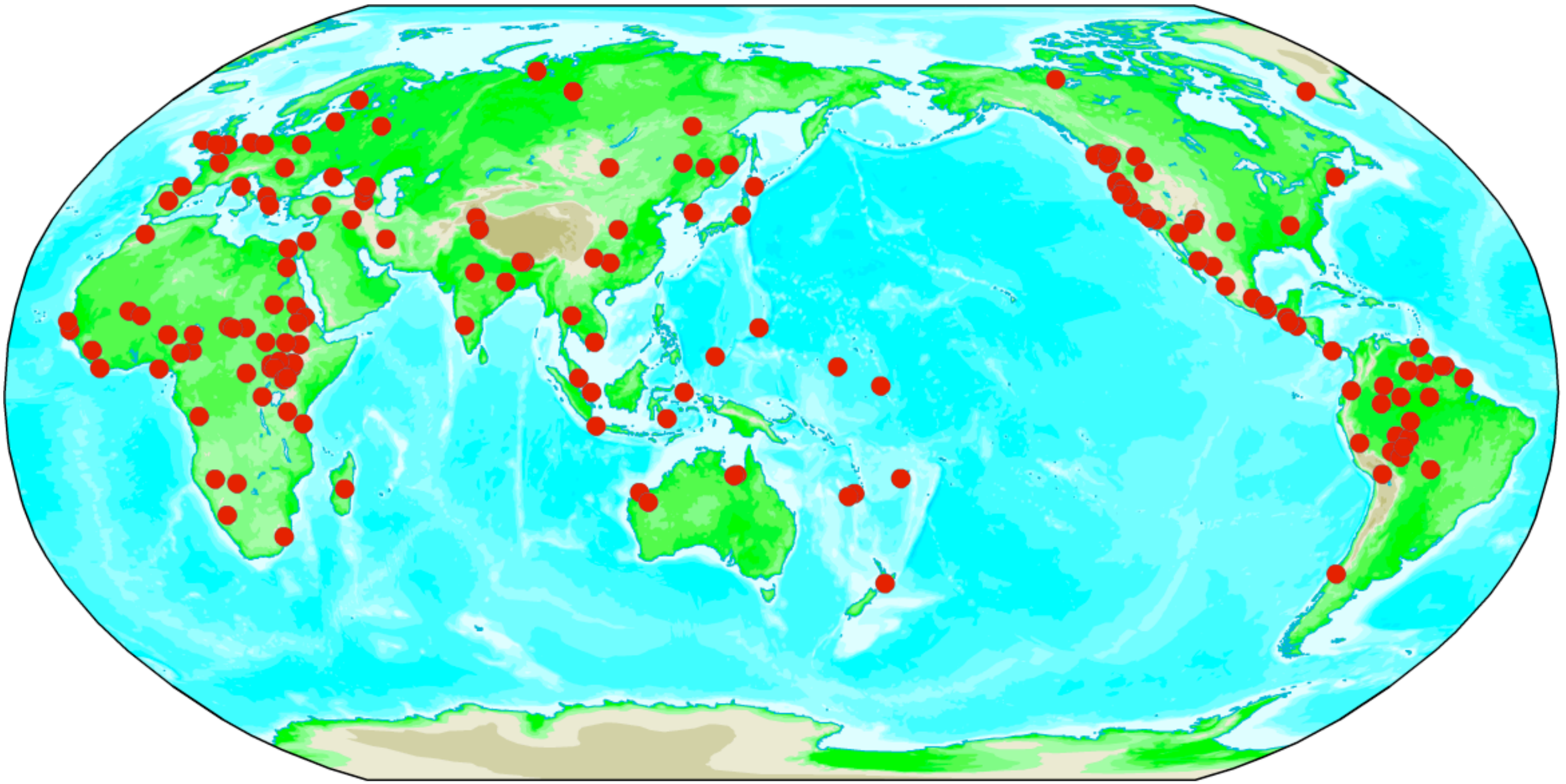
Passive Constructions

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:

"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

Radical Relativism

- Constructions (including lexicon) are always language-specific

Radical Relativism

- Constructions (including lexicon) are always language-specific
- In principle, each construction in each language should be uniquely named

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

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?

How to compare unique constructions across languages?

- Similarity Semantics: no identity, only similarity

How to compare unique constructions across languages?

- Similarity Semantics: no identity, only similarity
- Meaning/function is a continuous space, without universal semantic meta-language

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

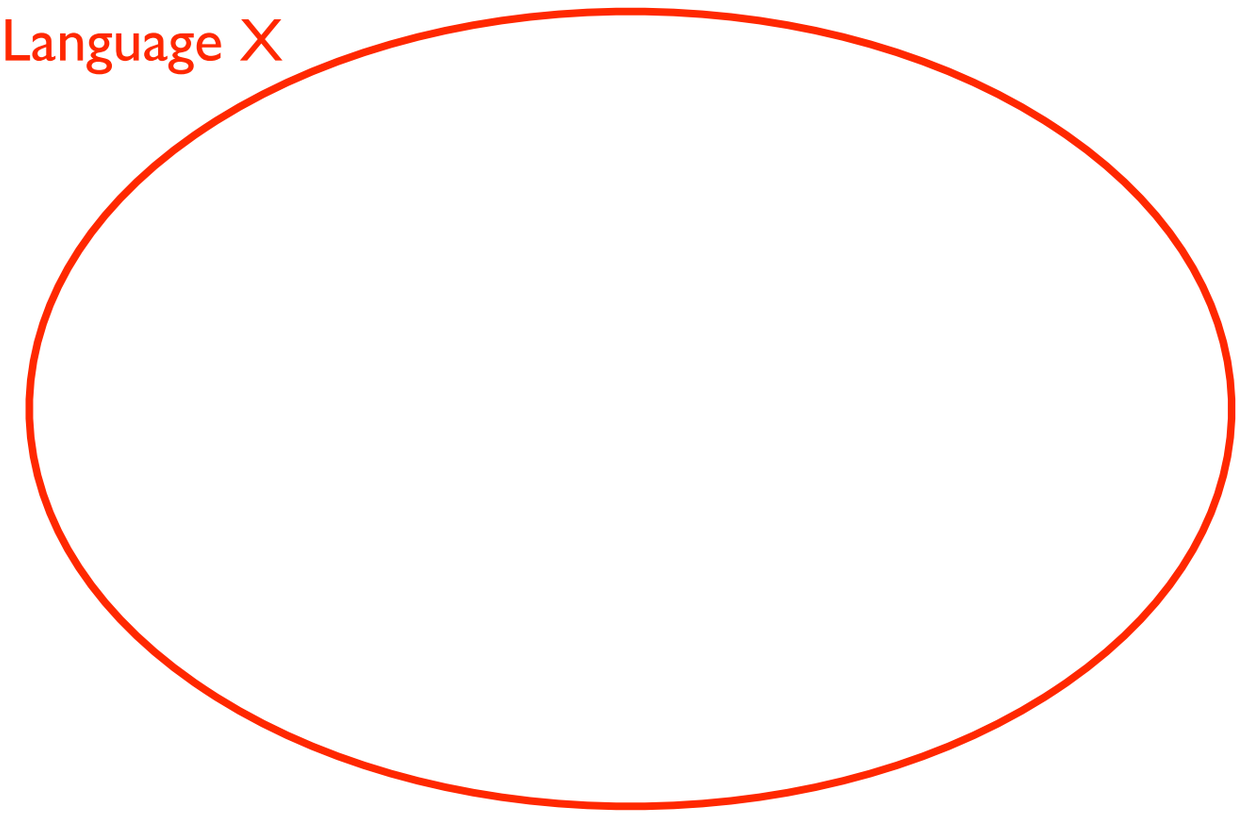
How to compare unique constructions across languages?

- 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

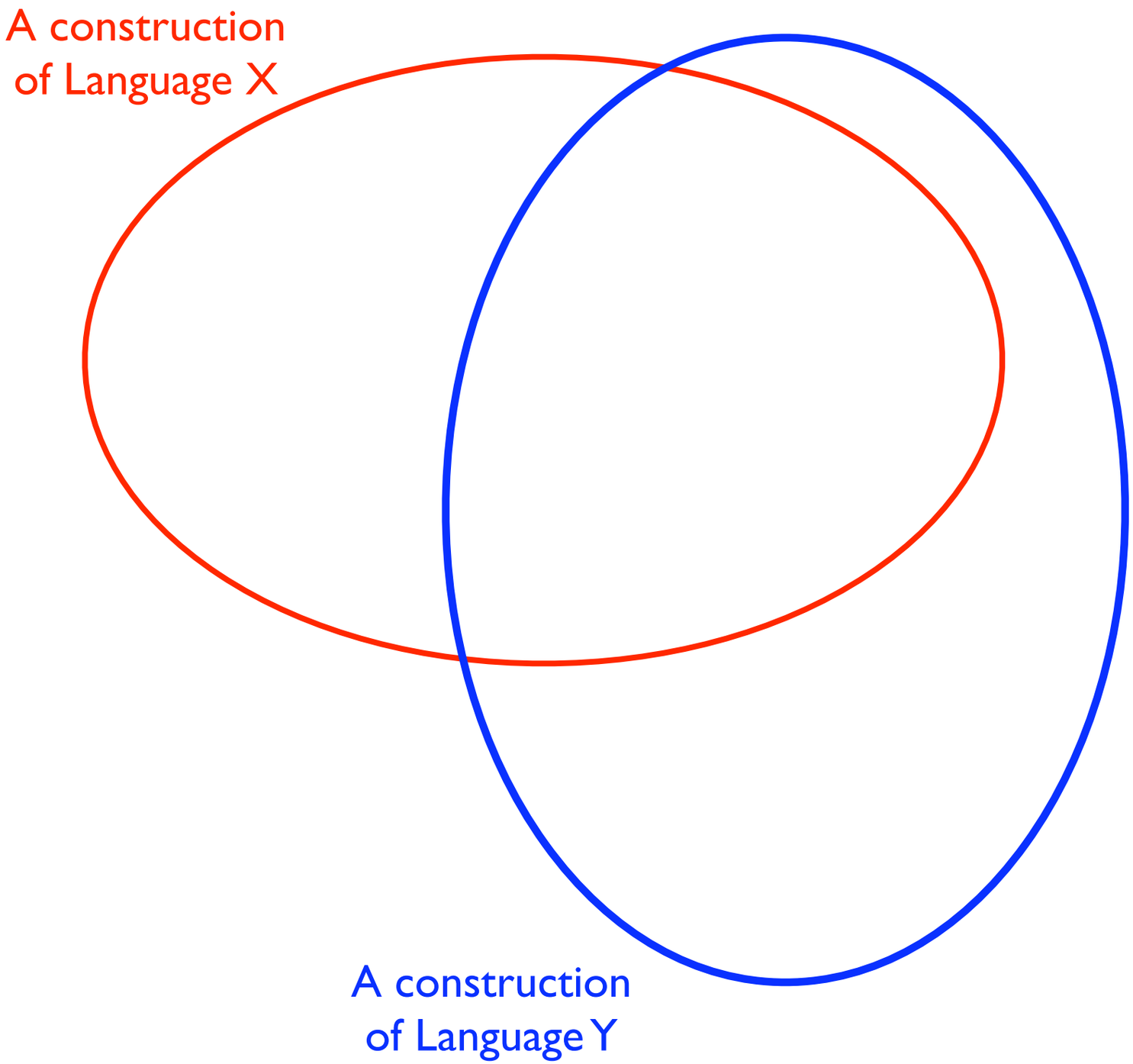
Meaning/Function-space

Meaning/Function-space

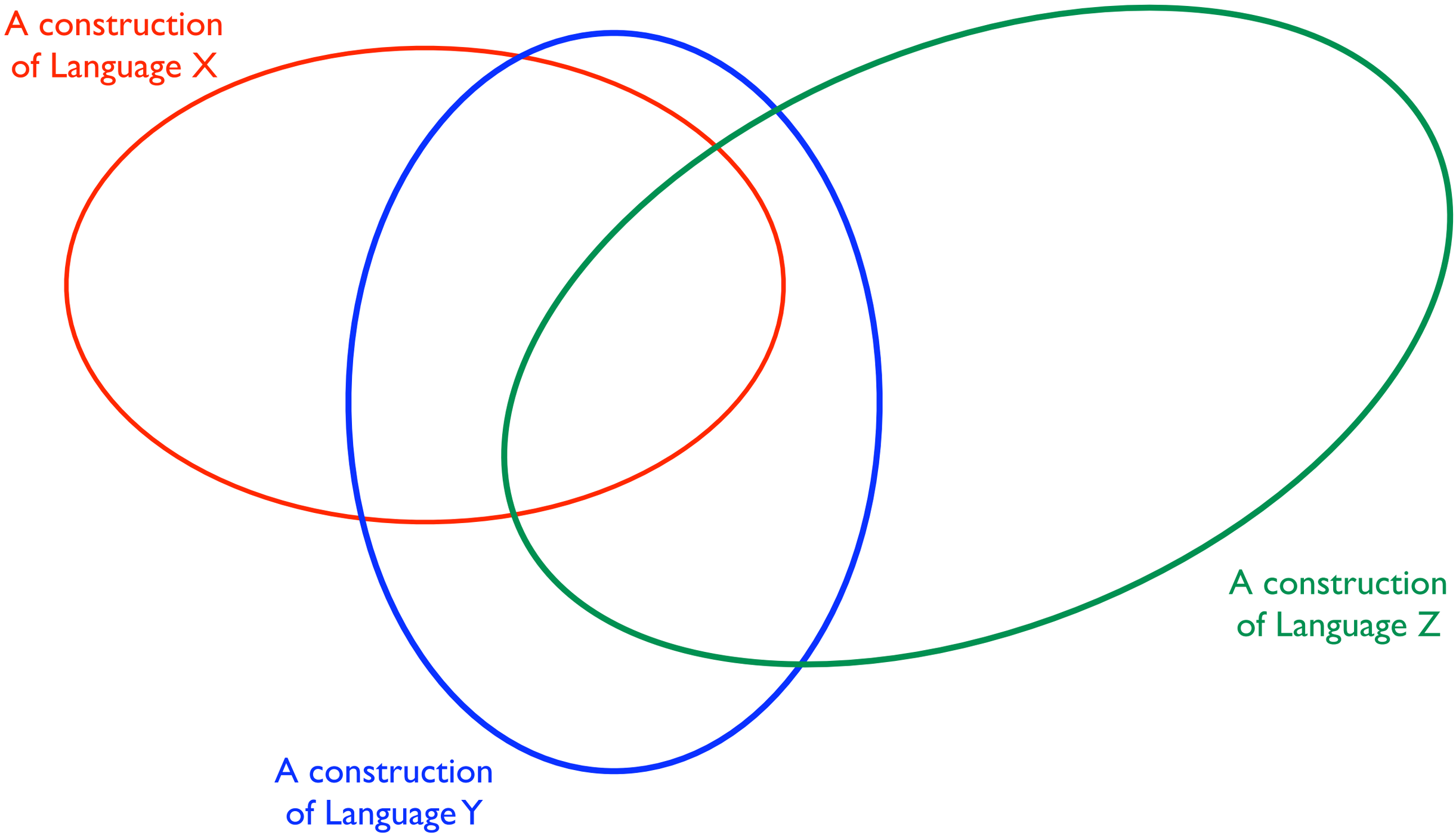
A construction
of Language X



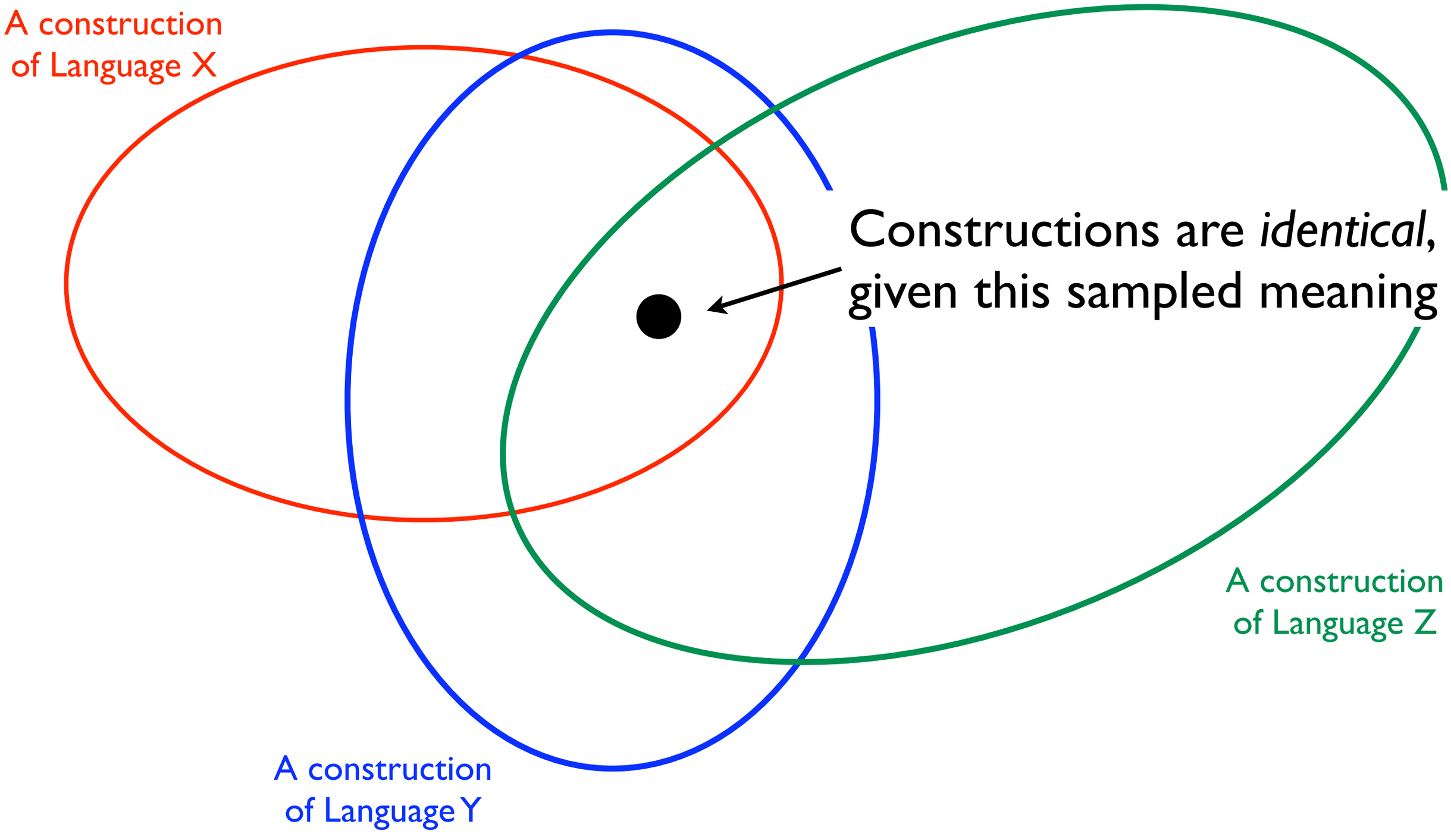
Meaning/Function-space



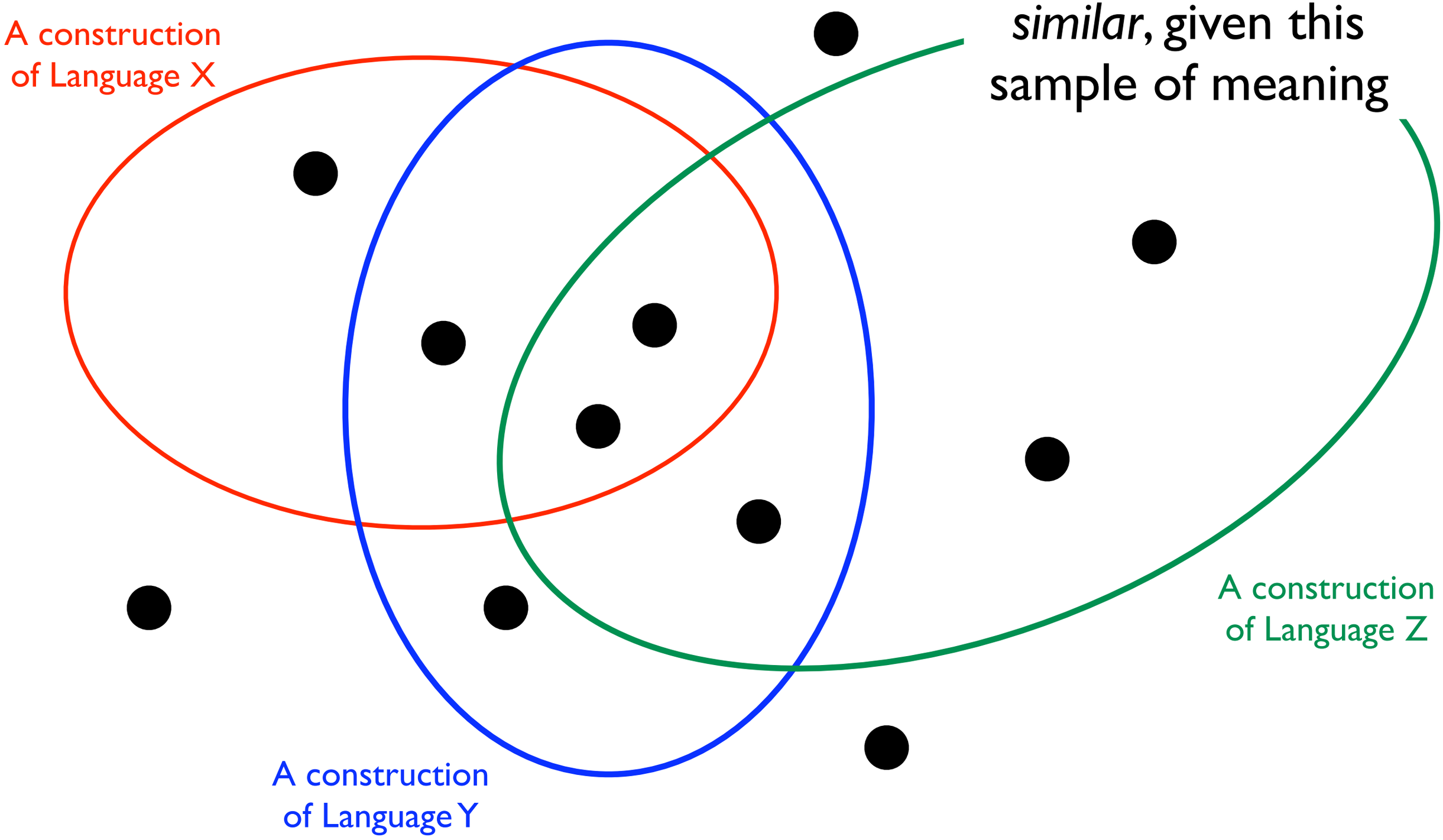
Meaning/Function-space



Meaning/Function-space



Meaning/Function-space



Sampling Meaning

Sampling Meaning

- The meaning/function-space can be sampled by collecting *contextually situated expressions*

Sampling Meaning

- The meaning/function-space can be sampled by collecting *contextually situated expressions*
  - items in parallel texts

Sampling Meaning

- The meaning/function-space can be sampled by collecting *contextually situated expressions*
  - items in parallel texts
  - pictures, videos

Sampling Meaning

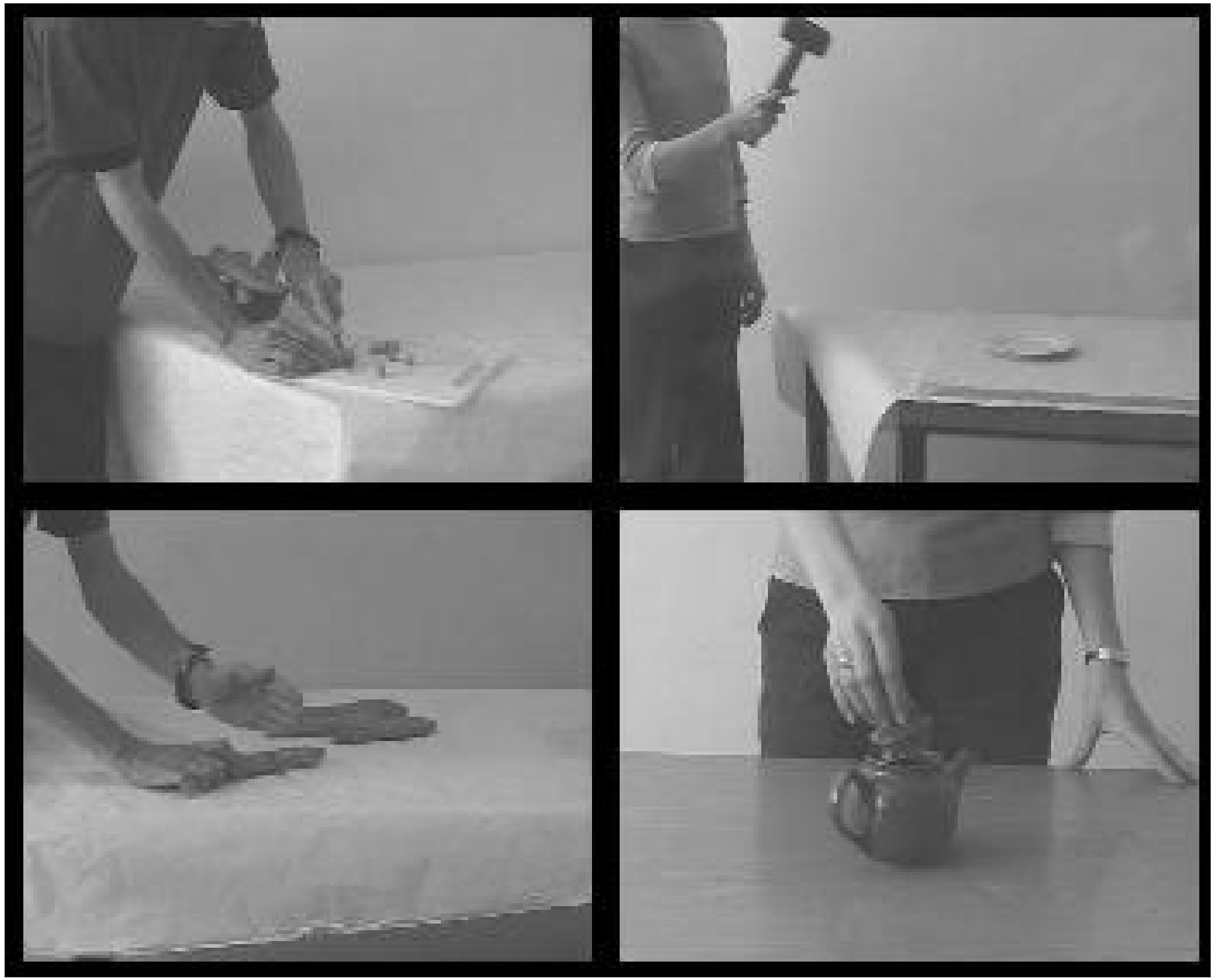
- The meaning/function-space can be sampled by collecting *contextually situated expressions*
  - items in parallel texts
  - pictures, videos
  - translational questionnaires

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”

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.

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.

Comparing expressions

Comparing expressions

- Compare languages based on *concrete expressions in context*

Comparing expressions

- Compare languages based on *concrete expressions in context*
- Compare similarity between expressions

Comparing expressions

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

Comparing expressions

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

Comparing expressions

- Compare languages based on *concrete expressions in context*
- Compare similarity between expressions
  - within each language (“constructions”)
  - between languages (“strategies”)
- Combine expressions per language, genus, or other factor for higher-order comparison

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)

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

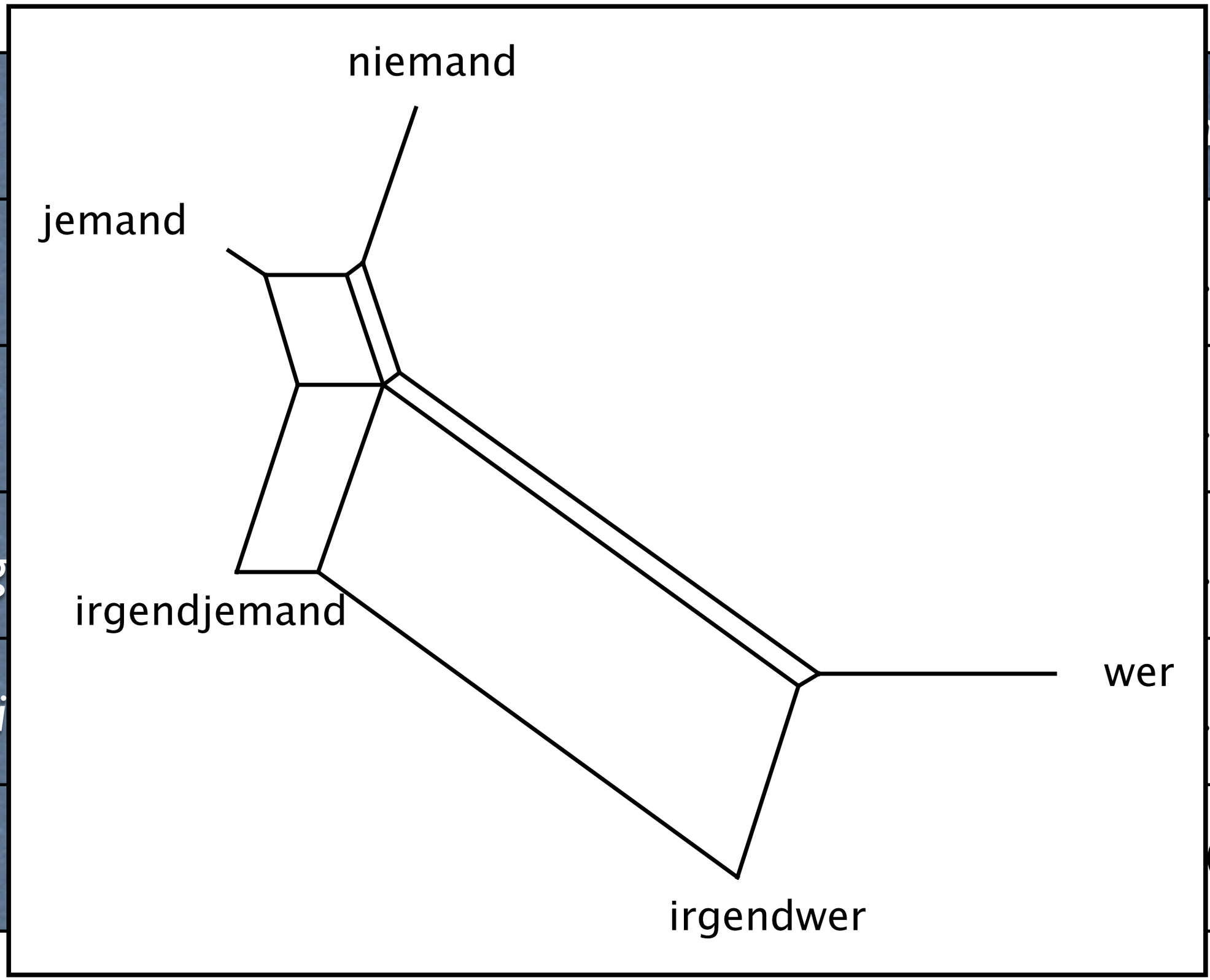
8. standard of comparison
9. free choice

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



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

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
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

	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

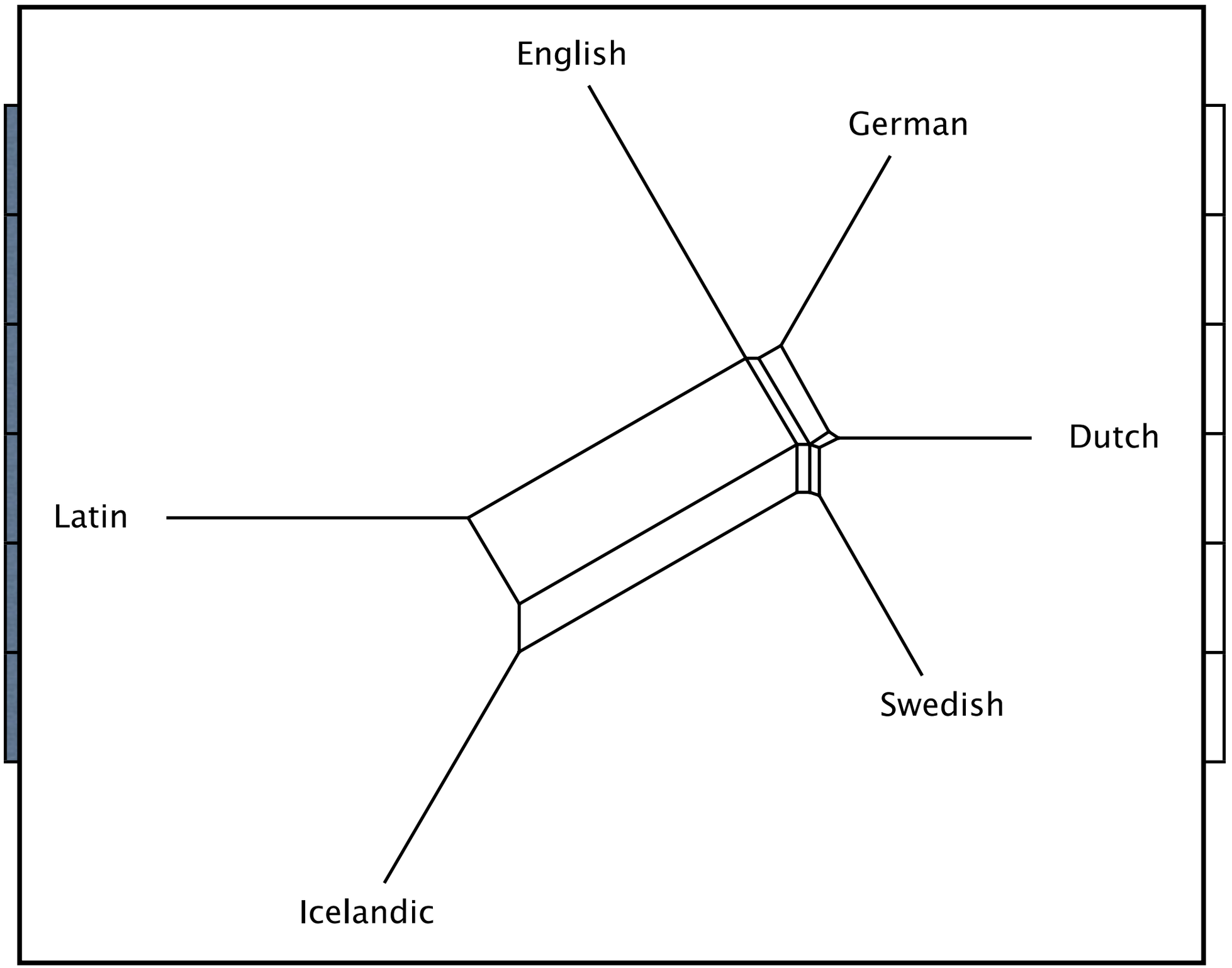
Language-specific perspective on sampled functions

correlate language-specific perspectives
with each other …

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 …

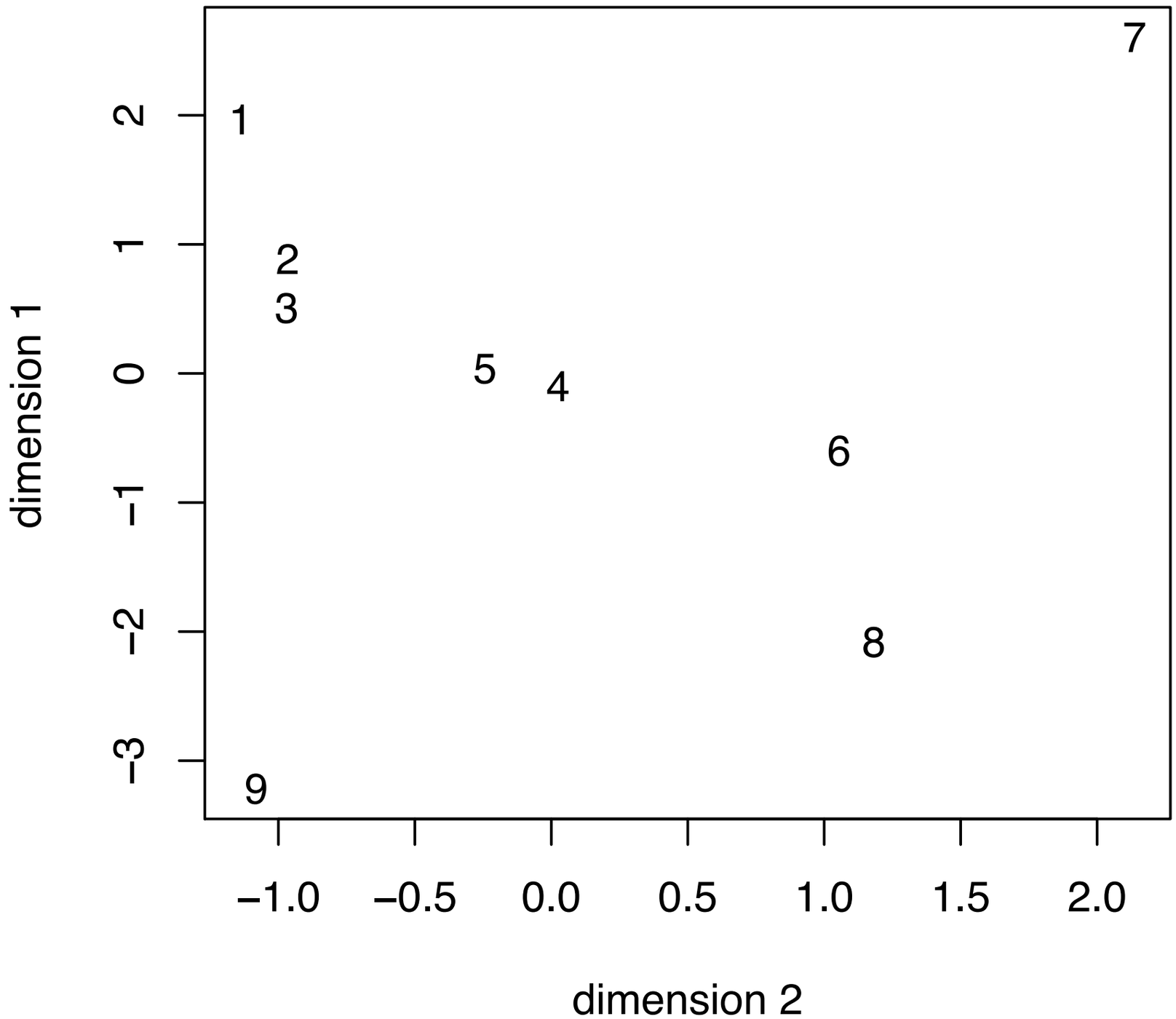


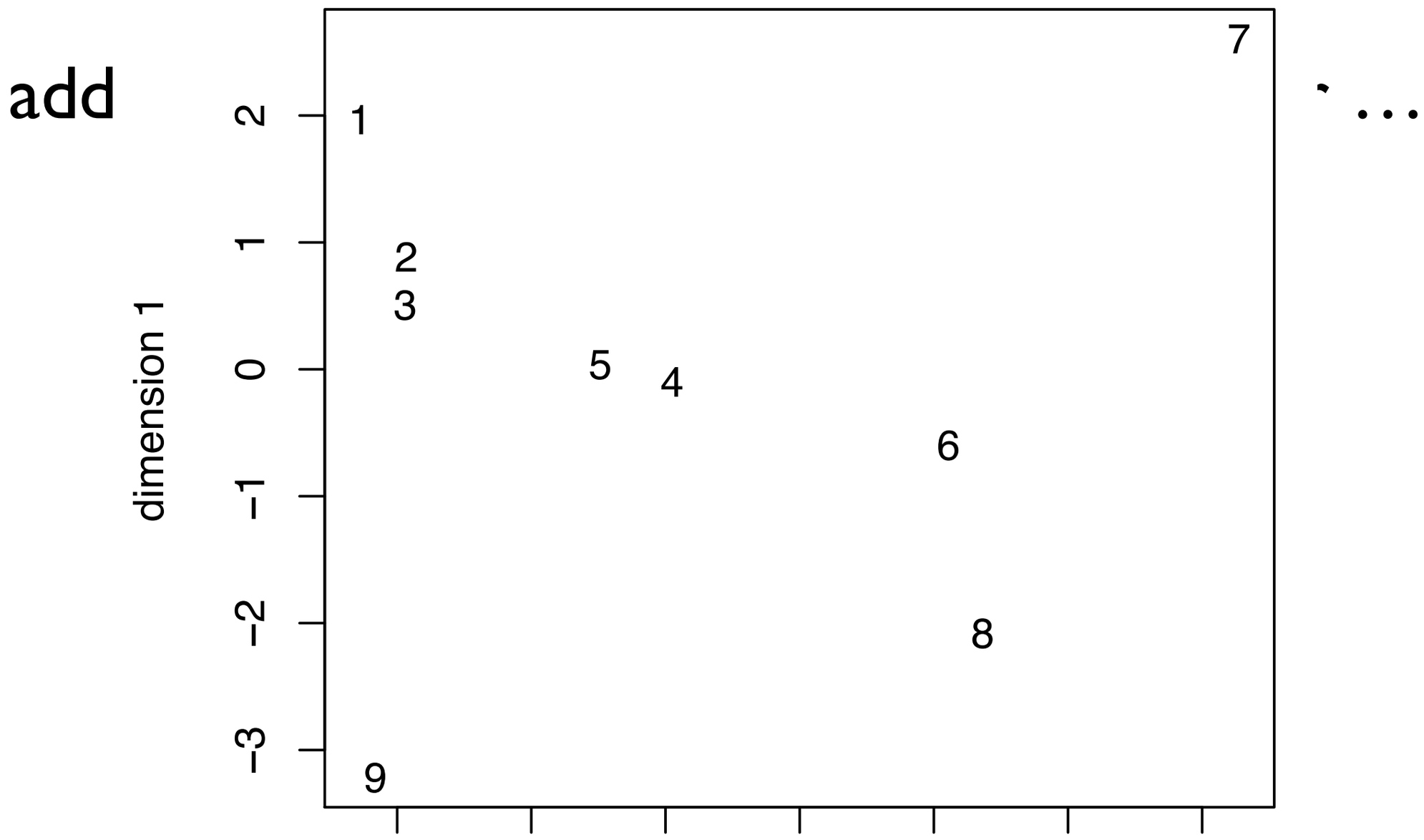
correlate language-specific perspectives with each other …

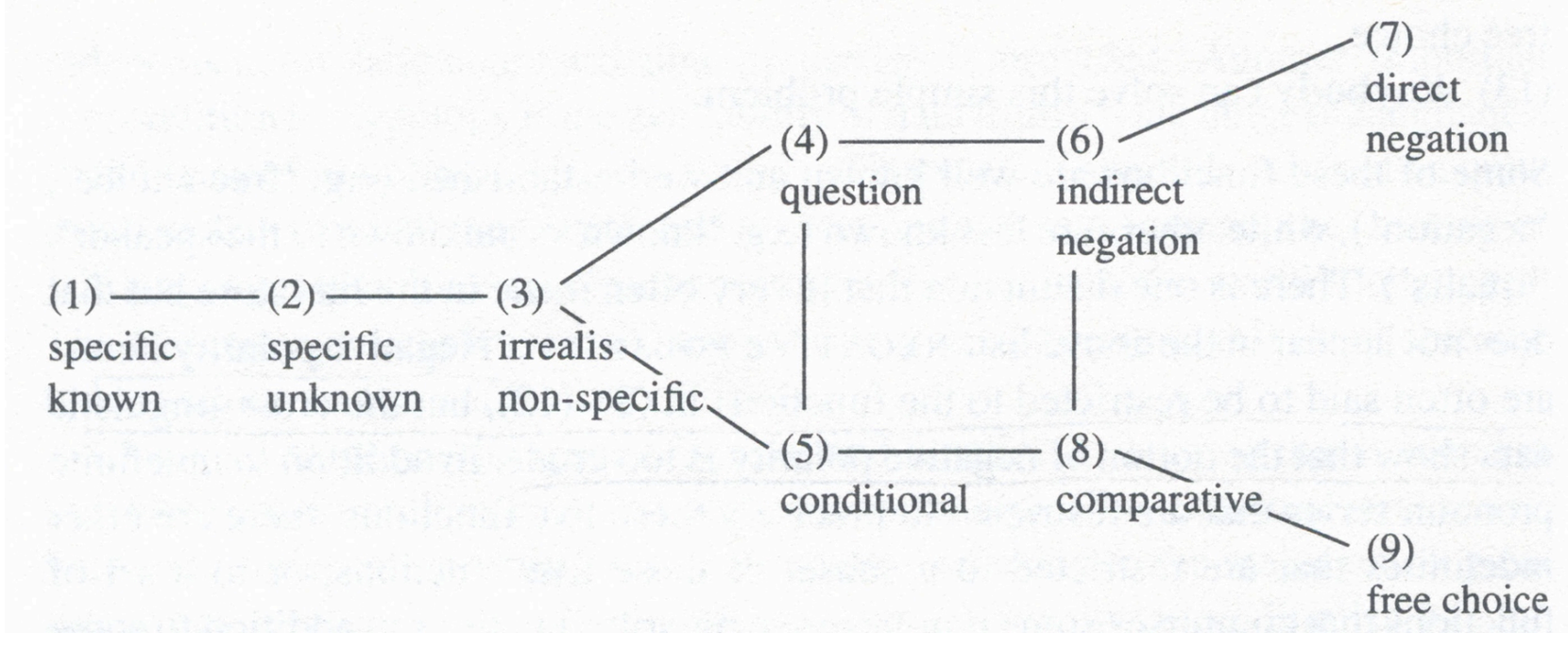
Dutch	0	0.53	0.40	0.67	0.71	0.36
English	0.53	0	0.44	0.70	0.82	0.50
German	0.40	0.44	0	0.86	0.57	0.47
Icelandic	0.67	0.70	0.86	0	0.55	0.63
Latin	0.71	0.82	0.57	0.55	0	0.80
Swedish	0.36	0.50	0.47	0.63	0.80	0

add language-specific perspectives together …

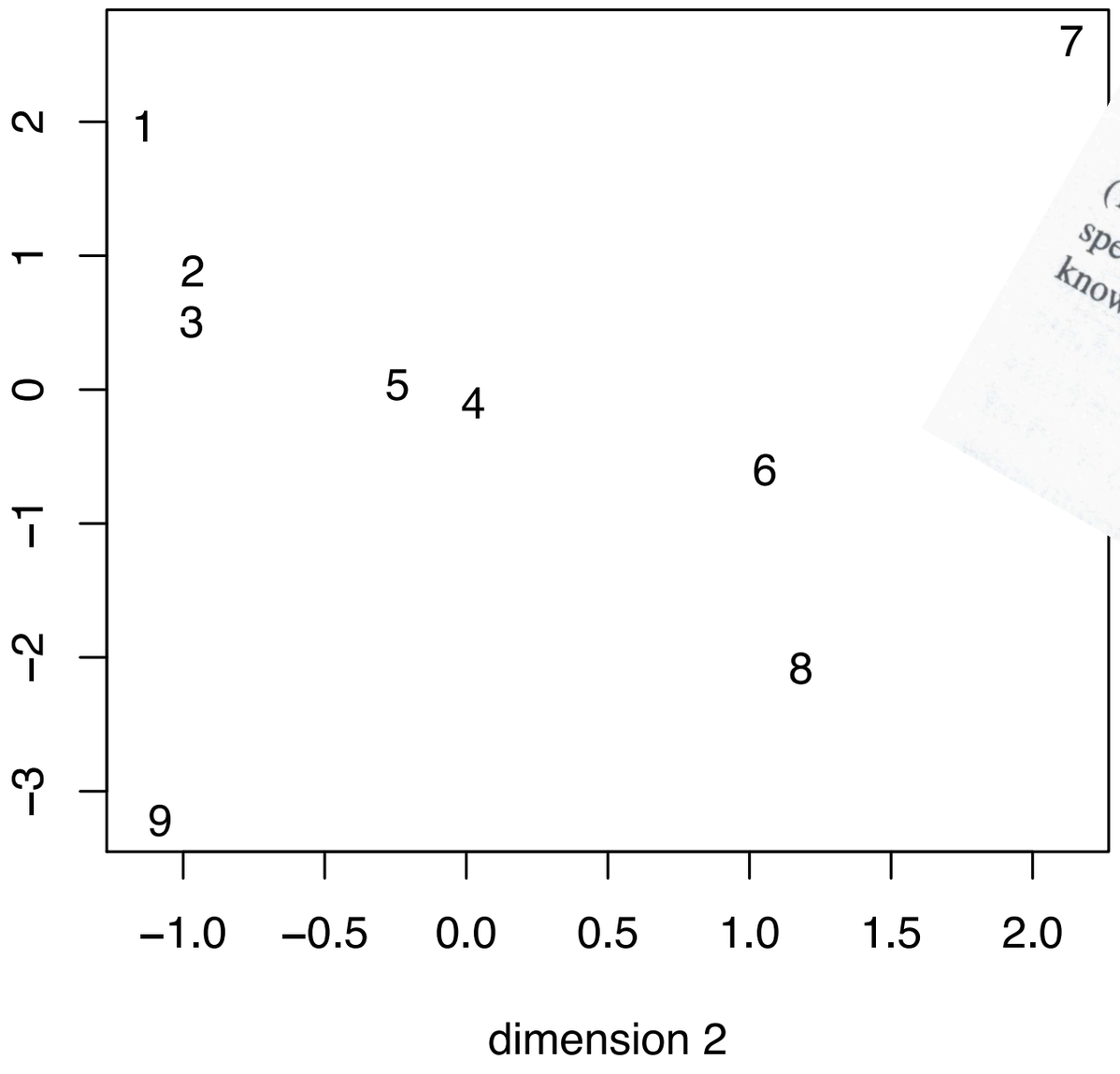
add language-specific perspectives together …

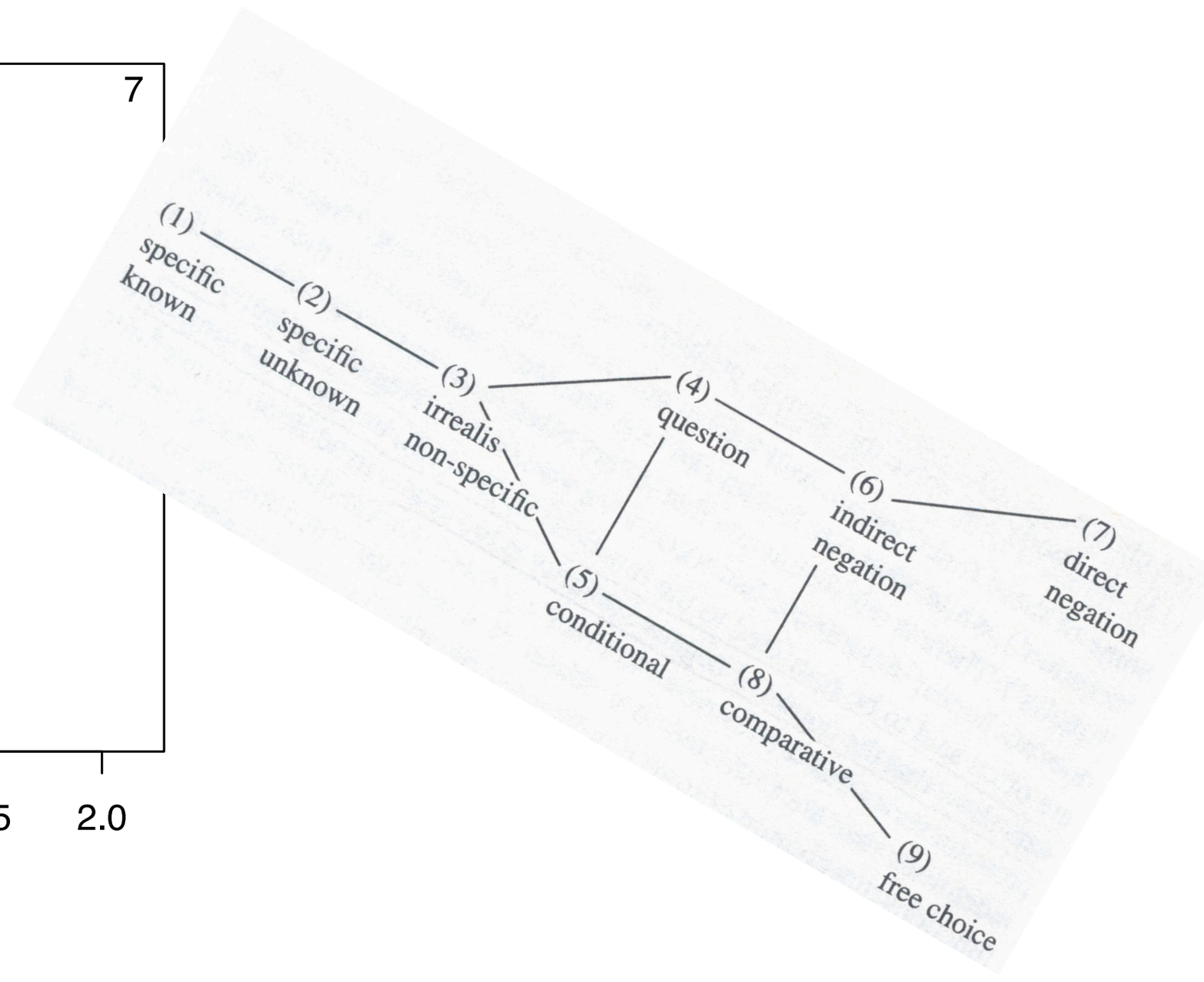



add
dimension 1
2
1
0
−1
−2
−3
1
2
3
5
4
6
8
7
9
...


(1)
specific
known
(2)
specific
unknown
(3)
irrealis
non-specific
(4)
question
(5)
conditional
(6)
indirect
negation
(7)
direct
negation
(8)
comparative
(9)
free choice

add language-specific perspectives together …





Interim Summary

Interim Summary

- Start with language-specific analysis:

Interim Summary

- Start with language-specific analysis:
  - Establish language-specific similarity between expressions

Interim Summary

- Start with language-specific analysis:
  - Establish language-specific similarity between expressions
  - Establish mapping of expressions to sample of functions

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

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

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

Inchoative - causative verb pairs

Inchoative - causative verb pairs

- Inchoative

  “The stick broke.”

Inchoative - causative verb pairs

- Inchoative
  "The stick broke."
- Causative
  "The girl broke the stick."

Inchoative - causative verb pairs

- Inchoative
  "The stick broke."

- Causative
  "The girl broke the stick."

Based on data from: Haspelmath, Martin. 1993. "More on the typology of inchoative/causative verb alternations." In: Comrie, B. & Polinsky, M. (eds.) *Causatives and transitivity*. Amsterdam: Benjamins, 87-120.

Some non-identical verb pairs in English

Some non-identical verb pairs in English

- *die - kill*
- *learn - teach*
- *rise - raise*
- *get lost - lose*
- *go out - put out*

Verb pairs investigated

Verb pairs investigated

begin	dry	melt
boil	fill	open
break	finish	rise/raise
burn	freeze	rock
change	gather	roll
close	get lost/	sink
connect	lose	split
destroy	go out/put	spread
develop	out	stop
die/kill	improve	turn
dissolve	learn/teach	wake up

Strategies for encoding inchoative-causative relation

(Terminology of: Haspelmath 1993)

Strategies for encoding inchoative-causative relation

- Different roots:
  Suppletive strategy (e.g. *die* - *kill*)

(Terminology of: Haspelmath 1993)

Strategies for encoding inchoative-causative relation

- Different roots:
  Suppletive strategy (e.g. *die* - *kill*)
- No differentiation:
  Labile strategy (e.g. *break*)

(Terminology of: Haspelmath 1993)

Strategies for encoding inchoative-causative relation

- Different roots:
  Suppletive strategy (e.g. *die* - *kill*)
- No differentiation:
  Labile strategy (e.g. *break*)
- Causative derived from Inchoative:
  Causative strategy (e.g. German *enden* - *beenden)*

(Terminology of: Haspelmath 1993)

Strategies for encoding inchoative-causative relation

- Different roots:
  Suppletive strategy (e.g. *die - kill*)
- No differentiation:
  Labile strategy (e.g. *break*)
- Causative derived from Inchoative:
  Causative strategy (e.g. German *enden - beenden)*
- Inchoative derived from Causative:
  Anticausative strategy (e.g. *be destroyed - destroy*)

(Terminology of: Haspelmath 1993)

Strategies for encoding inchoative-causative relation

- Different roots:
  Suppletive strategy (e.g. *die - kill*)
- No differentiation:
  Labile strategy (e.g. *break*)
- Causative derived from Inchoative:
  Causative strategy (e.g. German *enden - beenden)*
- Inchoative derived from Causative:
  Anticausative strategy (e.g. *be destroyed - destroy*)
- No (synchronic) primacy for one or the other:
  Equipollent strategy (e.g. German *versinken - versenken*)

(Terminology of: Haspelmath 1993)

Cross-linguistic proportion of causative strategies

Cross-linguistic proportion of causative strategies

split	0.04	get lost/lose	0.28	turn	0.48
close	0.06	develop	0.33	stop	0.62
break	0.07	roll	0.35	melt	0.68
open	0.10	spread	0.35	learn/teach	0.68
gather	0.12	begin	0.38	sink	0.70
change	0.12	finish	0.38	go out/put out	0.71
connect	0.14	fill	0.38	wake up	0.75
rock	0.25	destroy	0.39	dry	0.77
improve	0.26	burn	0.42	freeze	0.86
rise/raise	0.27	dissolve	0.42	boil	0.96

Cross-linguistic proportion of causative strategies

split	0.04
close	0.06
break	0.07
open	0.10
gather	0.12
change	0.12
connect	0.14
rock	0.25
improve	0.26
rise/raise	0.27

Scale of likelihood of spontaneous occurence

turn	0.48
stop	0.62
melt	0.68
learn/teach	0.68
sink	0.70
go out/put out	0.71
wake up	0.75
dry	0.77
freeze	0.86
boil	0.96

Arabic

Class A: C / CC
1. saʻaa / saʻʻaa
8. darasa / darrasa
14. damara / dammara
31. waqafa / waqqafa

Class B: in / ø
2. inkasara / kasara
5. infataʻa / fataʻa
6. inqafala / qafala
13. inṣahara / ṣahara
30. inšaqqa / šaqqa

Class C: t / ʔ
3. iʻtaraqa / ʔaʻraqa
22. intahaa / ʔanhaa

Class D: t / ø
9. iltamma / lamma
10. intašara / našara
17. irtabaṭa / rabaṭa
21. irtafaʕa / rafaʕa
27. imtalaʔa / malaʔa

Class E: ø / ʔ
11. ġariqa / ʔaġraqa
18. ġalaa / ʔaġlaa
23. daara / ʔadaara
26. ðaaba / ʔaðaaba

Class F: ta / ø
12. tabaddala / baddala
16. taṭawwara / ṭawwara
19. taʔarjaʻa / ʔarjaʻa
24. tadaʻraja / daʻraja
25. tajammada / jammada
28. taʻassana / ʻassana

Singular cases:
4. maata / qatala
7. badaʔa
15. daaʕa / xasira
20. inṭafaʔa / ʔaṭfaʔa
29. jaffa / jaffafa

English

Class A: Identical
1. wake up
2. break
3. burn
5. open
6. close
7. begin
9. gather
10. spread
11. sink
12. change
13. melt
16. develop
17. connect
18. boil
19. rock
22. finish
23. turn
24. roll
25. freeze
26. dissolve
27. fill
28. improve
29. dry
30. split
31. stop

Singular cases:
4. die / kill
8. learn / teach
14. be destroyed / destroy
15. get lost / lose
20. go out / put out
21. rise / raise

Finnish

Class A: ø / ttA
1. herätä / herättää
3. palaa / polttaa
8. oppia / opettaa
10. levitä / levittää
13. sulaa / sulattaa
18. kiehua / kiehuttaa
19. kiikkua / kiikuttaa
20. sammua / sammuttaa
21. kohota / kohottaa
22. loppua / lopettaa
24. vieriä / vierittää
25. jäätyä / jäädyttää
26. liueta / liuottaa
31. pysähtyä / pysähdyttää

Class B: U / A
2. murtua / murtaa
12. muuttua / muuttaa
16. kehittyä / kehittää
23. vääntyä / vääntää
27. täyttyä / täyttää
28. parantua / parantaa

Class C: UtU / ø
5. avautua / avata
6. sulkeutua / sulkea
14. tuhoutua / tuhota

Singular cases:
4. kuolla / tappaa
7. alkaa / aloitaa
9. kokoontua / koota
11. laskea
15. hukkaantua / hukata
17. yhtyä / yhdistää
29. kuivaa / kuivata
30. haljeta / halkaista

French

Class A: se / ø
1. se réveiller / réveiller
2. se briser / briser
5. s’ouvrir / ouvrir
6. se fermer / fermer
9. s’assembler / assembler
10. s’étendre / étendre
11. s’enfoncer / enfoncer
15. se perdre / perdre
16. se développer / développer
17. se lier / lier
19. se balancer / balancer
20. s’éteindre / éteindre
21. se lever / lever
23. se tourner / tourner
26. se dissoudre / dissoudre
27. se remplir / remplir
28. s’améliorer / améliorer
30. se fendre / fendre
31. s’arrêter / arrêter

Class B: Identical
3. brûler
7. commencer
8. apprendre
12. changer
22. finir
24. rouler
25. geler
29. sécher

Class C: ø / faire
13. fondre / faire fondre
18. bouillir / faire bouillir

Singular cases:
4. mourir / tuer
14. être détruit / détruir

Arabic

Class A: C / CC
1. saʿaa / saʿʿaa
8. darasa / darrasa
14. damara / dammara
31. waqafa / waqqafa

Class B: in / ø
2. inkasara / kasara
5. infataʿa / fataʿa
6. inqafala / qafala
13. inṣahara / ṣahara
30. inšaqqa / šaqqa

Class C: t / ʔ
3. iʿtaraqa / ʔaʿraqa
22. intahaa / ʔanhaa

Class D: t / ø
9. iltamma / lamma
10. intašara / našara
17. irtabaṭa / rabaṭa
21. irtafaʕa / rafaʕa
27. imtalaʔa / malaʔa

Class E: ø / ʔ
11. ġariqa / ʔaġraqa
18. ġalaa / ʔaġlaa
23. daara / ʔadaara
26. ðaaba / ʔaðaaba

Class F: ta / ø
12. tabaddala / baddala
16. taṭawwara / ṭawwara
19. taʔarjaʿa / ʔarjaʿa
24. tadaʿraja / daʿraja
25. tajammada / jammada
28. taʿassana / ʿassana

Singular cases:
4. maata / qatala
7. badaʔa
15. daaʕa / xasira
20. inṭafaʔa / ʔaṭfaʔa
29. jaffa / jaffafa

English

Class A: Identical
1. wake up
2. break
3. burn
5. open
6. close
7. begin
9. gather
10. spread
11. sink
12. change
13. melt
16. develop
17. connect
18. boil
19. rock
22. finish
23. turn
24. roll
25. freeze
26. dissolve
27. fill
28. improve
29. dry
30. split
31. stop

Singular cases:
4. die / kill
8. learn / teach
14. be destroyed / destroy
15. get lost / lose
20. go out / put out
21. rise / raise

Finnish

Class A: ø / ttA
1. herätä / herättää
3. palaa / polttaa
8. oppia / opettaa
10. levitä / levittää
13. sulaa / sulattaa
18. kiehua / kiehuttaa
19. kiikkua / kiikuttaa
20. sammua / sammuttaa
21. kohota / kohottaa
22. loppua / lopettaa
24. vieriä / vierittää
25. jäätyä / jäädyttää
26. liueta / liuottaa
31. pysähtyä / pysähdyttää

Class B: U / A
2. murtua / murtaa
12. muuttua / muuttaa
16. kehittyä / kehittää
23. vääntyä / vääntää
27. täyttyä / täyttää
28. parantua / parantaa

Class C: UtU / ø
5. avautua / avata
6. sulkeutua / sulkea
14. tuhoutua / tuhota

Singular cases:
4. kuolla / tappaa
7. alkaa / aloitaa
9. kokoontua / koota
11. laskea
15. hukkaantua / hukata
17. yhtyä / yhdistää
29. kuivaa / kuivata
30. haljeta / halkaista

French

Class A: se / ø
1. se réveiller / réveiller
2. se briser / briser
5. s’ouvrir / ouvrir
6. se fermer / fermer
9. s’assembler / assembler
10. s’étendre / étendre
11. s’enfoncer / enfoncer
15. se perdre / perdre
16. se développer / développer
17. se lier / lier
19. se balancer / balancer
20. s’éteindre / éteindre
21. se lever / lever
23. se tourner / tourner
26. se dissoudre / dissoudre
27. se remplir / remplir
28. s’améliorer / améliorer
30. se fendre / fendre
31. s’arrêter / arrêter

Class B: Identical
3. brûler
7. commencer
8. apprendre
12. changer
22. finir
24. rouler
25. geler
29. sécher

Class C: ø / faire
13. fondre / faire fondre
18. bouillir / faire bouillir

Singular cases:
4. mourir / tuer
14. être détruit / détruir

Arabic

Class A: C / CC
1. saʿaa / saʿʿaa
8. darasa / darrasa
14. damara / dammara
31. waqafa / waqqafa

Class B: in / ø
2. inkasara / kasara
5. infataʿa / fataʿa
6. inqafala / qafala
13. inṣahara / ṣahara
30. inšaqqa / šaqqa

Class C: t / ʔ
3. iʿtaraqa / ʔaʿraqa
22. intahaa / ʔanhaa

Class D: t / ø
9. iltamma / lamma
10. intašara / našara
17. irtabaṭa / rabaṭa
21. irtafaʕa / rafaʕa
27. imtalaʔa / malaʔa

Class E: ø / ʔ
11. ġariqa / ʔaġraqa
18. ġalaa / ʔaġlaa
23. daara / ʔadaara
26. ðaaba / ʔaðaaba

Class F: ta / ø
12. tabaddala / baddala
16. taṭawwara / ṭawwara
19. taʔarjaʿa / ʔarjaʿa
24. tadaʿraja / daʿraja
25. tajammada / jammada
28. taʿassana / ʿassana

Singular cases:
4. maata / qatala
7. badaʔa
15. daaʕa / xasira
20. inṭafaʔa / ʔaṭfaʔa
29. jaffa / jaffafa

English

Class A: Identical
1. wake up
2. break
3. burn
5. open
6. close
7. begin
9. gather
10. spread
11. sink
12. change
13. melt
16. develop
17. connect
18. boil
19. rock
22. finish
23. turn
24. roll
25. freeze
26. dissolve
27. fill
28. improve
29. dry
30. split
31. stop

Singular cases:
4. die / kill
8. learn / teach
14. be destroyed / destroy
15. get lost / lose
20. go out / put out
21. rise / raise

Finnish

Class A: ø / ttA
1. herätä / herättää
3. palaa / polttaa
8. oppia / opettaa
10. levitä / levittää
13. sulaa / sulattaa
18. kiehua / kiehuttaa
19. kiikkua / kiikuttaa
20. sammua / sammuttaa
21. kohota / kohottaa
22. loppua / lopettaa
24. vieriä / vierittää
25. jäätyä / jäädyttää
26. liueta / liuottaa
31. pysähtyä / pysähdyttää

Class B: U / A
2. murtua / murtaa
12. muuttua / muuttaa
16. kehittyä / kehittää
23. vääntyä / vääntää
27. täyttyä / täyttää
28. parantua / parantaa

Class C: UtU / ø
5. avautua / avata
6. sulkeutua / sulkea
14. tuhoutua / tuhota

Singular cases:
4. kuolla / tappaa
7. alkaa / aloitaa
9. kokoontua / koota
11. laskea
15. hukkaantua / hukata
17. yhtyä / yhdistää
29. kuivaa / kuivata
30. haljeta / halkaista

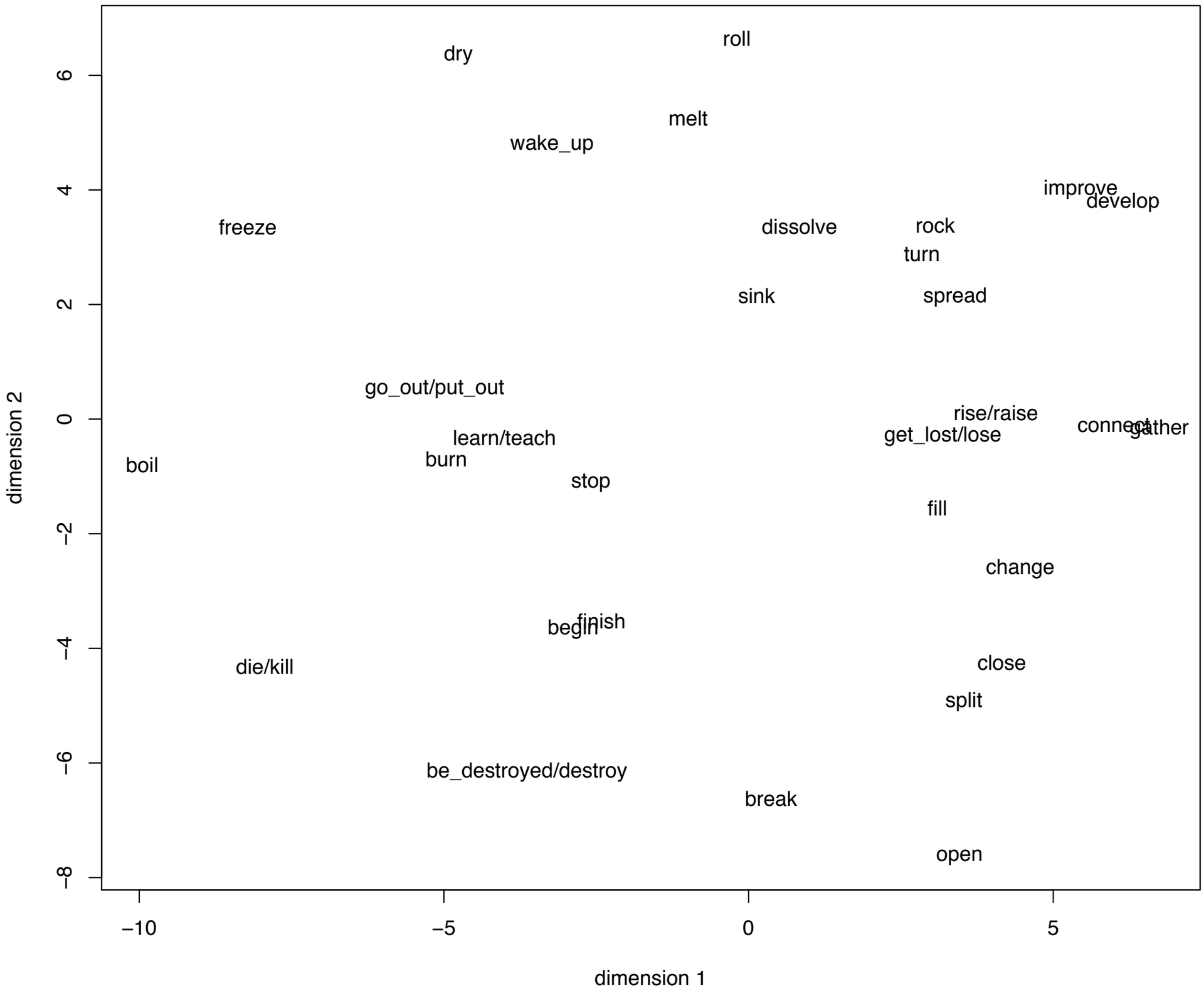
French

Class A: se / ø
1. se réveiller / réveiller
2. se briser / briser
5. s’ouvrir / ouvrir
6. se fermer / fermer
9. s’assembler / assembler
10. s’étendre / étendre
11. s’enfoncer / enfoncer
15. se perdre / perdre
16. se développer / développer
17. se lier / lier
19. se balancer / balancer
20. s’éteindre / éteindre
21. se lever / lever
23. se tourner / tourner
26. se dissoudre / dissoudre
27. se remplir / remplir
28. s’améliorer / améliorer
30. se fendre / fendre
31. s’arrêter / arrêter

Class B: Identical
3. brûler
7. commencer
8. apprendre
12. changer
22. finir
24. rouler
25. geler
29. sécher

Class C: ø / faire
13. fondre / faire fondre
18. bouillir / faire bouillir

Singular cases:
4. mourir / tuer
14. être détruit / détruir


roll
dry
melt
wake_up
improve
develop
freeze
dissolve
rock
turn
sink
spread
go_out/put_out
rise/raise
connect
gather
learn/teach
get_lost/lose
boil
burn
stop
fill
change
finish
begin
die/kill
close
split
be_destroyed/destroy
break
open
6
4
2
0
−2
−4
−6
−8
−10
−5
0
5
dimension 2
dimension 1

Cross-linguistic judgments

Cross-linguistic judgments

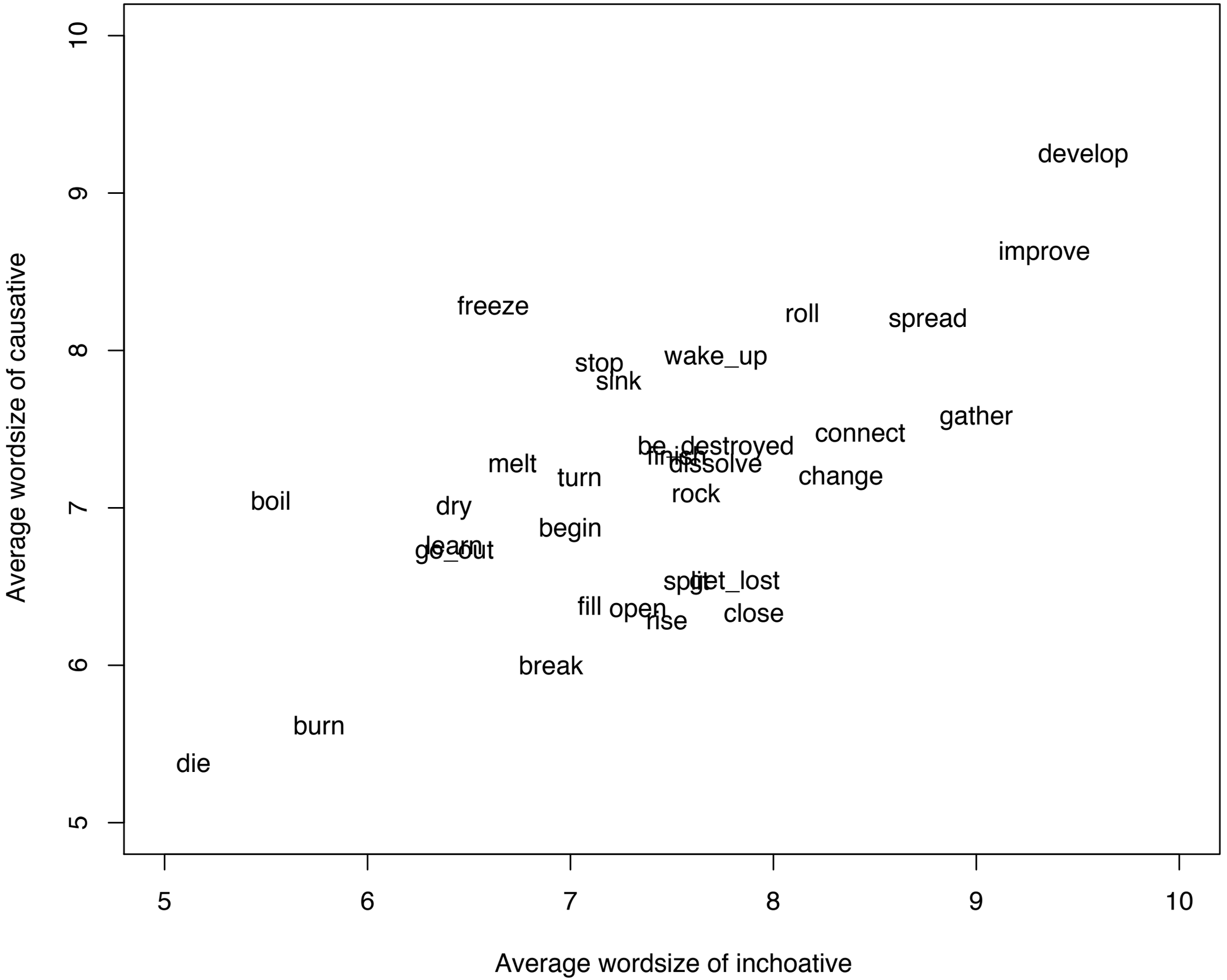
- Cross-linguistic analysis of expressions

Cross-linguistic judgments

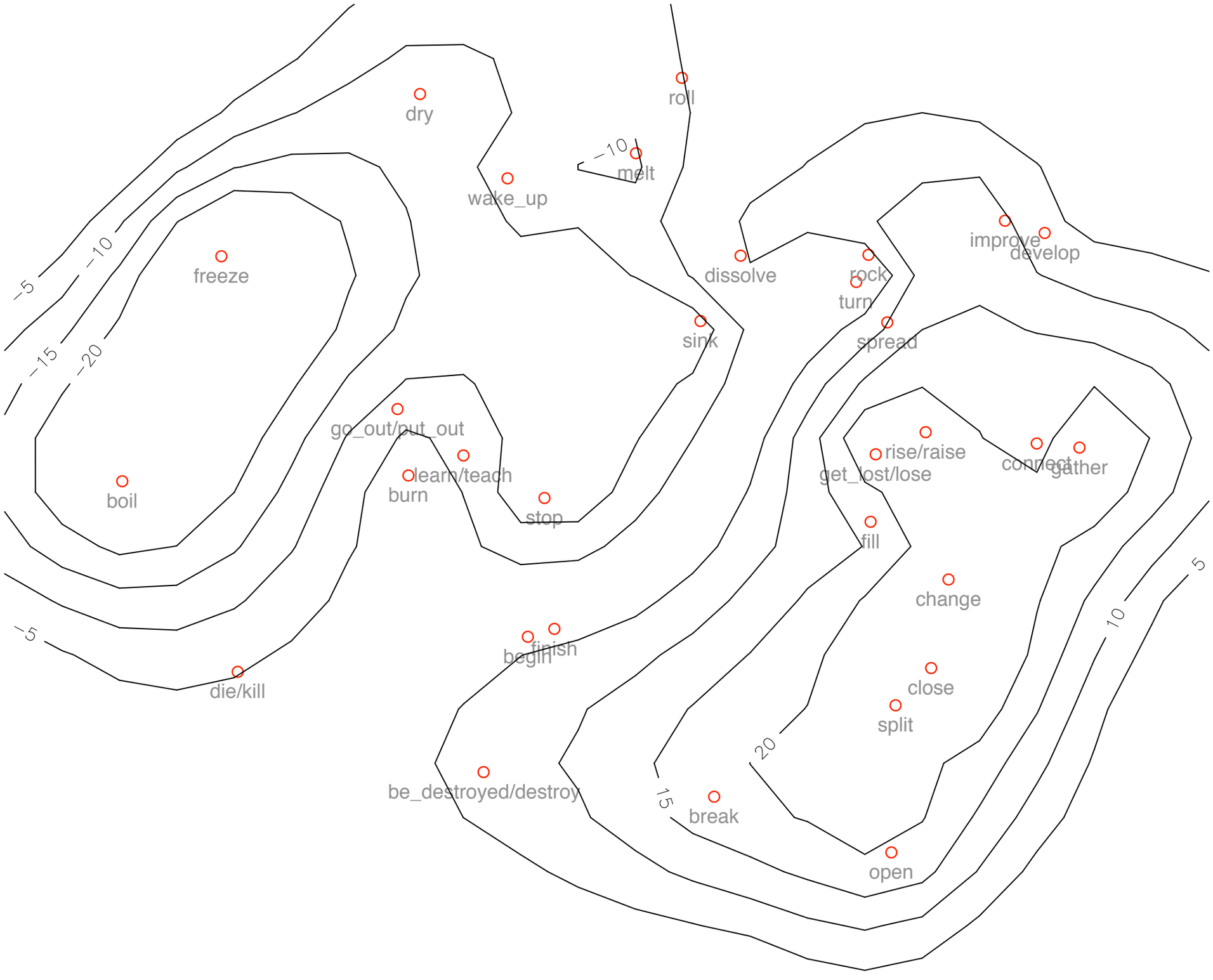
- Cross-linguistic analysis of expressions
- Keep it simple

Cross-linguistic judgments

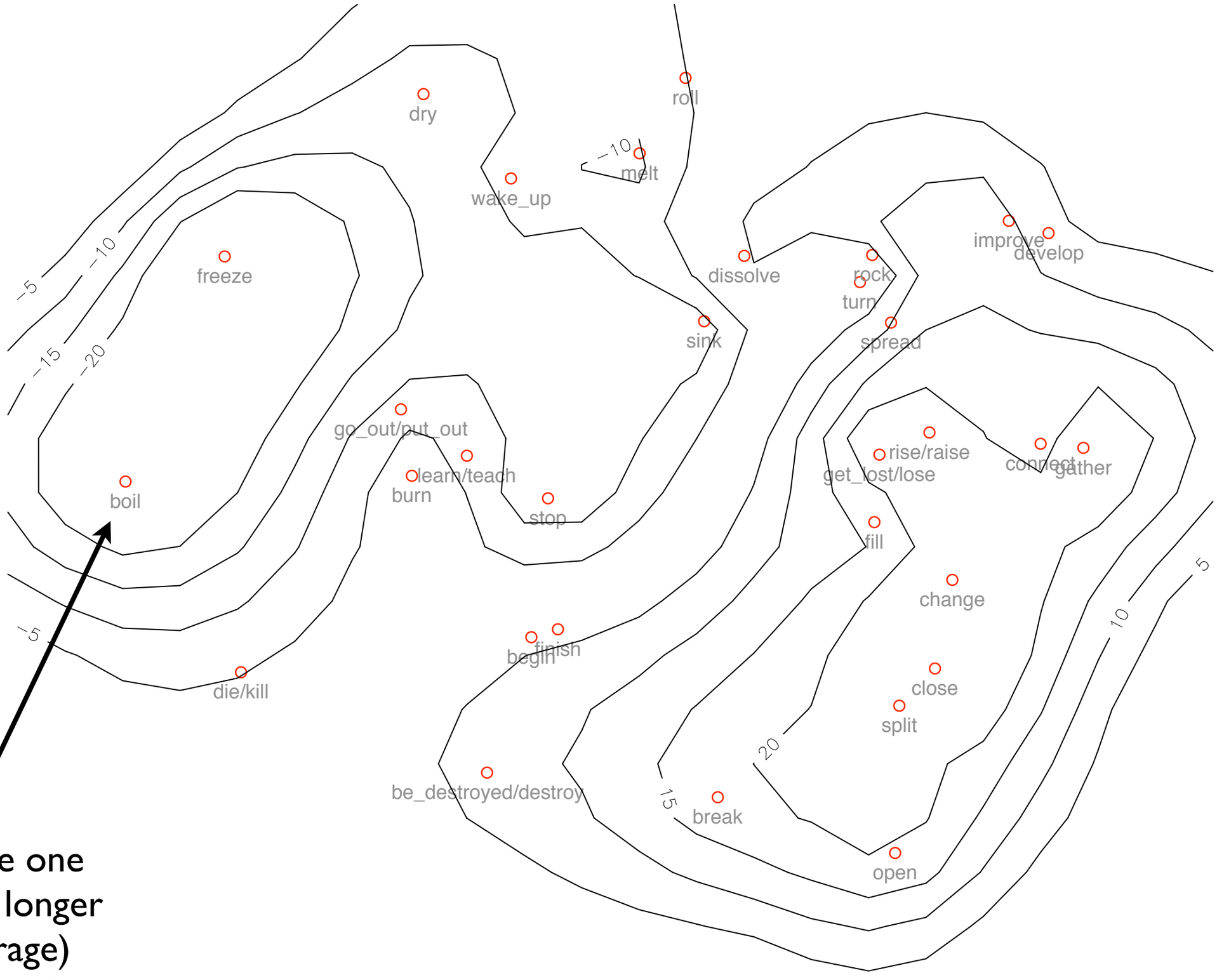
- Cross-linguistic analysis of expressions
- Keep it simple
- e.g. count length of strings


develop
improve
freeze
roll
spread
stop
wake_up
sink
gather
connect
be_destroyed
finish
dissolve
melt
turn
change
rock
boil
dry
begin
learn
go_out
get_lost
fill
open
close
rise
break
burn
die
Average wordsize of causative
Average wordsize of inchoative
5
6
7
8
9
10

Inchoative–Causative Character Count

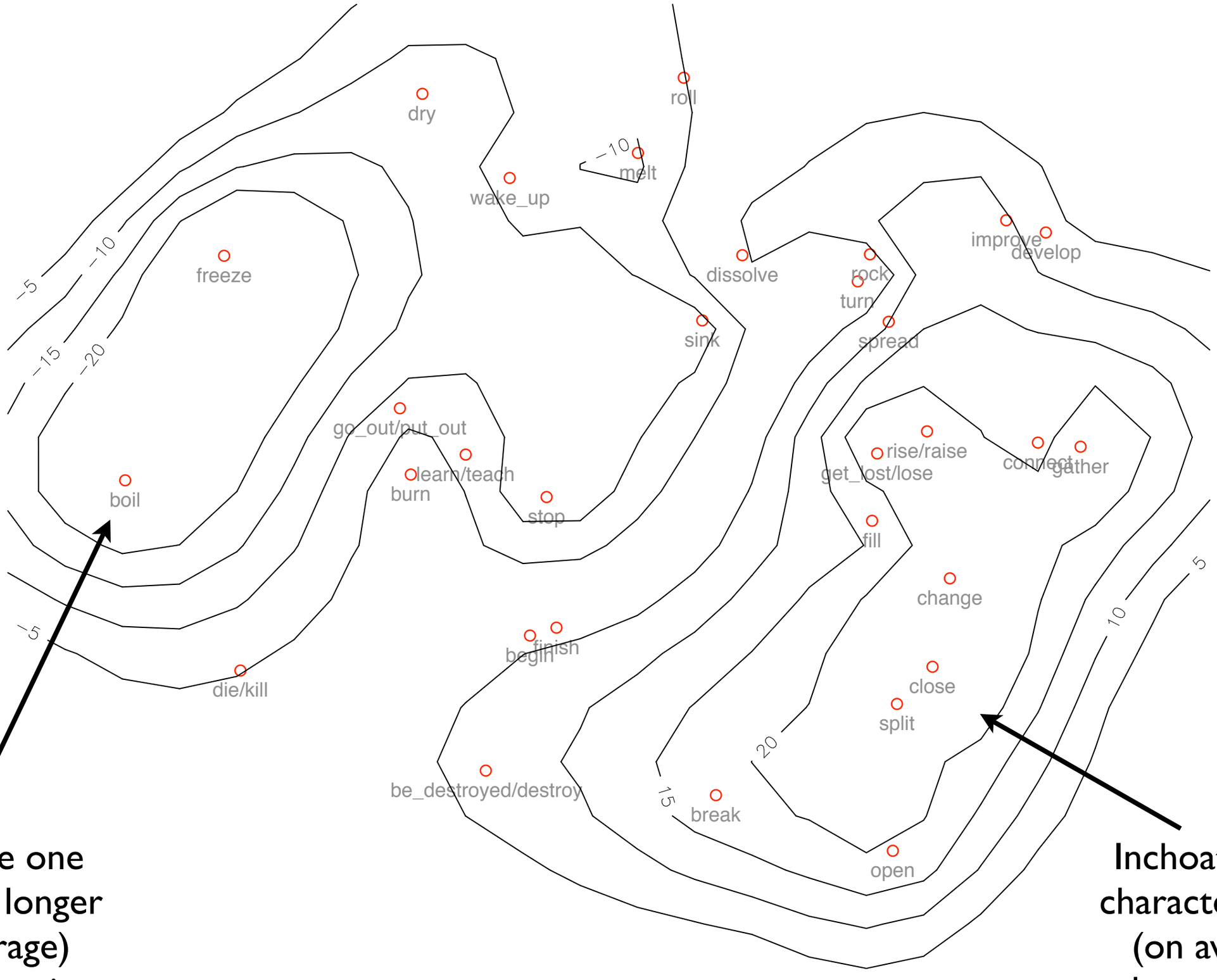


Inchoative–Causative Character Count



Causative one character longer (on average) than inchoative

Inchoative–Causative Character Count



Causative one character longer (on average) than inchoative

Inchoative one character longer (on average) than causative

Multivariate Matrix Regression

Zapala, M.A. and J. Schork (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. PNAS 103(51): 19430–19435

Multivariate Matrix Regression

	Sums of Sqs	Mean Sqs	F Model	R^2
difference inchoative-causative	0.041	0.041	9.594	0.211 ***
average length of inchoative+causative	0.030	0.030	6.922	0.152 ***
combined effect	0.008	0.008	1.758	0.039
Residuals	0.116	0.004		0.596

Zapala, M.A. and J. Schork (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. PNAS 103(51): 19430–19435

Multivariate Matrix Regression

	Sums of Sqs	Mean Sqs	F Model	R^2
length of causative	0.022	0.022	5.449	0.116 ***
length of inchoative	0.048	0.048	11.759	0.249 ***
combined effect	0.012	0.012	2.962	0.063
Residuals	0.111	0.004		0.572

Zapala, M.A. and J. Schork (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. PNAS 103(51): 19430–19435

Language comparison using minimal comparative judgments

Language comparison using minimal comparative judgments

- Select concrete expressions in context

Language comparison using minimal comparative judgments

- Select concrete expressions in context
- Annotate these expressions with

Language comparison using minimal comparative judgments

- Select concrete expressions in context
- Annotate these expressions with
  - language-specific characteristics (don’t worry !)

Language comparison using minimal comparative judgments

- Select concrete expressions in context
- Annotate these expressions with
  - language-specific characteristics (don’t worry !)
  - sample of function/meaning (tricky !)

Language comparison using minimal comparative judgments

- Select concrete expressions in context
- Annotate these expressions with
  - language-specific characteristics (don’t worry !)
  - sample of function/meaning (tricky !)
  - cross-linguistic characteristics (minimal !)

Language comparison using minimal comparative judgments

- Select concrete expressions in context
- Annotate these expressions with
  - language-specific characteristics (don’t worry !)
  - sample of function/meaning (tricky !)
  - cross-linguistic characteristics (minimal !)
- That seems to be sufficient for accurate large-scale language comparison