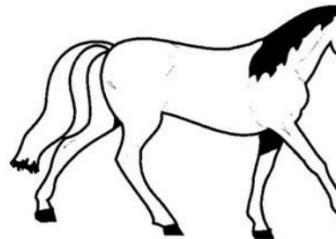


# Learnability pressures influence information density in the lexicon

*Molly Lewis and Michael C. Frank  
Stanford University*

23 March 2016

*“The linguistic sign is arbitrary” – Saussure (1916)*



horse

kalë	حصان	āḥ	at	zaldi	коń	શાડી	konj	кон	cavall	kabayo	马	馬	konj
kůň	hest	paard	čevalo		hobune		kabayo		hevonen		cheval		cabalo
Pferd	ଅଲୋଗ୍ୟ	ଘୋଡ଼ି	chwal	doki	ଧୀର	ଘୋଡ଼ା	nees	ló	hestur		anyinya		kuda
capall	cavallo		馬	jaran	କୁଦୁର୍ବ		ସ୍ତେ	ମଲ୍ଲ୍ୟା	equo	zirgs	arklys	କୋନ୍ହ	kuda
hoiho	ଘୋଡ଼	ادୁୱ	ଘୋଡ଼ା	hest	اسବ	କୋନ୍ହ	cavalo	ଷେଡା	cal	ଲୋଶାଦ୍ୟ	କୋନ୍ହ	kôň	
konj	faras	caballo		farasi	hăst	କୁତୀଷ		ଗୁରୁମୁ	ମାଣ୍ଡା	କିନ୍ହ	କୁନ୍ହା	ng̪uṛା	ceffyl

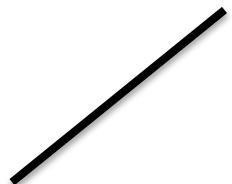
However, limits to arbitrariness

Köhler, 1929; Maurer et al., 2006; Ramachandran & Hubbard, 2001;  
Dingemanse et al., 2015; Farmer, Christiansen, & Monaghan, 2006; Zipf,  
1936; Piantadosi, Tily, & Gibson, 2011)

# Complexity Bias

A bias to map longer words (in terms of phonemes, morphemes, syllables) to more complex referents

*tupabugorn*



# Complexity Bias

Theories of communication predict tradeoff between length and predictability

## Horn Implicatures

(Horn, 1984)

*I turned on the car.*

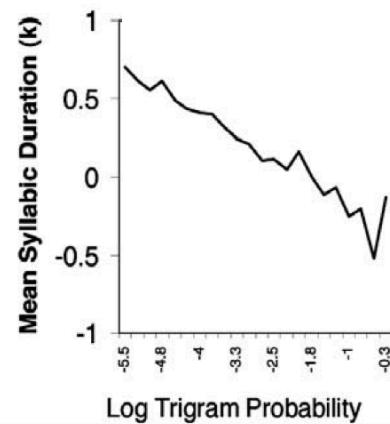
→ TYPICAL

*I got the car to turn on.*

→ ATYPICAL

## Uniform Information Density

(Aylett & Turk, 2004; A. Frank & Jaeger, 2008)



# Outline

**Study 1:** Do speakers have a productive complexity bias? \*

**Study 2:** Is there a complexity bias in the lexicon? \*

**Study 3:** Where does the lexical bias come from?

\* (Lewis & Frank; under review)

# Study 1a:

## Explicit complexity judgment

How complicated is this object?



simple

complicated



Next

Least complex



Most complex

N=60

# Study 1b: Mapping task

Map novel word to novel object, given 2 alternatives

Look at the two objects below.

Imagine you just heard someone say **fepolopus**. Which object do you think **fepolopus** refers to? Choose an object by clicking the button below it.



# Study 1b: Design

Referent complexity x word length (within subject)

Linguistic stimuli:

- short words (e.g., "bugorn," "ratum," "lopus")
- long words (e.g., "tupabugorn," "gaburatum," "fepolopus")

Referent stimuli:

- Divided objects into quintiles, based on explicit complexity norms
- Tested every pairing of quintiles (15 conditions): 1/1, 1/2, 1/3, 1/4, 1/5, 2/2, 2/3, etc.

Procedure: 8 trials/participant

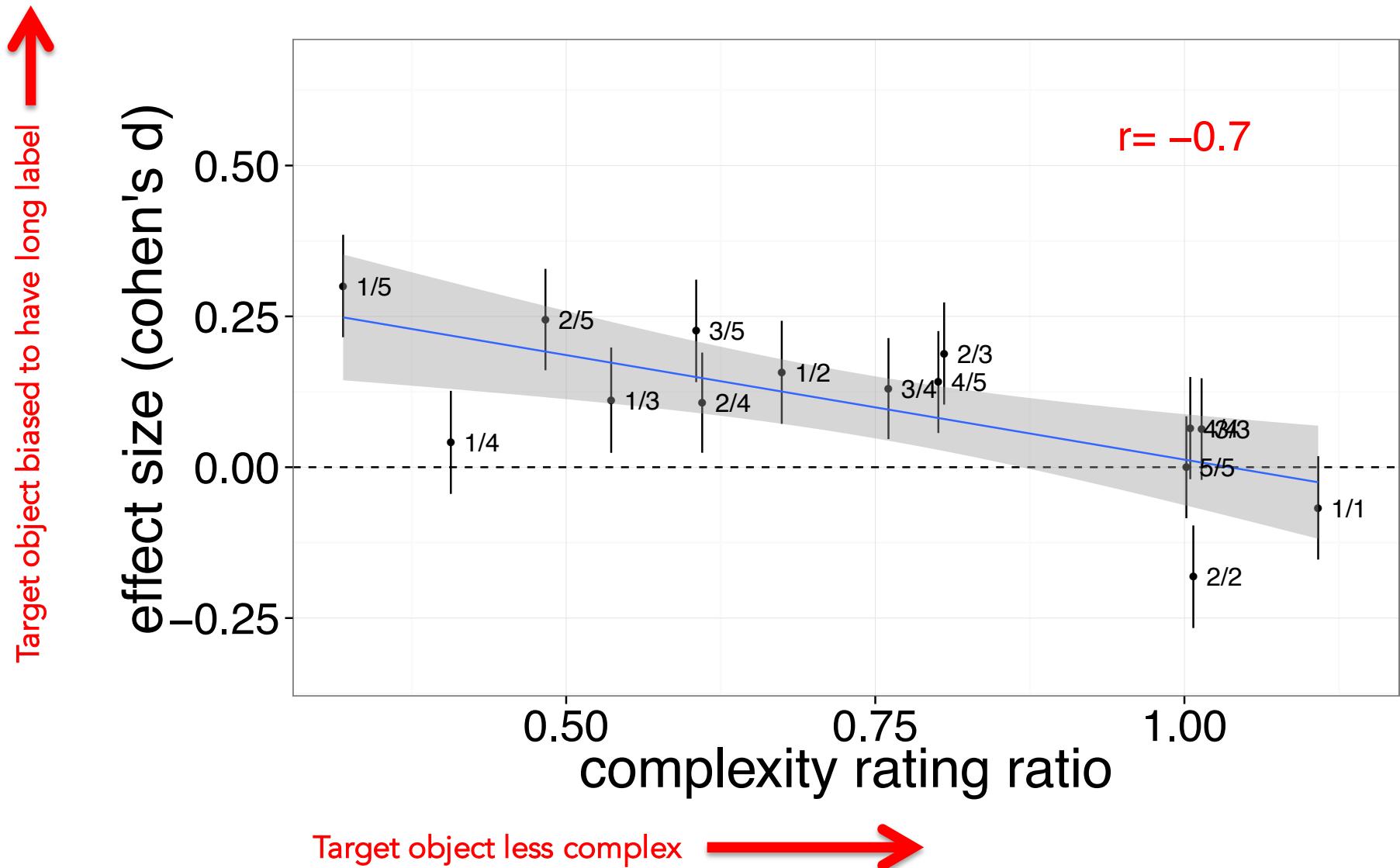
Quintile



Example 1/5 Trial



# Study 1b: Results



N = 1500

# Discussion

Study 1: A productive complexity bias with novel words and objects

Also holds for artificial objects.

Next: Is this bias present in natural languages?

Study 2: Explicit complexity norms for real English words

# Study 2a: English complexity norms

Rate words for complexity

How complex is the meaning of this word?

**alphabet**

simple       complex

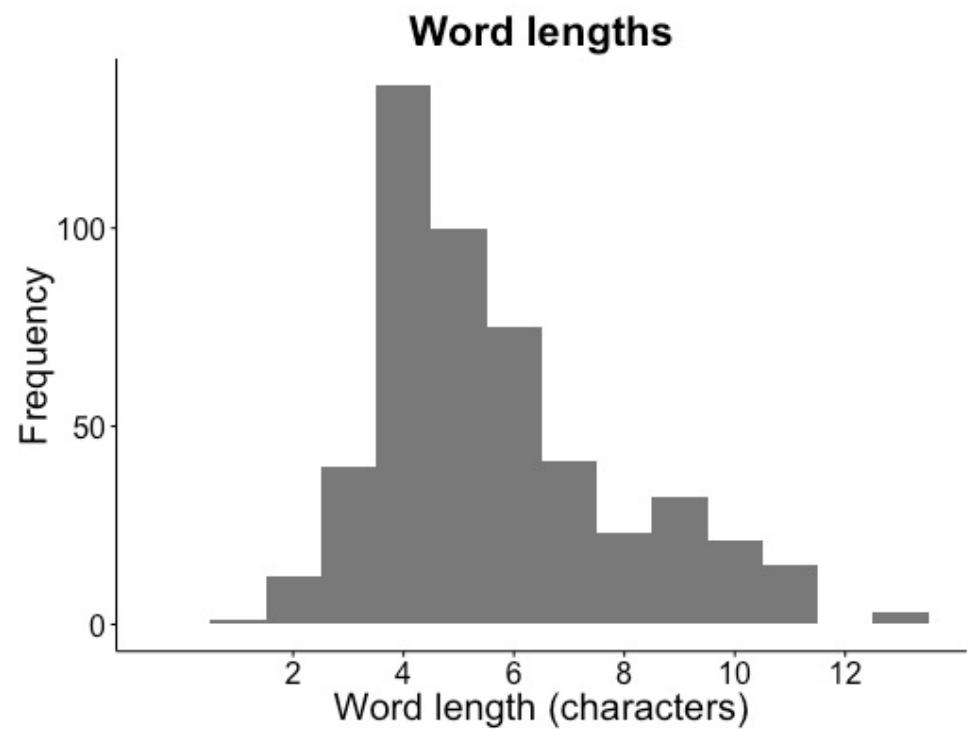
Next

# Complexity norms

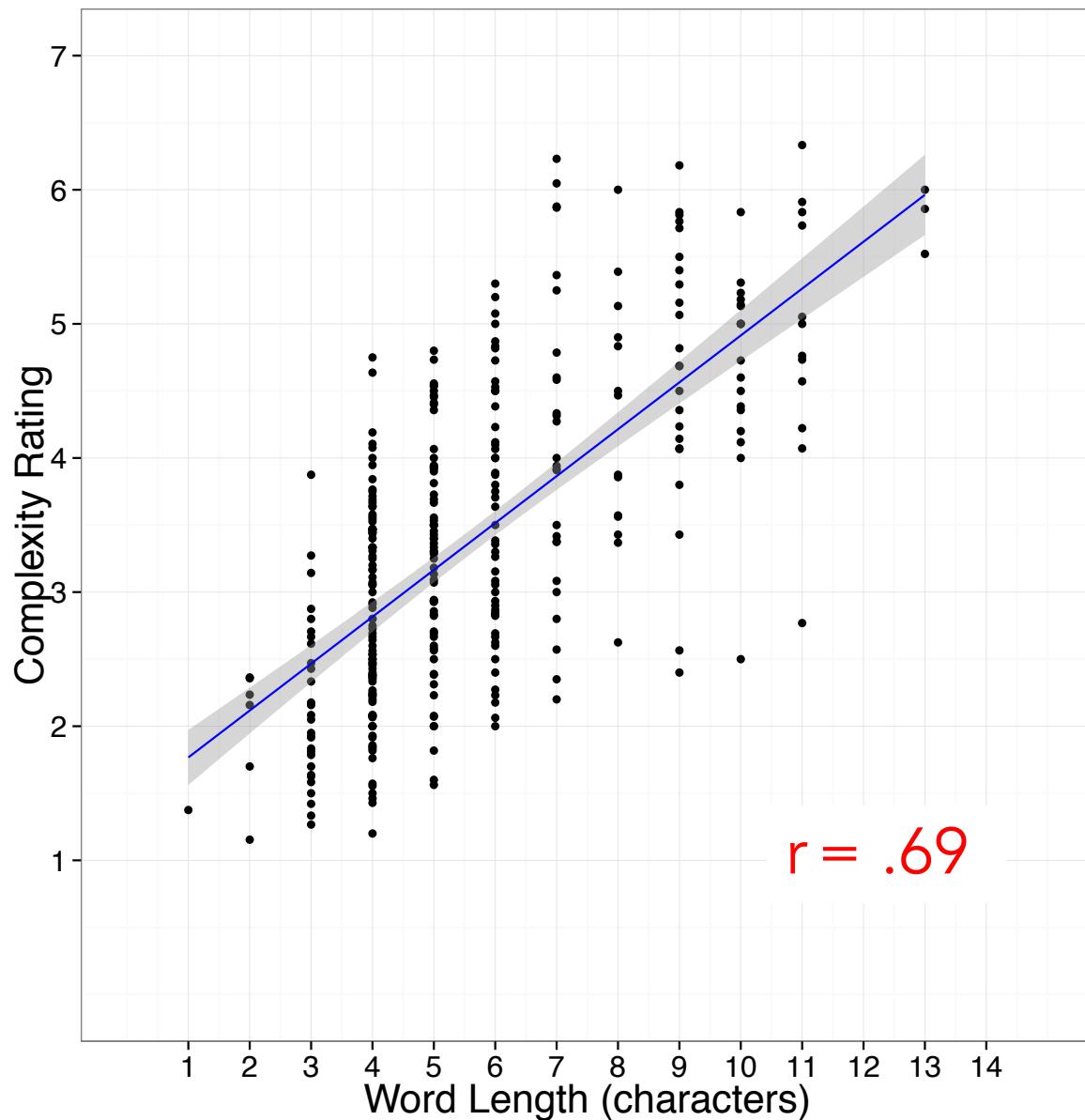
Normed 499 English words

30 words/participant

N = 250 participants



# Study 2a: Results



Log frequency  
 $r_{CL \bullet F} = .60$

Surprisal  
 $r_{CL \bullet S} = .59$

Reliable controlling for  
concreteness, familiarity  
and imagability

N = 250

# Study 2b: Cross-linguistic

Evidence that complexity is related to length in English (controlling for other semantic variables)

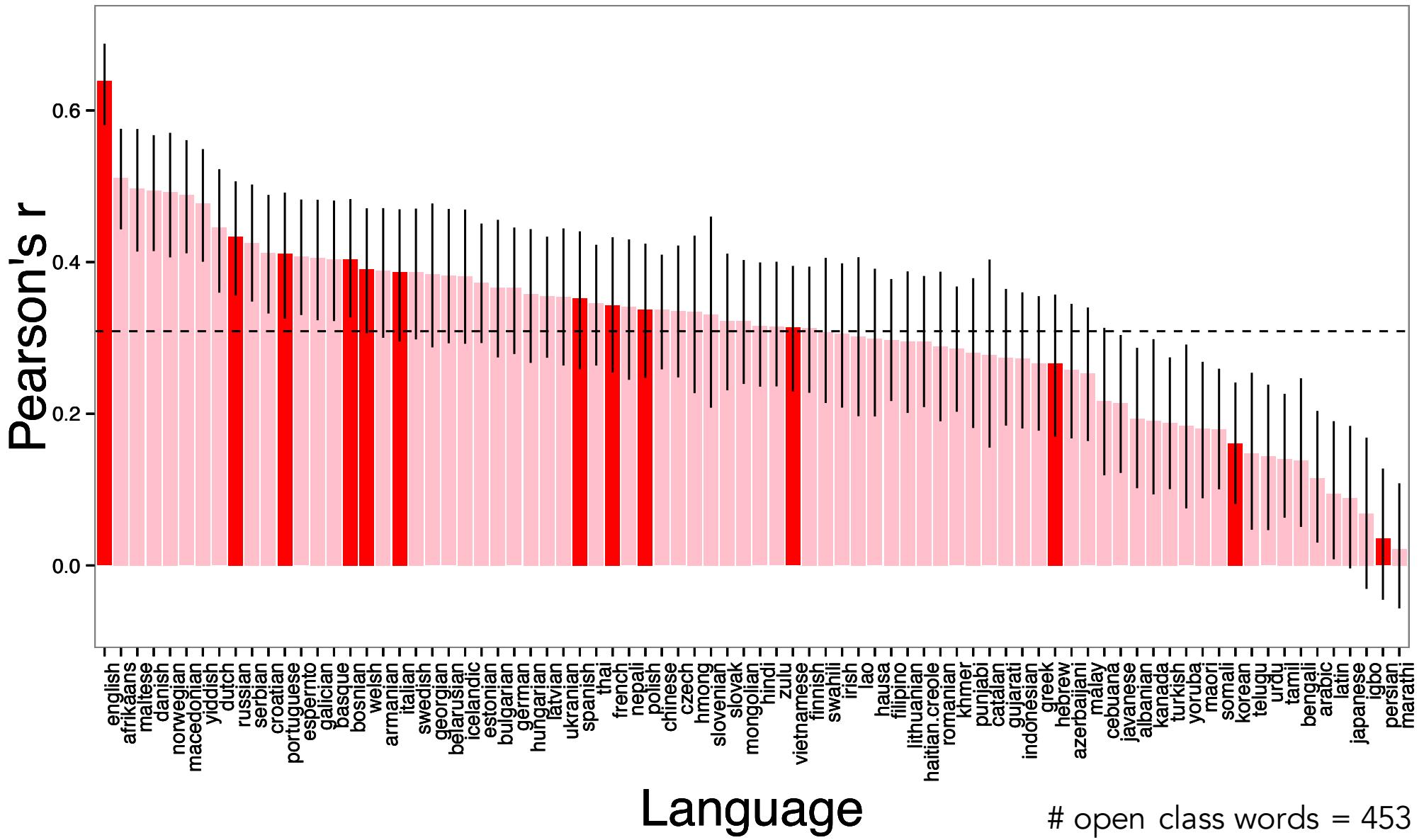
But: does this extend to other languages?

Examined relationship between word lengths for normed words in 80 languages

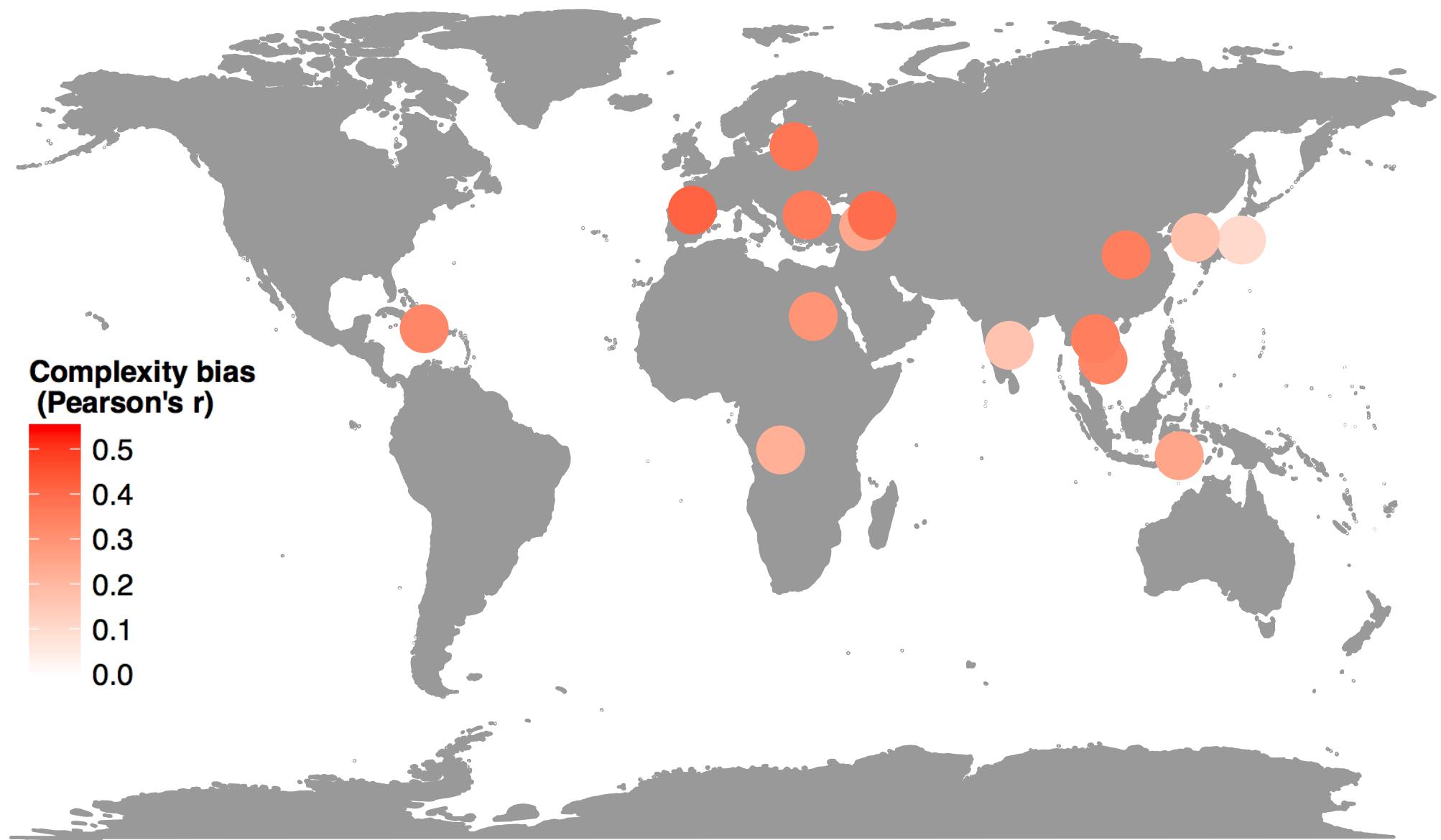
Google translate

- Native speakers hand-checked 12 languages
- Accuracy: 92%

# Correlation between complexity norm and word length



# Bias holds across language families

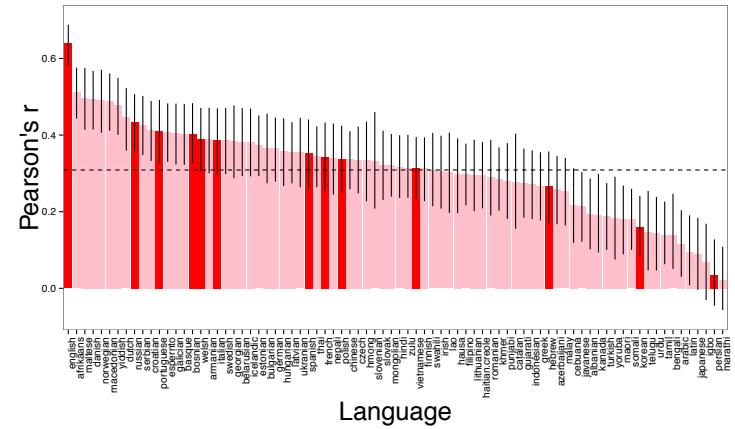


# Where does the bias come from?

Productive complexity bias



Lexical complexity bias

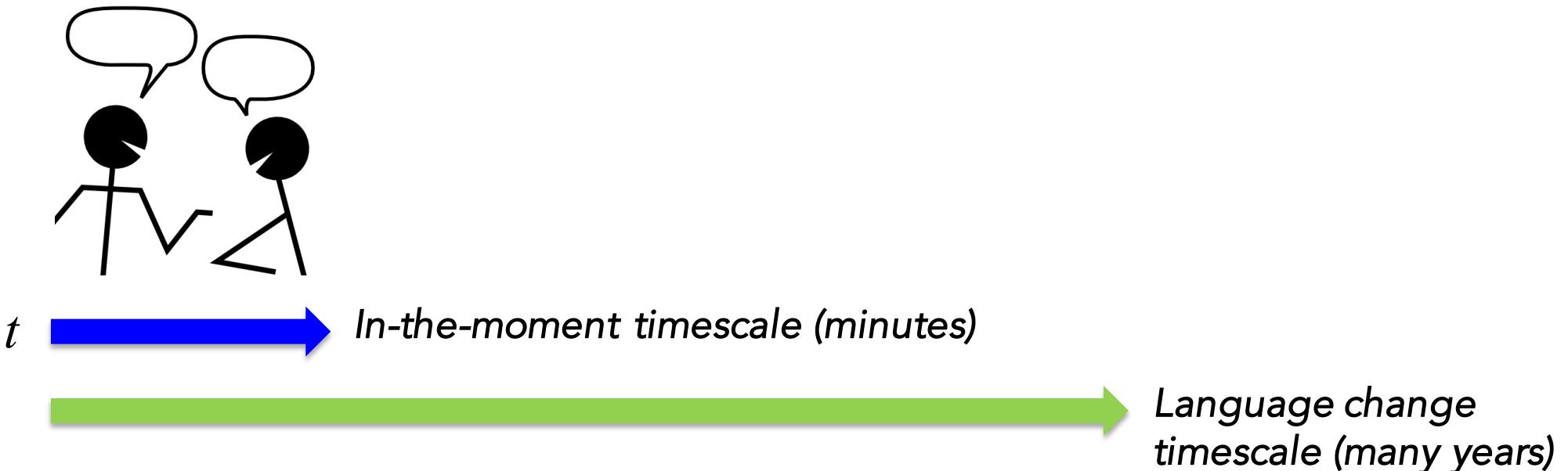


Over time, complexity bias in *individual speakers* leads to the same regularity emerging in *the structure of the lexicon*

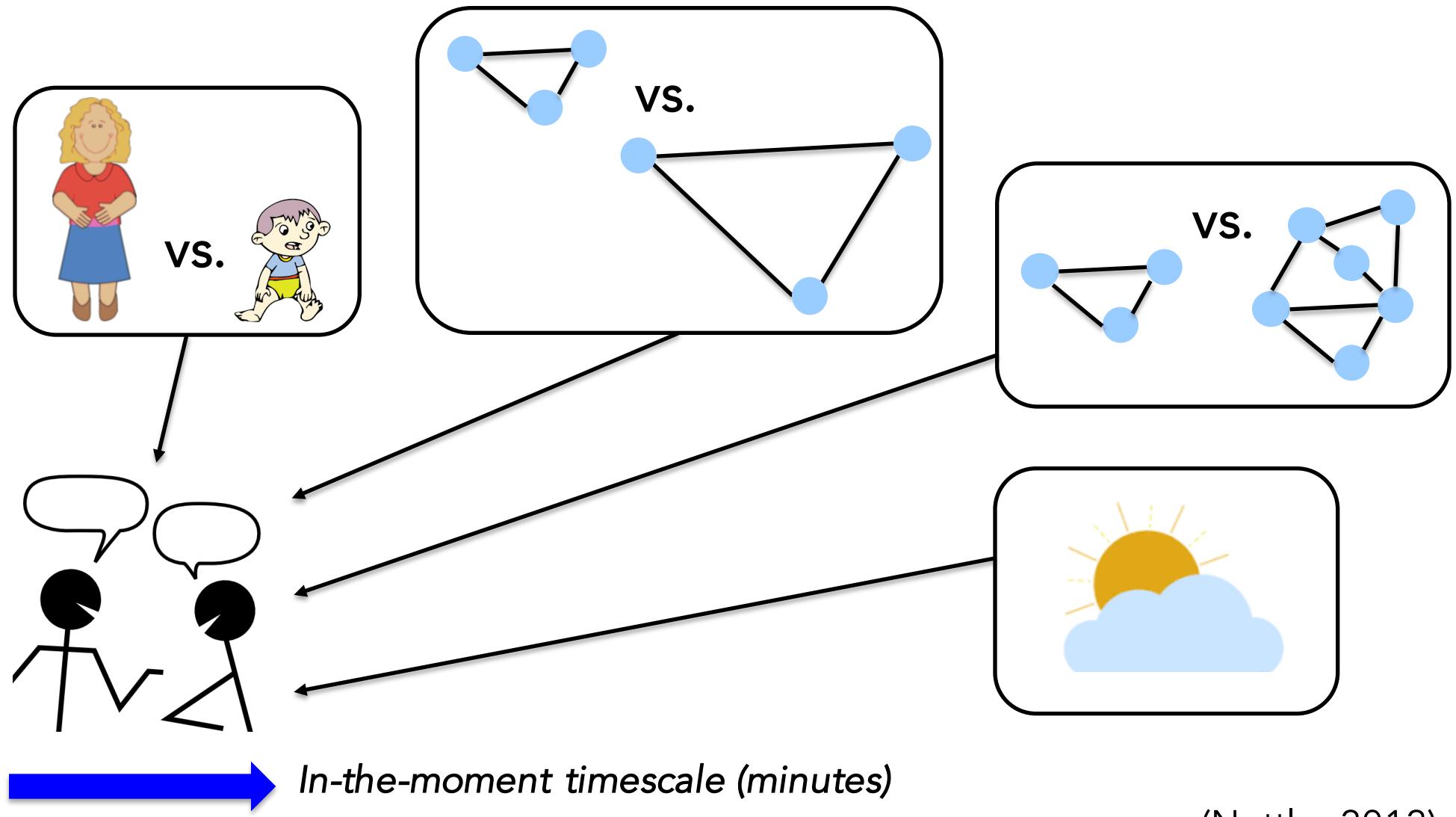
# Where does the bias come from?

Differences in cross-linguistic pressures at the in-the-moment timescale

...lead to differences at the language change timescale. (*Linguistic Niche Hypothesis*; Lupyan & Dale, 2010)

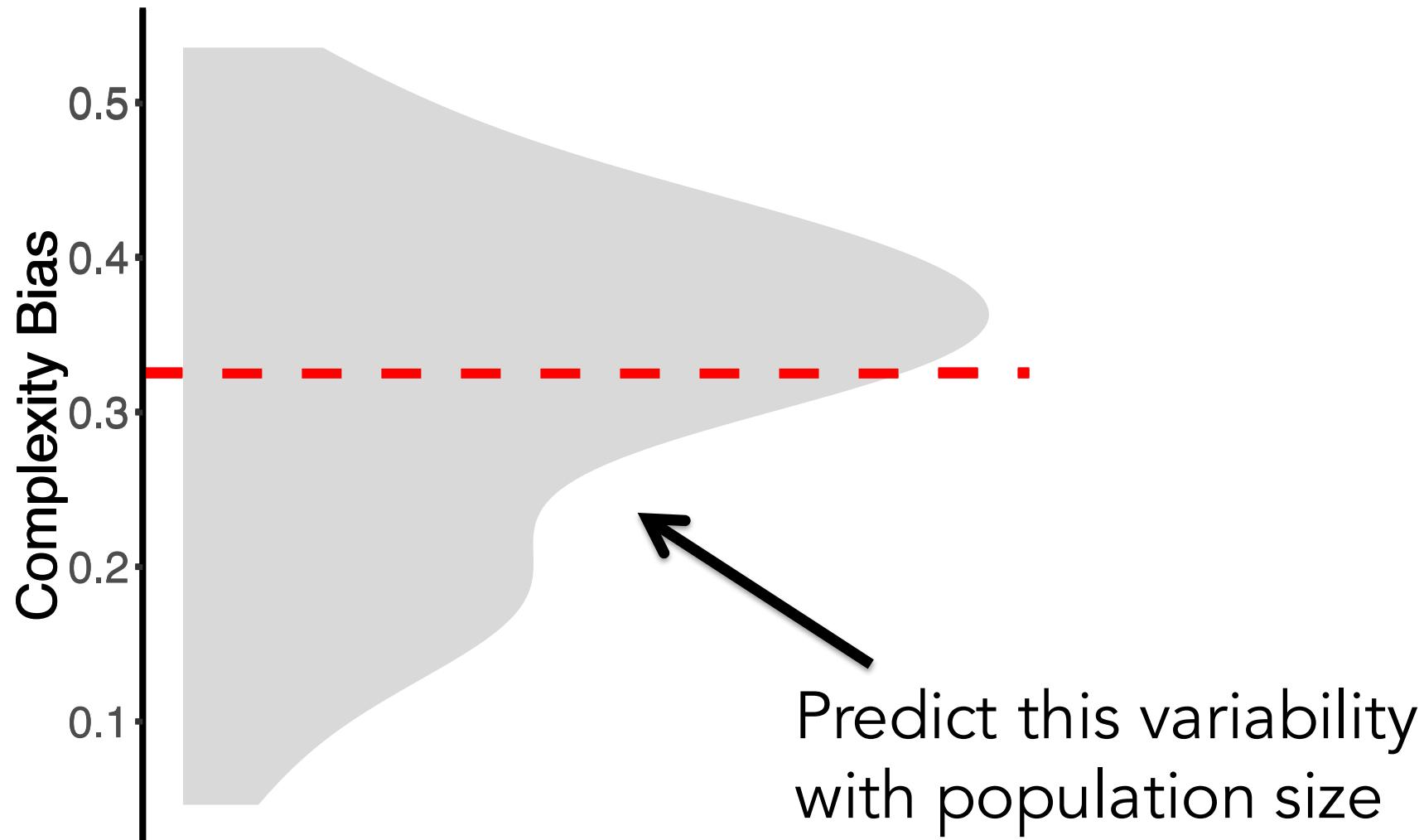


# Linguistic Niche Hypothesis



(Nettle, 2012)

# Study 3: Origins of complexity bias



# Method

Test for linear relationship between complexity bias and population size, estimated from Ethnologue (Gordon, 2005)

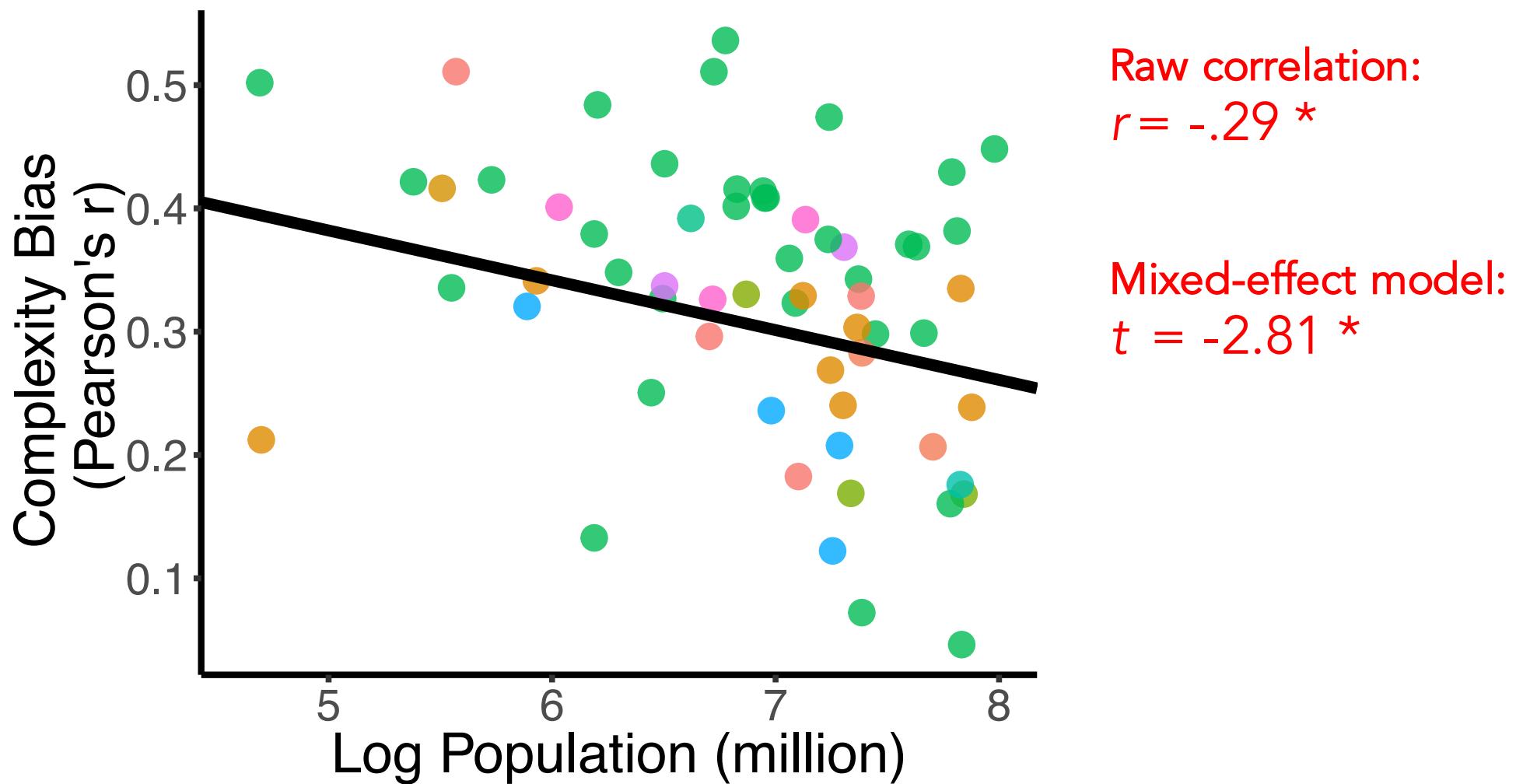
A challenge: non-independence

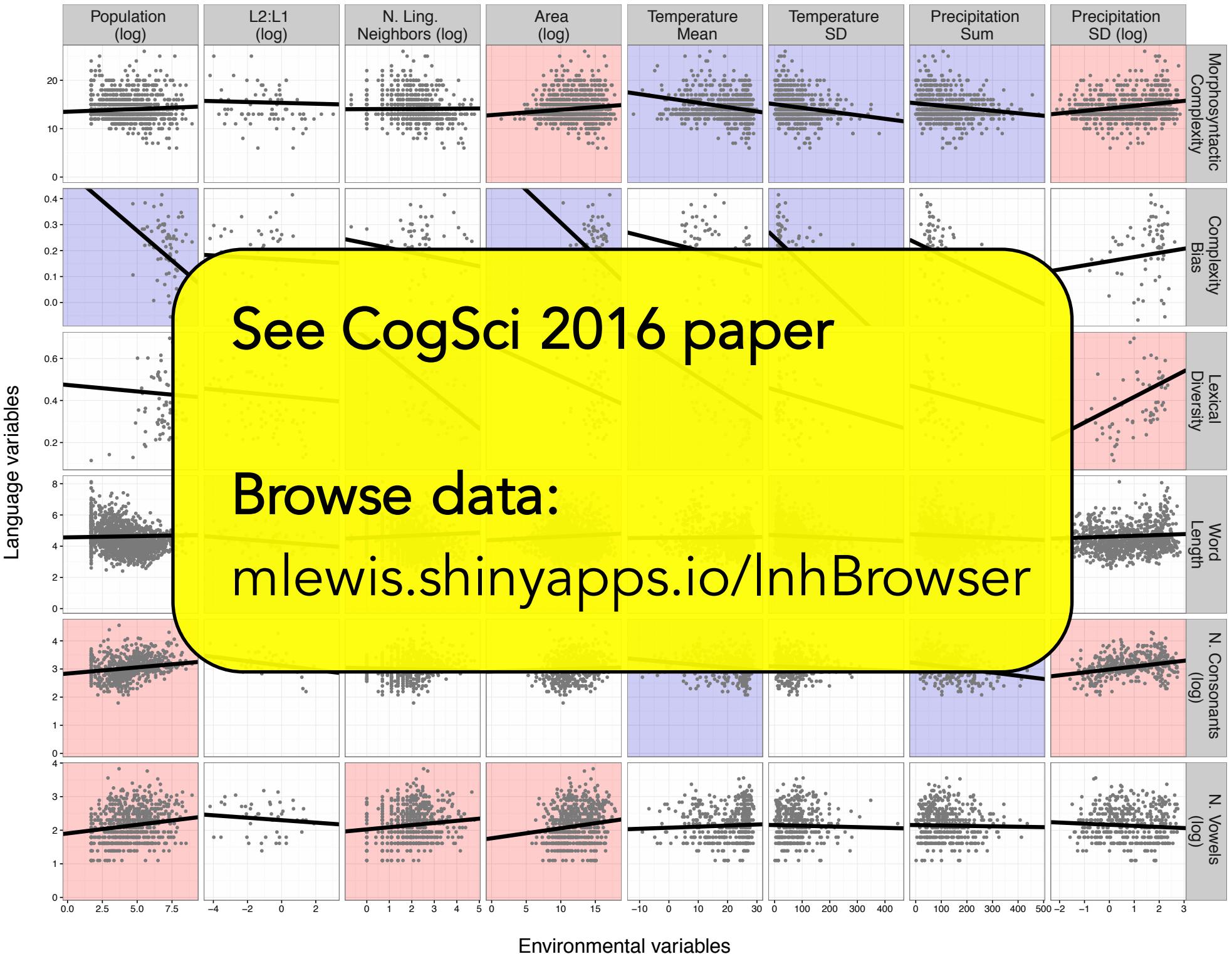
- Genetic relatedness
- Language contact

One solution: Mixed-effect models (Jaeger et al., 2011)

```
complexity.bias ~ population +  
  (population | language.family) +  
  (1 | origin.country)
```

# Languages spoken by more people have smaller bias





# Discussion

Languages with smaller complexity bias spoken in communities with fewer speakers

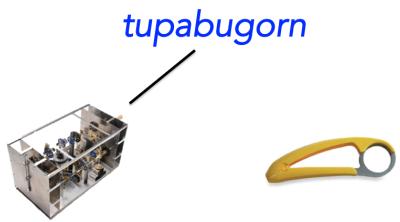
Why?

- Holds controlling for morphological complexity and word length
- Maybe related to variability in word length?

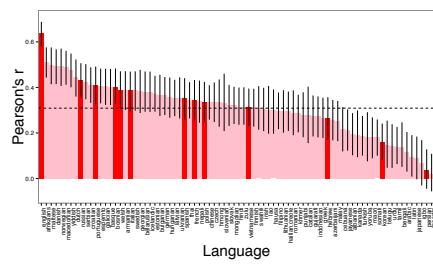
Bigger languages less information-uniform?

- Might rely on other strategies (e.g. prosody; Pellegrino, Coupe, & Marsico, 2015)

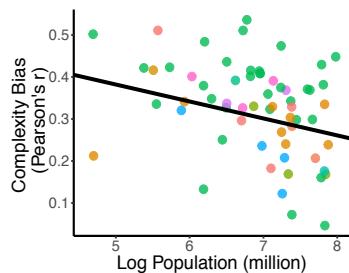
# Conclusion



**Study 1:** Speakers have a productive complexity bias.



**Study 2:** This bias is present in the lexicon of natural language.



**Study 3:** Learnability pressures shape the bias.

*Thank you*

[mlewis.shinyapps.io/lnhBrowser](http://mlewis.shinyapps.io/lnhBrowser)