1 May, 2025.

A Grounded Typology of Word Classes

COLEMAN HALEY SHARON GOLDWATER EDOARDO PONTI

Typology:

The study of how languages *vary* from one another.

To talk about how languages *vary* from one another, we need something which is *the same*.

John
s

John loves
S V

John loves Paul
S V O

John loves Paul

S V O

ジョーンは

JOHN

S

John loves Paul

S V O

ジョーンは　ポールを

JOHN PAUL

S O

John loves Paul

S V O

ジョーンは　ポールを　愛している

JOHN PAUL LOVES

S O V

Comparative Concepts

(Haspelmath 2010)

John loves Paul

S V O

ジョーンは　ポールを　愛している

JOHN PAUL LOVES

S O V

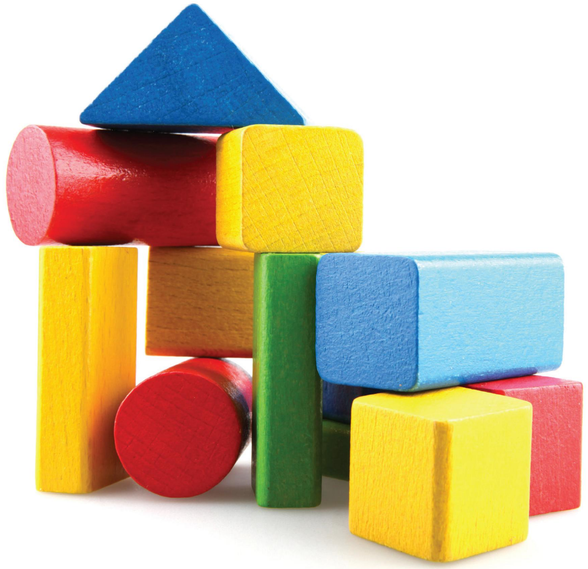
What makes a verb a verb?

What makes a verb a verb?

Semantics

What makes a verb a verb?

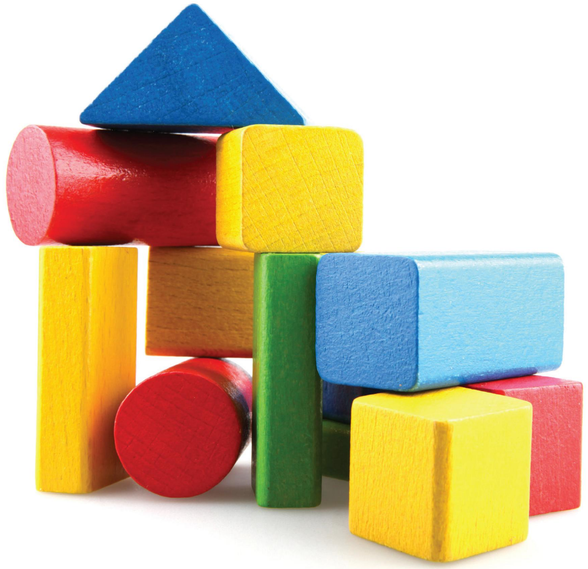
Semantics



Nouns: Objects

What makes a verb a verb?

Semantics



Nouns: Objects

Verbs: Events/actions

What makes a verb a verb?

Semantics

What makes a verb a verb?

What makes a verb a verb?

(Morpho)syntax

What makes a verb a verb?

(Morpho)syntax

"Giving gifts is my favourite thing to do."

What makes a verb a verb?

(Morpho)syntax

"Giving gifts is my favourite thing to do."

What about other languages?

Goal: *Empirically ground* the semantics of comparative concepts across languages.

Start simple: semantic *contentfulness*
(how much meaning a unit carries)

How can contentfulness be quantified?

How can contentfulness be quantified?

Prior work: ask people (Spreen & Schultz 1966; Connell & Lynott 2012; Paivio et al., 1968)

How can contentfulness be quantified?

Prior work: ask people (Spreen & Schultz 1966; Connell & Lynott 2012; Paivio et al., 1968)

- ✳ Difficult to scale, subjective

How can contentfulness be quantified?

Prior work: ask people (Spreen & Schultz 1966; Connell & Lynott 2012; Paivio et al., 1968)

- ✳ Difficult to scale, subjective
- ✳ What to ask? *Imageability, Concreteness, Strength of sensory experience...*

How can contentfulness be quantified?

Prior work: ask people (Spreen & Schultz 1966; Connell & Lynott 2012; Paivio et al., 1968)

- ✳ Difficult to scale, subjective
- ✳ What to ask? *Imageability, Concreteness, Strength of sensory experience...*
- ✳ Almost always type level

How can contentfulness be quantified?

Prior work: ask people (Spreen & Schultz 1966; Connell & Lynott 2012; Paivio et al., 1968)

- ✳ Difficult to scale, subjective
- ✳ What to ask? *Imageability, Concreteness, Strength of sensory experience...*
- ✳ Almost always type level
  - ✳ But variation isn't! *"An explosion at the factory"* vs. *"An explosion of ideas"*

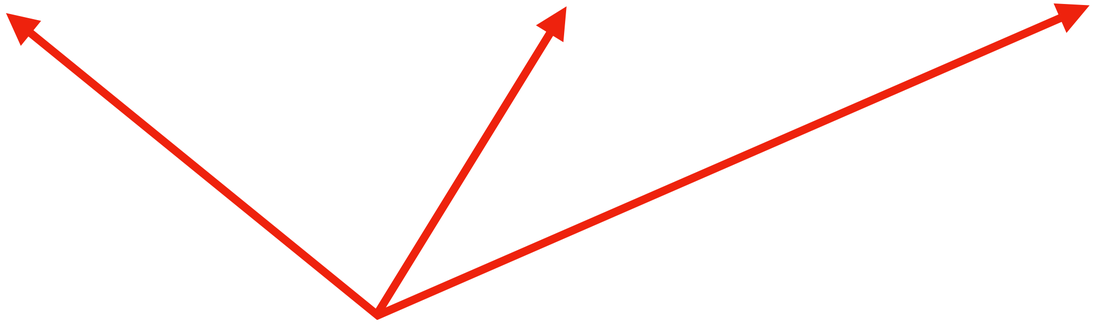
A new method: groundedness

A new method: groundedness

$$\mathrm{PMI}(\mathbf{w}_t, \mathbf{m} \mid w_{<t}) = \log p(\mathbf{w}_t \mid \mathbf{m}, \mathbf{w}_{<t}) - \log p(\mathbf{w}_t \mid \mathbf{w}_{<t})$$

A new method: groundedness

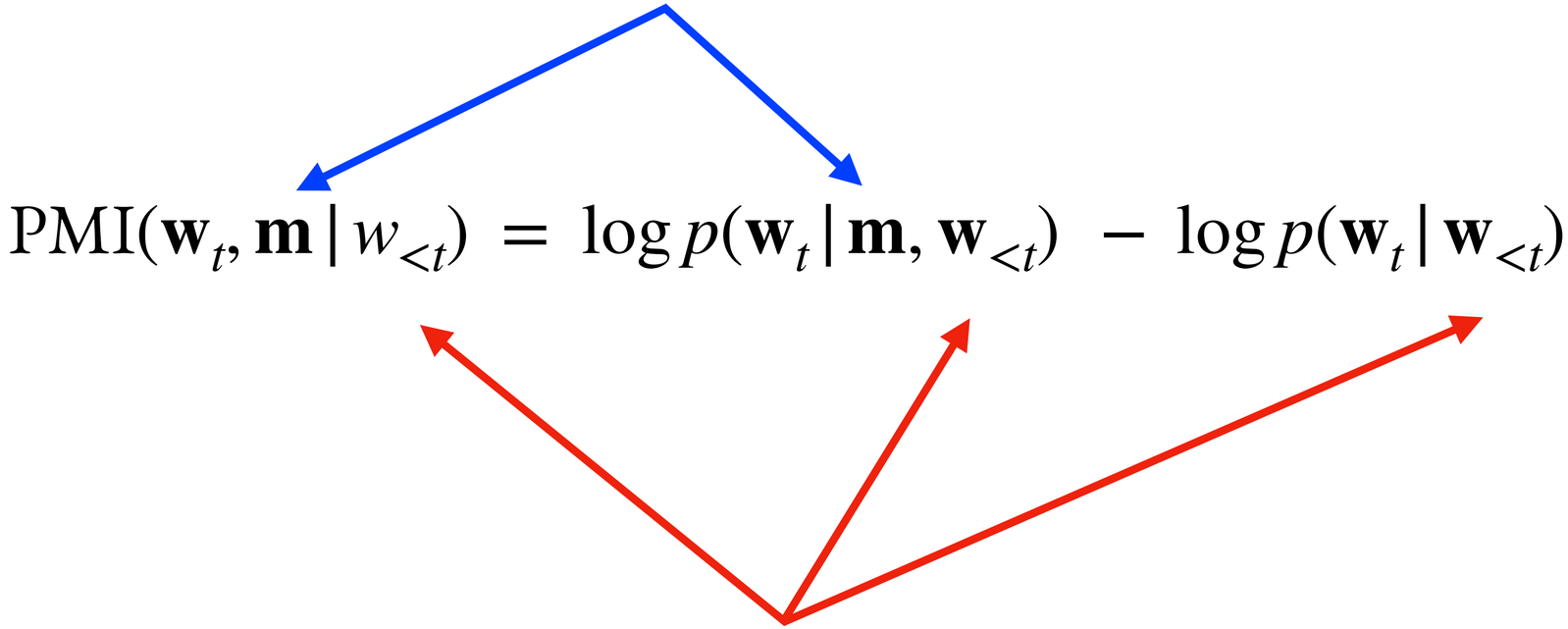
$$\mathrm{PMI}(\mathbf{w}_t, \mathbf{m} \mid w_{<t}) = \log p(\mathbf{w}_t \mid \mathbf{m}, \mathbf{w}_{<t}) - \log p(\mathbf{w}_t \mid \mathbf{w}_{<t})$$



(word) token-level!

A new method: groundedness

Where to get the meaning from?



(word) token-level!

Let's get meaning from (visual) *grounding!*

Let's get meaning from (visual) *grounding!*

Let's get meaning from (visual) *grounding!*

"A cat plays with a toy banana."

Let's get meaning from (visual) *grounding!*

"A cat plays with a toy banana."

$$\text{groundedness} = \text{PMI}(\mathbf{w}_t, \mathbf{m} \mid w_{<t}) = \log p(\mathbf{w}_t \mid \mathbf{m}, \mathbf{w}_{<t}) - \log p(\mathbf{w}_t \mid \mathbf{w}_{<t})$$

Let's get meaning from (visual) *grounding!*

"A cat plays with a toy banana."

$$\text{groundedness} = \text{PMI}(\mathbf{w}_t, \mathbf{m} \mid w_{<t}) = \log p(\mathbf{w}_t \mid \mathbf{m}, \mathbf{w}_{<t}) - \log p(\mathbf{w}_t \mid \mathbf{w}_{<t})$$

$p(\text{cat} \mid \text{A},$  $)$

Let's get meaning from (visual) *grounding!*

"A cat plays with a toy banana."

$$\text{groundedness} = \text{PMI}(\mathbf{w}_t, \mathbf{m} \mid w_{<t}) = \log p(\mathbf{w}_t \mid \mathbf{m}, \mathbf{w}_{<t}) - \log p(\mathbf{w}_t \mid \mathbf{w}_{<t})$$

$p(\text{cat} \mid \text{A},$  $)$

Image captioning model

Let's get meaning from (visual) *grounding!*

"A cat plays with a toy banana."

$$\text{groundedness} = \text{PMI}(\mathbf{w}_t, \mathbf{m} \mid w_{<t}) = \log p(\mathbf{w}_t \mid \mathbf{m}, \mathbf{w}_{<t}) - \log p(\mathbf{w}_t \mid \mathbf{w}_{<t})$$

$p(\text{cat} \mid \text{A},$  $)$ $\qquad$ $p(\text{cat} \mid \text{A})$

Image captioning model

Let's get meaning from (visual) *grounding!*

"A cat plays with a toy banana."

$$\text{groundedness} = \text{PMI}(\mathbf{w}_t, \mathbf{m} \mid w_{<t}) = \log p(\mathbf{w}_t \mid \mathbf{m}, \mathbf{w}_{<t}) - \log p(\mathbf{w}_t \mid \mathbf{w}_{<t})$$

$p(\text{cat} \mid \text{A},$  $)$

Image captioning model

$p(\text{cat} \mid \text{A})$

Language model

GROUNDEDNESS

4.89

LOG P(📄) − LOG P(📄|🖼)

7.96 − 3.07

4.89							
kanepede	doldurulmuş	bir	muzla	oynayan	bir	kedi
ON-THE-COUCH NOUN	STUFF+PAST+PART VERB	A DET	BANANA NOUN	PLAY+PART VERB	A DET	CAT NOUN

GROUNDEDNESS

3.59

LOG P(📄) – LOG P(📄|🖼)

10.03 – 6.44

	3.59						
kanepede	**doldurulmuş**	**bir**	**muzla**	**oynayan**	**bir**	**kedi**
ON-THE-COUCH	STUFF+PAST+PART	A	BANANA	PLAY+PART	A	CAT
NOUN	VERB	DET	NOUN	VERB	DET	NOUN

GROUNDEDNESS

0.25

LOG P(📄) − LOG P(📄|🖼)

0.38 − 0.13

kanepede	doldurulmuş	0.25 bir	muzla	oynayan	bir	kedi
ON-THE-COUCH NOUN	STUFF+PAST+PART VERB	A DET	BANANA NOUN	PLAY+PART VERB	A DET	CAT NOUN

GROUNDEDNESS

7.83

LOG P(📄) − LOG P(📄|🖼)

8.89 − 1.06

kanepede	doldurulmuş	bir	**7.83** **muzla**	oynayan	bir	kedi
ON-THE-COUCH	STUFF+PAST+PART	A	BANANA	PLAY+PART	A	CAT
NOUN	VERB	DET	NOUN	VERB	DET	NOUN

GROUNDEDNESS

0.84

LOG P(📄) – LOG P(📄|🖼)

0.95 – 0.11

kanepede	doldurulmuş	bir	muzla	0.84 oynayan	bir	kedi
ON-THE-COUCH	STUFF+PAST+PART	A	BANANA	PLAY+PART	A	CAT
NOUN	VERB	DET	NOUN	VERB	DET	NOUN

GROUNDEDNESS

0.21

LOG P(📄) − LOG P(📄|🖼)

0.56 − 0.35

kanepede	doldurulmuş	bir	muzla	oynayan	0.21 bir	kedi
ON-THE-COUCH	STUFF+PAST+PART	A	BANANA	PLAY+PART	A	CAT
NOUN	VERB	DET	NOUN	VERB	DET	NOUN

GROUNDEDNESS

5.17

LOG P(📄) − LOG P(📄|🖼)

5.20 − 0.03

kanepede	doldurulmuş	bir	muzla	oynayan	bir	5.17 kedi
ON-THE-COUCH	STUFF+PAST+PART	A	BANANA	PLAY+PART	A	CAT
NOUN	VERB	DET	NOUN	VERB	DET	NOUN

4.89	3.59	0.25	7.83	0.84	0.21	5.17
kanepede	doldurulmuş	bir	muzla	oynayan	bir	kedi
ON-THE-COUCH	STUFF+PAST+PART	A	BANANA	PLAY+PART	A	CAT
NOUN	VERB	DET	NOUN	VERB	DET	NOUN

Lexical-functional distinction:

Lexical-functional distinction:

Lexical POS: **vs.** ***Functional POS:***

Lexical-functional distinction:

Lexical POS: **vs.** ***Functional POS:***

- *Example:* Nouns

Lexical-functional distinction:

Lexical POS:

- *Example:* Nouns

vs.

Functional POS:

- *Example:* Determiners (“the”, “a”)

Lexical-functional distinction:

Lexical POS:

- *Example:* Nouns
- Contentful

vs.

Functional POS:

- *Example:* Determiners (“the”, “a”)

Lexical-functional distinction:

Lexical POS:

- *Example:* Nouns
- Contentful

vs.

Functional POS:

- *Example:* Determiners (“the”, “a”)
- Grammatical

Lexical-functional distinction:

Lexical POS:

- *Example:* Nouns
- Contentful
- Admit new members

vs.

Functional POS:

- *Example:* Determiners (“the”, “a”)
- Grammatical

Lexical-functional distinction:

Lexical POS:

- *Example:* Nouns
- Contentful
- Admit new members

vs.

Functional POS:

- *Example:* Determiners (“the”, “a”)
- Grammatical
- Closed to new members

Lexical-functional distinction:

Lexical POS:

- *Example:* Nouns
- Contentful
- Admit new members
- Morphologically productive

vs.

Functional POS:

- *Example:* Determiners (“the”, “a”)
- Grammatical
- Closed to new members

Lexical-functional distinction:

Lexical POS:

- *Example:* Nouns
- Contentful
- Admit new members
- Morphologically productive

vs.

Functional POS:

- *Example:* Determiners (“the”, “a”)
- Grammatical
- Closed to new members
- Morphologically inert

Lexical-functional distinction:

Lexical POS:

- *Example:* Nouns
- Contentful
- Admit new members
- Morphologically productive

vs.

Functional POS:

- *Example:* Determiners (“the”, “a”)
- Grammatical
- Closed to new members
- Morphologically inert

Tools

Tools

Captioning model: PaliGemma

Tools

Captioning model: PaliGemma

Language model: matched to be trained on the same data as PaliGemma*

Tools

Captioning model: PaliGemma

Language model: matched to be trained on the same data as PaliGemma*

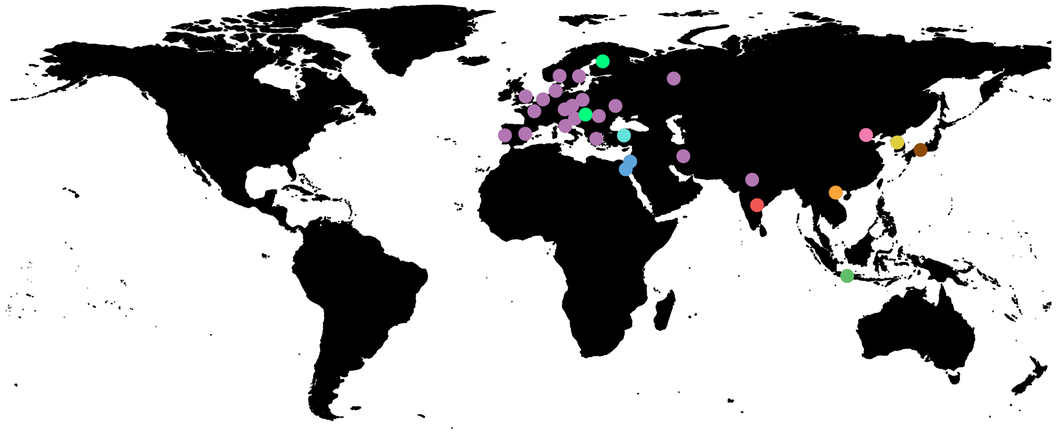
Datasets: COCO-35L; XM3600; Multi30K

Tools

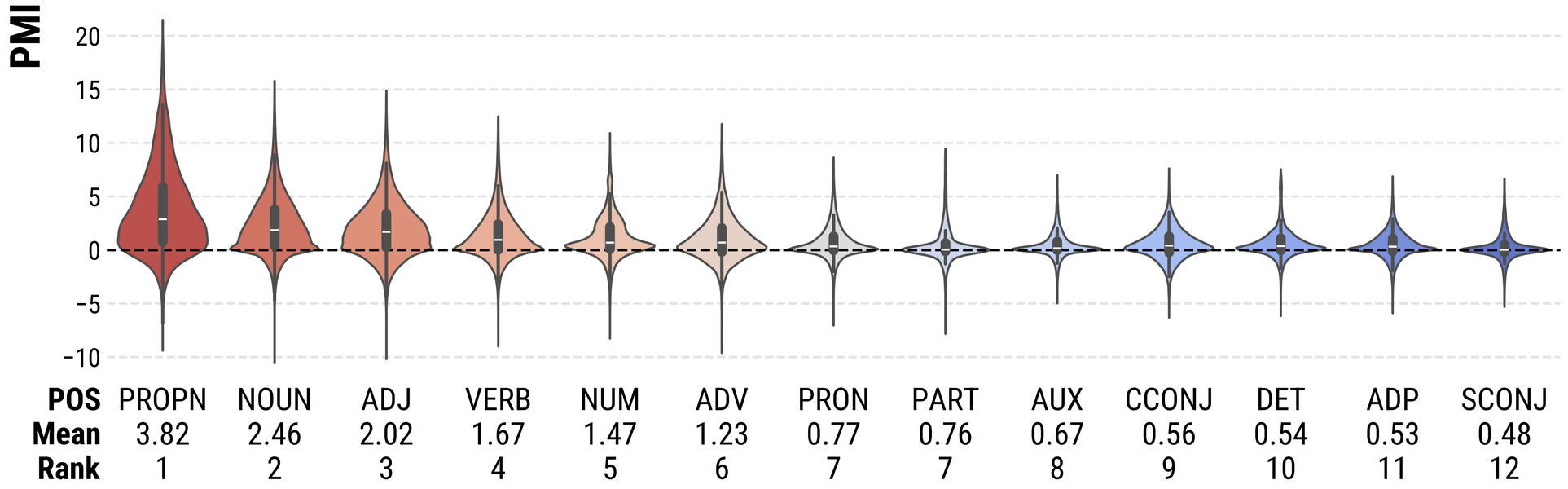
Captioning model: PaliGemma

Language model: matched to be trained on the same data as PaliGemma*

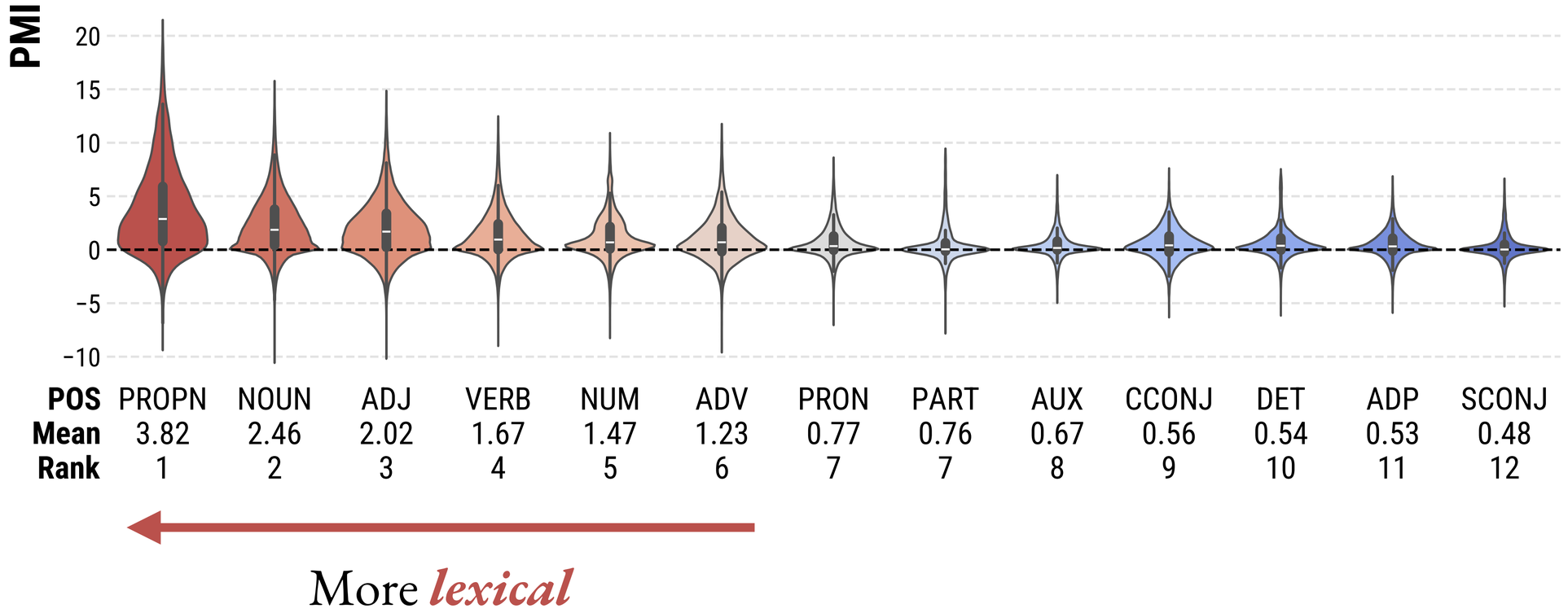
Datasets: COCO-35L; XM3600; Multi30K



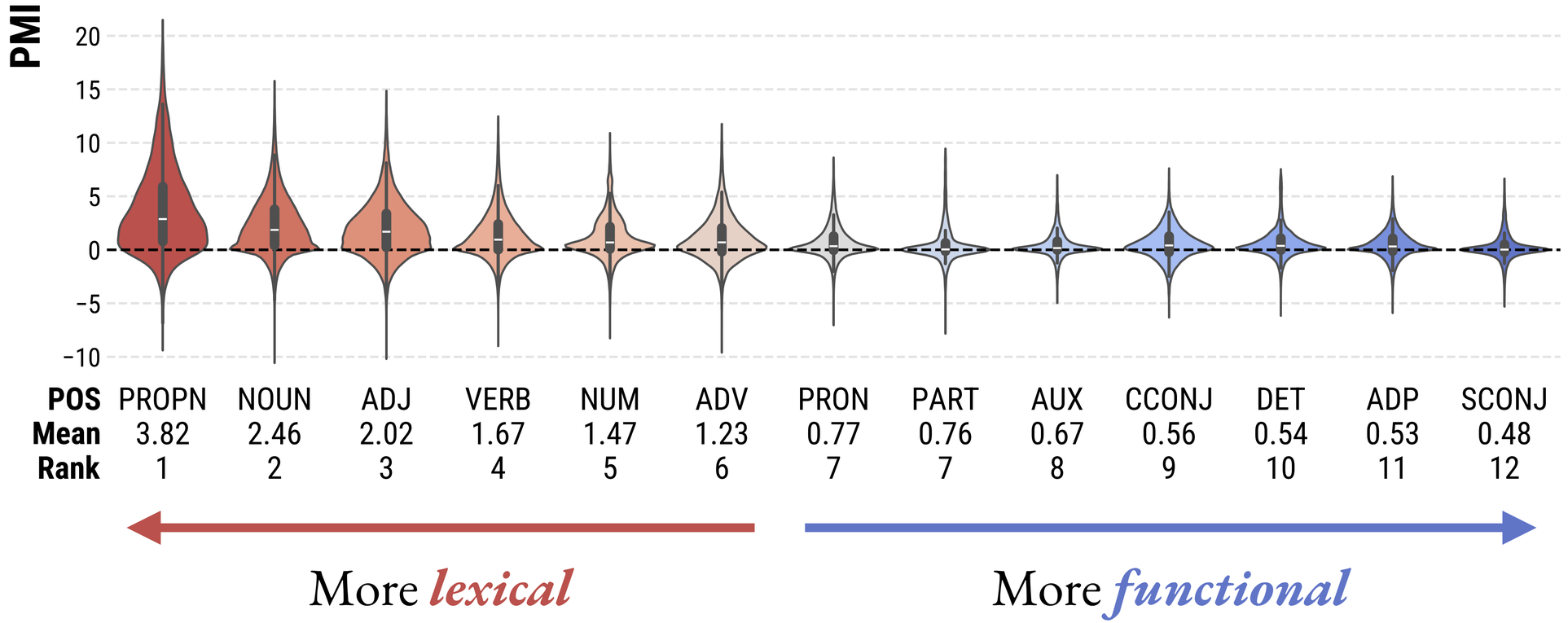
Groundedness recovers the lexical-functional distinction.



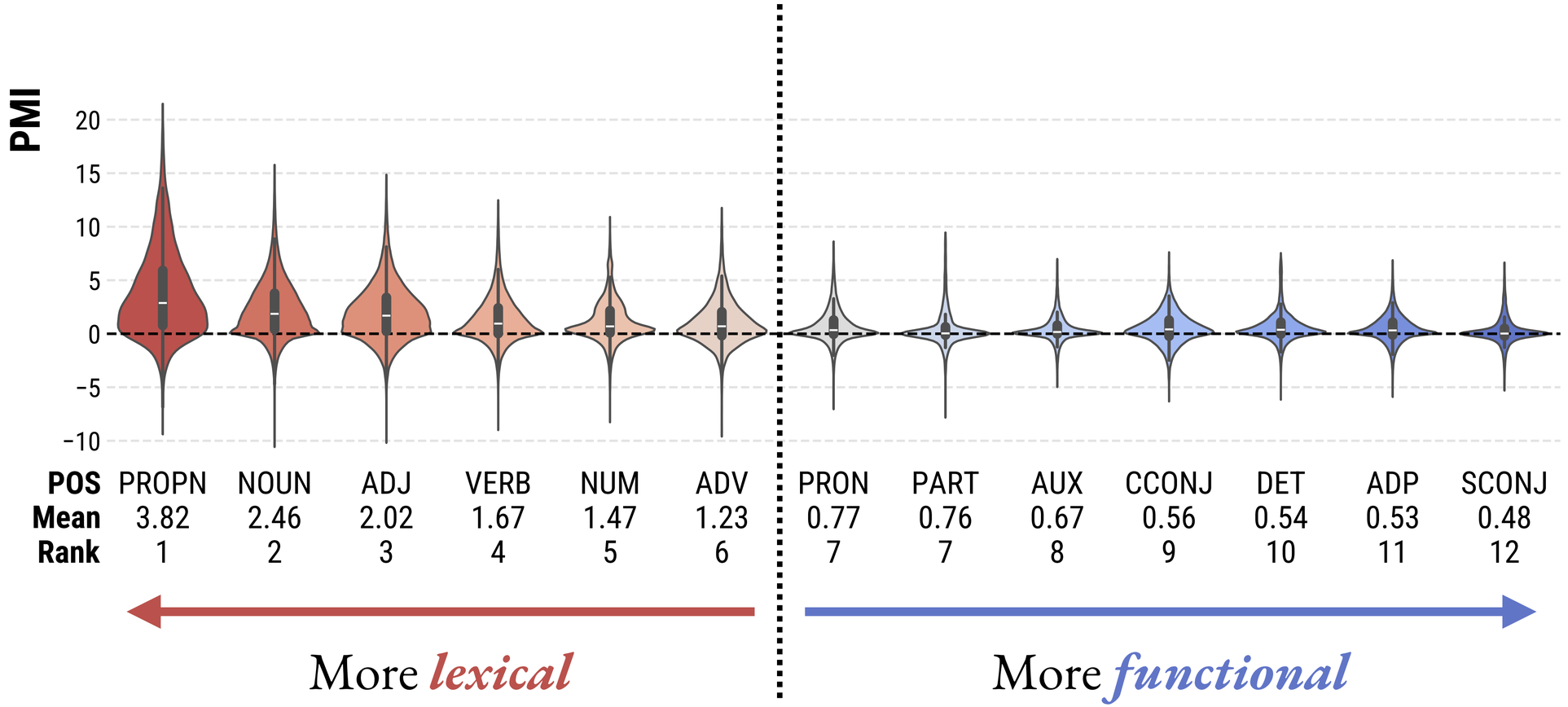
Groundedness recovers the lexical-functional distinction.



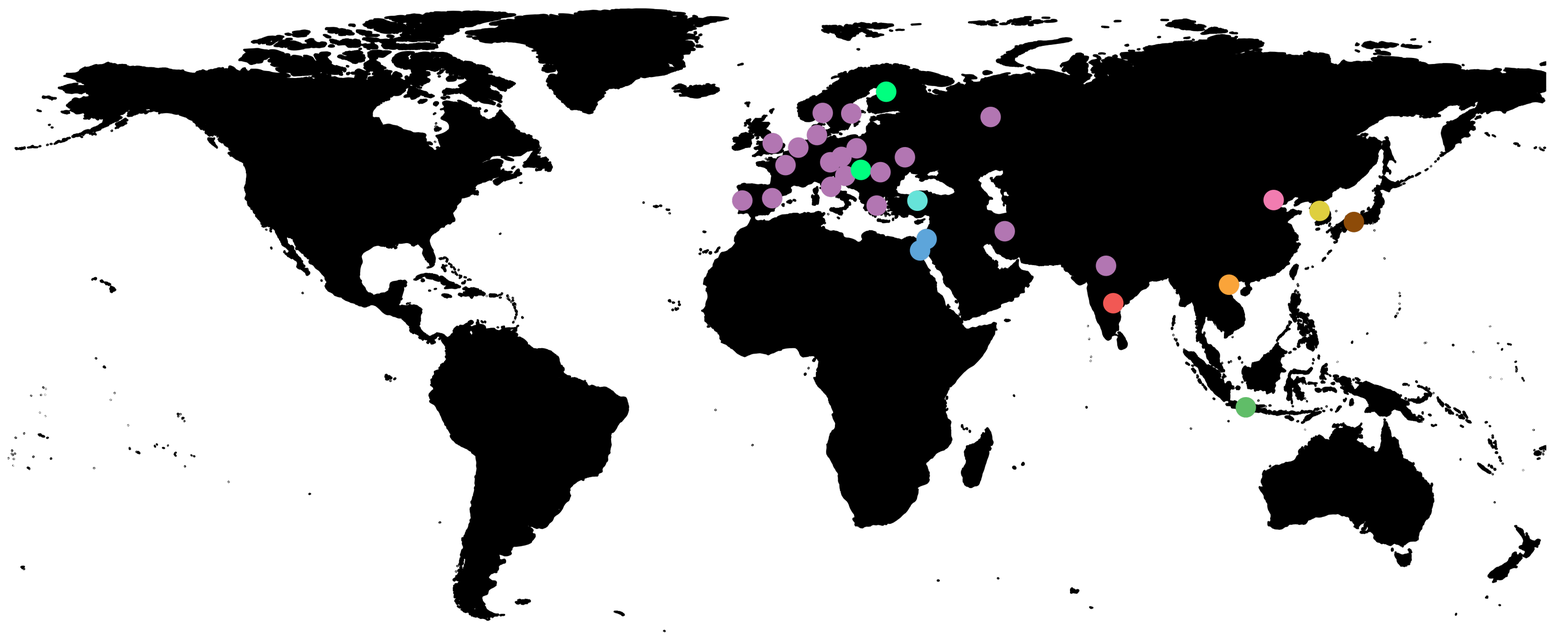
Groundedness recovers the lexical-functional distinction.



Groundedness recovers the lexical-functional distinction.

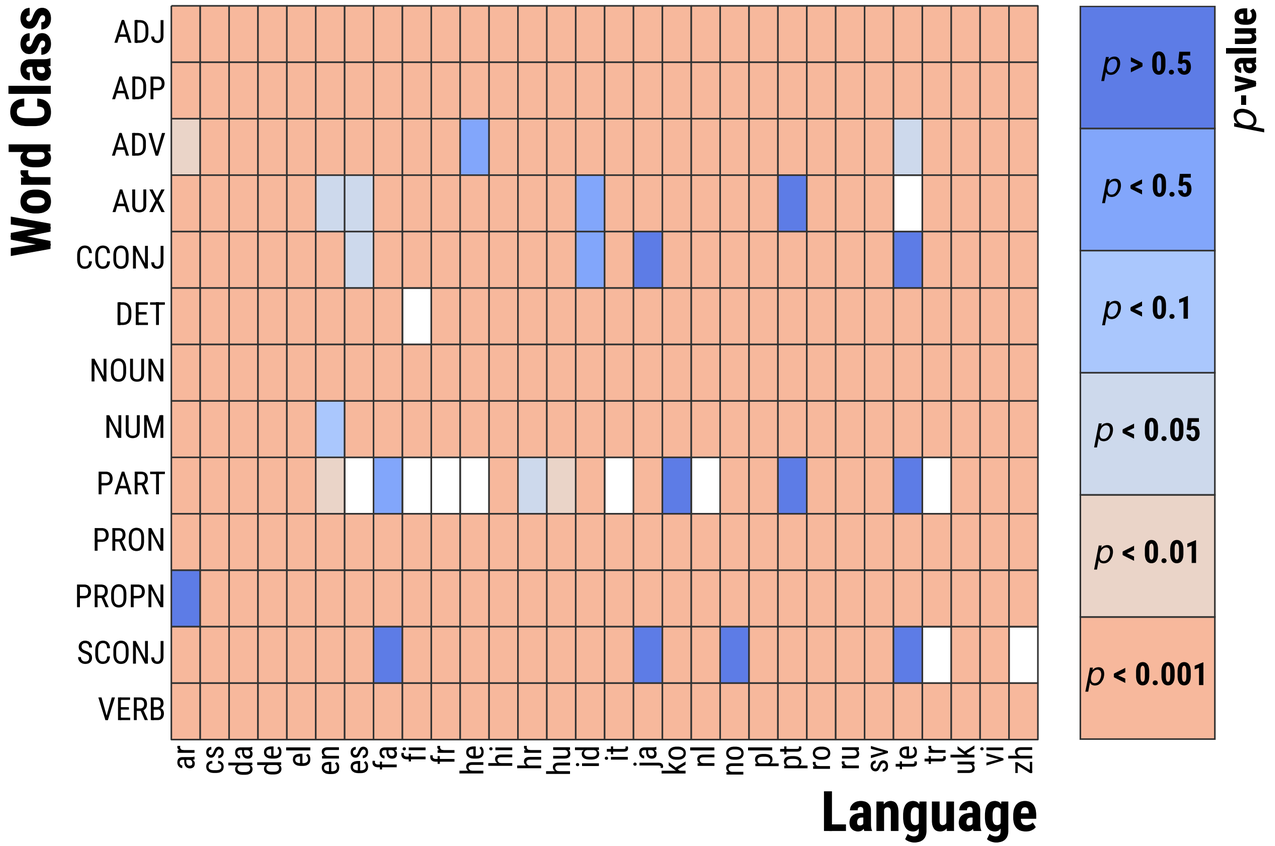


Broad cross-linguistic consistency.

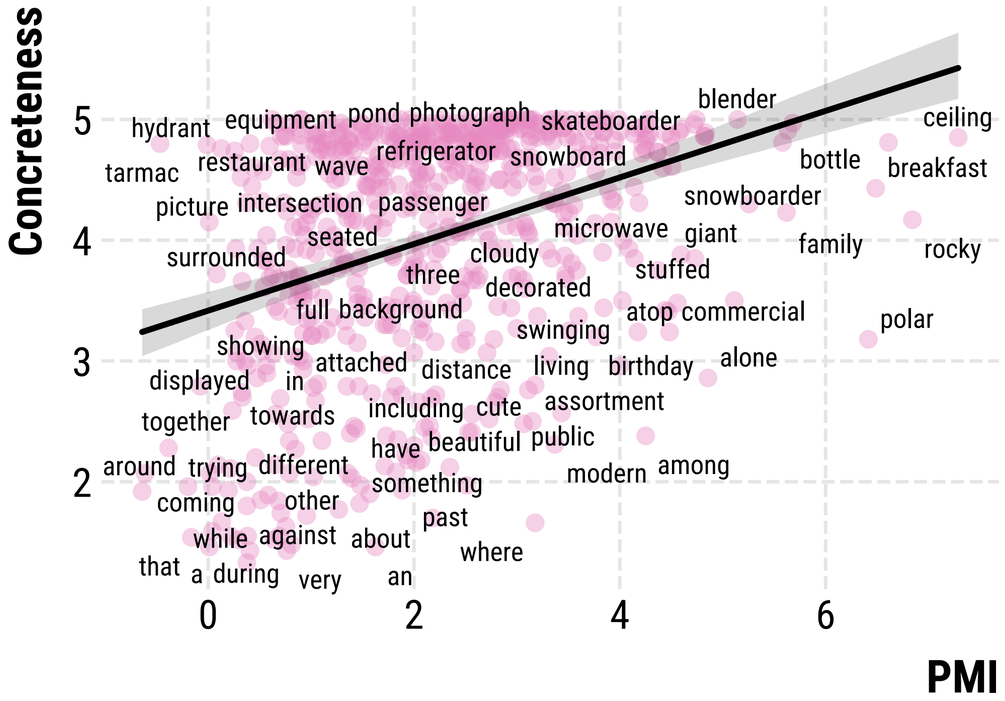


30 languages; 10 families

Even functional classes are grounded!

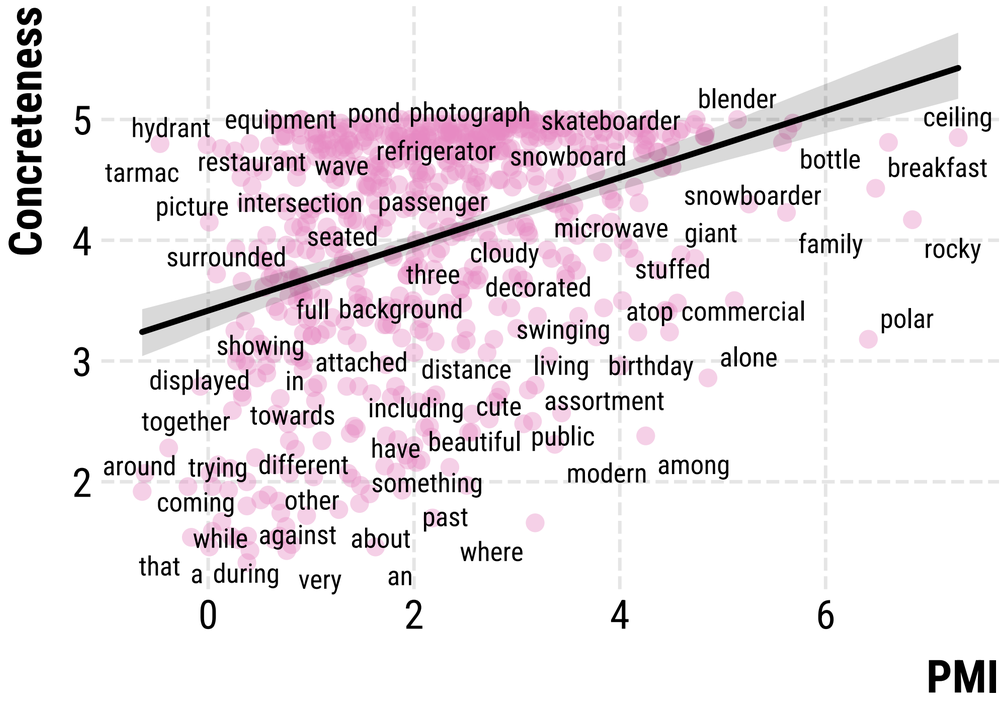


But different from existing measures...



$\rho = 0.368$

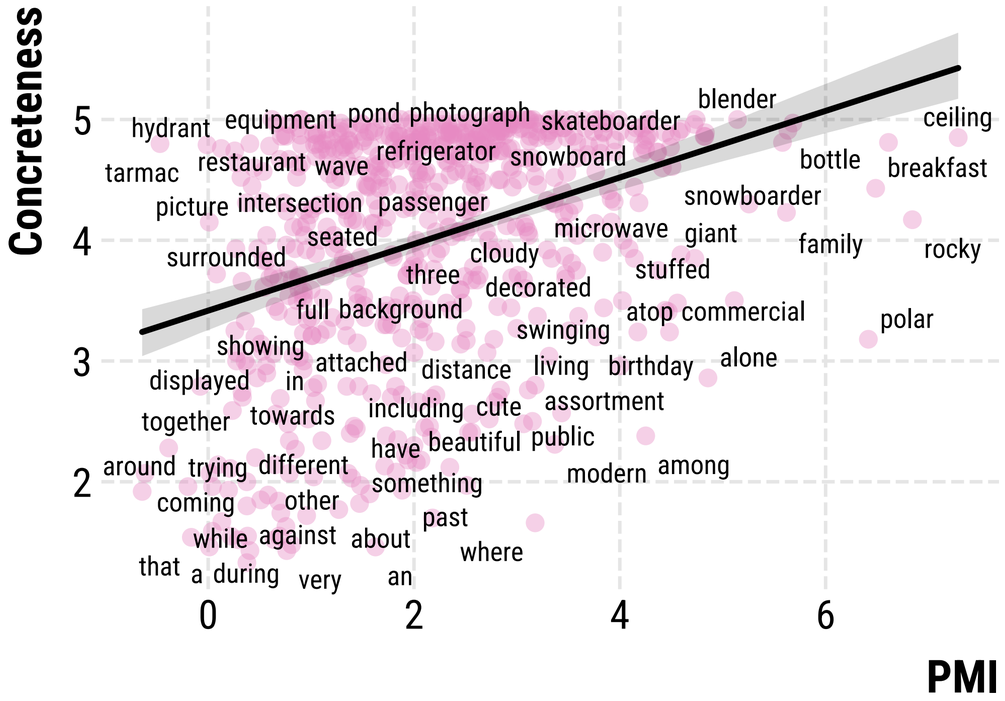
But different from existing measures...



$\rho = 0.368$

Hypothesis: caused by informativity

But different from existing measures...

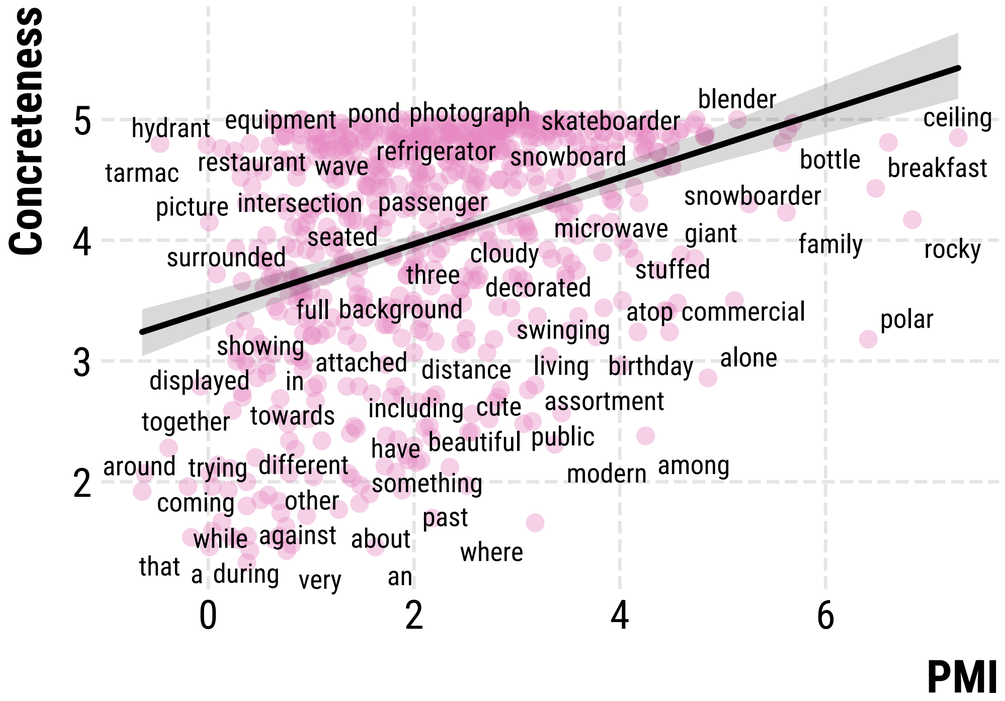


$\rho = 0.368$

Hypothesis: caused by informativity

e.g. skateboarder > person

But different from existing measures...



$\rho = 0.368$

Hypothesis: caused by informativity

e.g. skateboarder > person

Normalizing out* informativity: $\rho = 0.609$

Summary & Future directions

(1) We introduce groundedness, a token-level measure of contentfulness based on VLMs

(2) Groundedness captures the lexical-functional distinction

(3) Groundedness incorporates informativity, unlike psycholinguistic norms

(4) Potential applications:

- ✳ Analysis of items which humans struggle to score (contextual, highly grammaticalised)
- ✳ Grammaticalisation processes
- ✳ Different levels of linguistic structure (e.g. morphemes, syntax)

Thank you!

Data & Model

Paper

Me