Effects on the particle verb alternation in a global Twitter corpus

www.danielezrajohnson.com/johnson_probx2016.pdf

Daniel Ezra Johnson
danielezrajohnson@gmail.com
April 5, 2016

previous work on geographical aspects of the particle verb alternation

- Hughes & Trudgill 1979:25

  All speakers will accept both forms as normal English, but speakers in the south of England are more likely to employ the [VOP] forms in their own speech, whereas Scottish speakers almost invariably use [VPO] forms.

- Hughes, Trudgill & Watt 2013:23

  …….. whereas Scottish speakers very frequently use [VPO] forms.

- Trudgill p.c

  one of those things that is 'apparent to inspection' for anyone who has lived in Britain all their life
  [re: aux-neg contraction]

- Cappelle 2009, Lohse et al. 2004

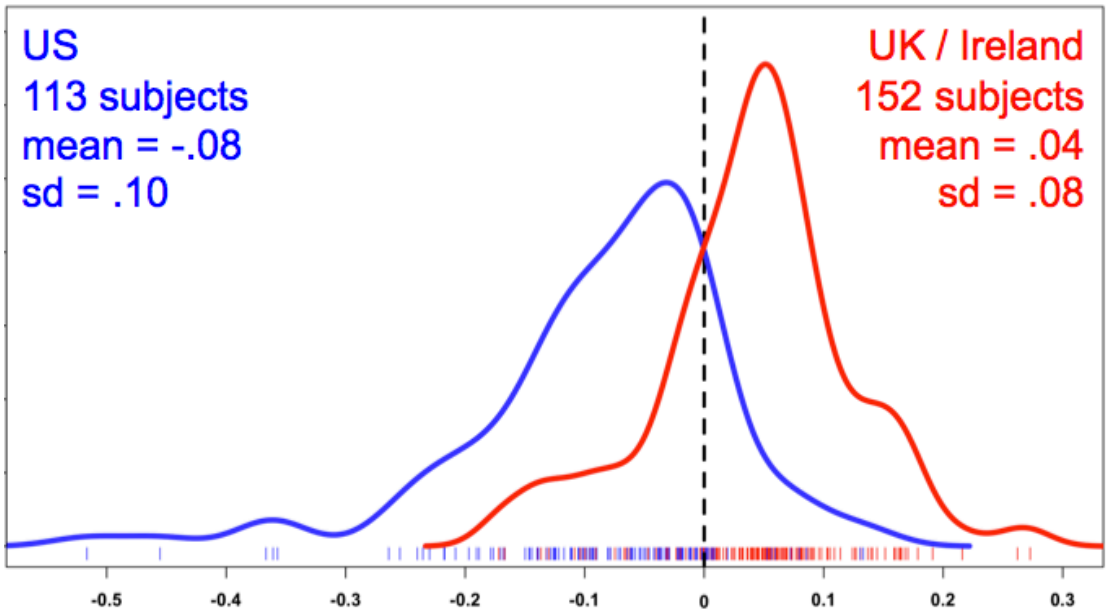
  I notice that their written British English material, taken from the Lancaster-Oslo/ Bergen corpus, contains a higher percentage of split orderings than their combined material from corpora of American English, more than half of which even consists of phone conversations (the Switchboard corpus).

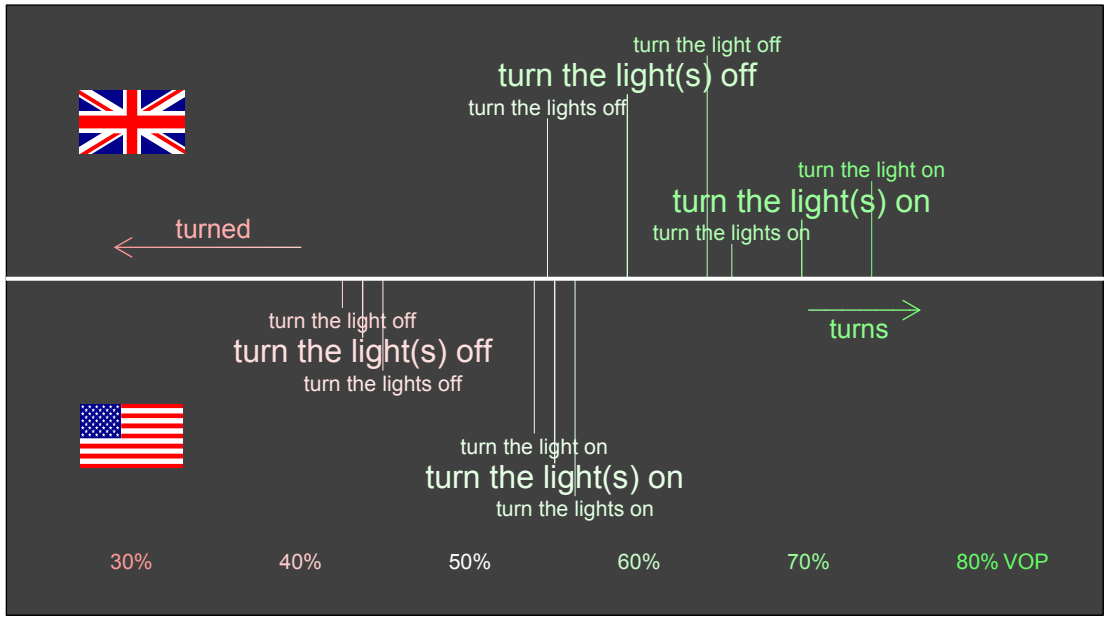
- Haddican & Johnson 2012

  judgment experiment (many PV)

  Twitter survey (few PV)

  diachronic story (important!)





design of the current study
use of common particle verbs on Twitter (12/15-3/16)

- Gardner & Davies 2008
  - found most frequent "phrasal verbs"
  - included any VP: GO OUT, COME BACK
  - these do not all show an alternation
  - top 25 PV: 30.4% of PV; top 50: 42.7%

  - chose 12 PVs from the top 100:
  - verb lemmas BRING, PUT, TAKE
  - particles BACK, DOWN, OUT, UP

  - together, 5.3% of PV in Gardner & Davies (but really more)

- Twitter search using *streamR* package
  - saved all geolocated tweets
  - filtered by V (…) P, then by country
  - manually checked for "alternability"

- eight countries with the most data
  - each PV: 1000 from USA, 200 from UK
  - 50 each from Australia, Canada, Ireland, South Africa; India, Philippines

THE PLAN	bring	put	take
back	1500	1500	1500
down	1500	1500	1500
out	1500	1500	1500
up	1500	1500	1500

design of the current study
use of common particle verbs on Twitter (12/15-3/16)

- 1000 from USA, 200 from UK
- 50 each from Canada, Australia, Ireland, South Africa; India, Philippines
- unclear why availability of data does not match (English-speaking) population
- did not deal with idiomatic meanings of each verb-particle combination (as Jason predicted), some of which are variety-specific
- did not deal with common collocations (e.g. random effects)
- did remove clear fixed idioms
- guided by principle of accountability (alternability)
- subjective, but hopefully consistent across varieties

THE REALITY	bring	put	take
back	1500	Aus. Ireland 1344 S.A. Phil. India	Aus. Ireland 1431 Phil.
down	Ireland 1456 S.A. .	Ireland 1476 India	1449 S.A. Phil. India
out	Ireland 1472 S.A. .	Ireland 1411 S.A. Phil. India	1498 Phil.
up	Ireland 1481	1500	1497 S.A. .

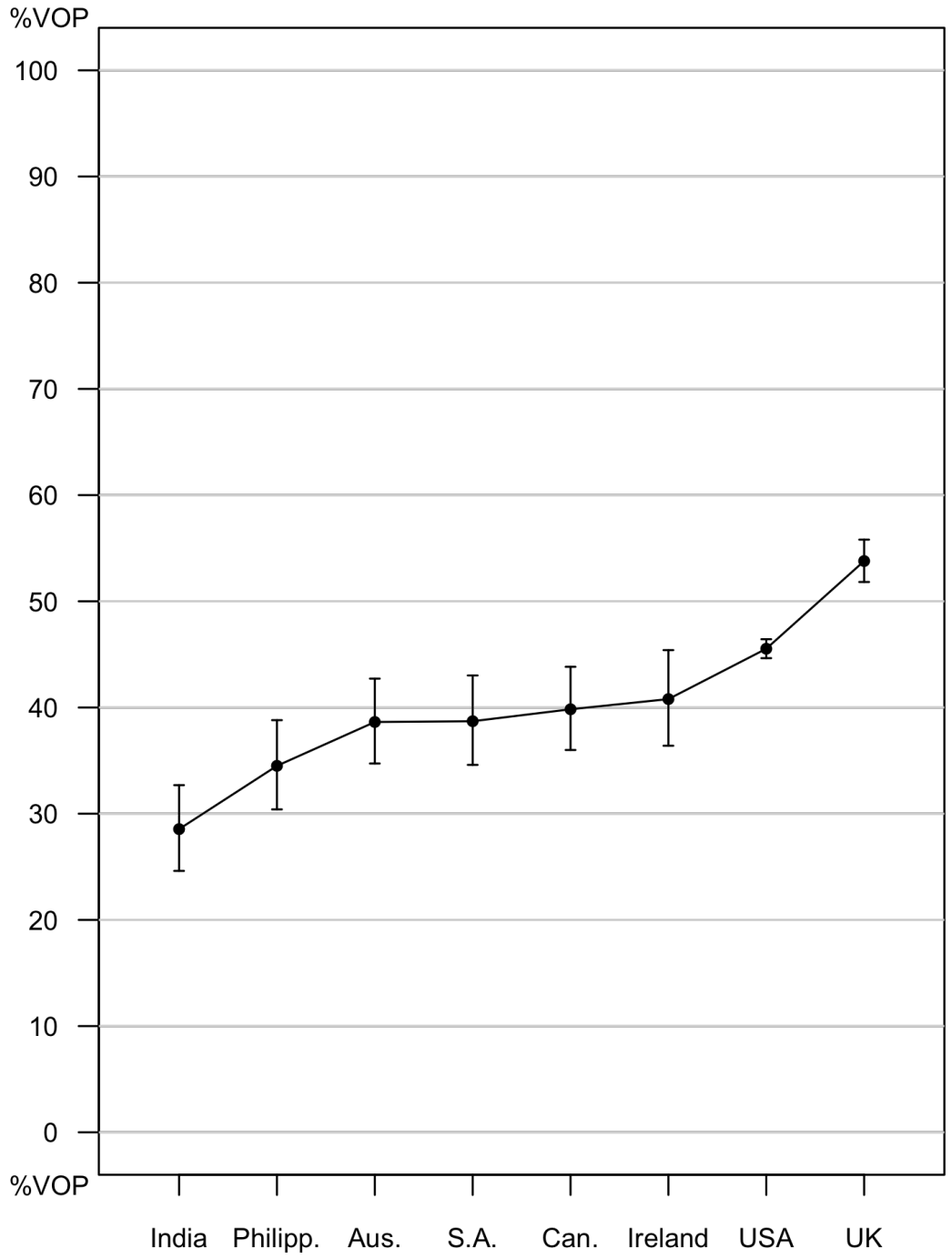
percentage of VOP order
average of 8 countries
by verb-particle combination

- no neat patterns here
- of the verbs, PUT is most associated with VOP order
- of the particles, UP is least associated with VOP order
- the highest level of VOP is found for PUT BACK (86.1%)
- the lowest level of VOP is found for TAKE UP (4.4%)
- VOP favored for discourse-old objects; high levels of PUT BACK and PUT DOWN make sense, at least with their literal meanings
- VOP disfavored for new objects; low levels of TAKE UP, BRING UP, BRING OUT make sense
- low BRING BACK, relatively high TAKE OUT more surprising
- “bring back memories” etc.

% VOP	bring	put	take
back	24.0	86.1	48.8
down	52.9	67.0	37.5
out	18.2	38.1	50.8
up	15.7	37.1	4.4

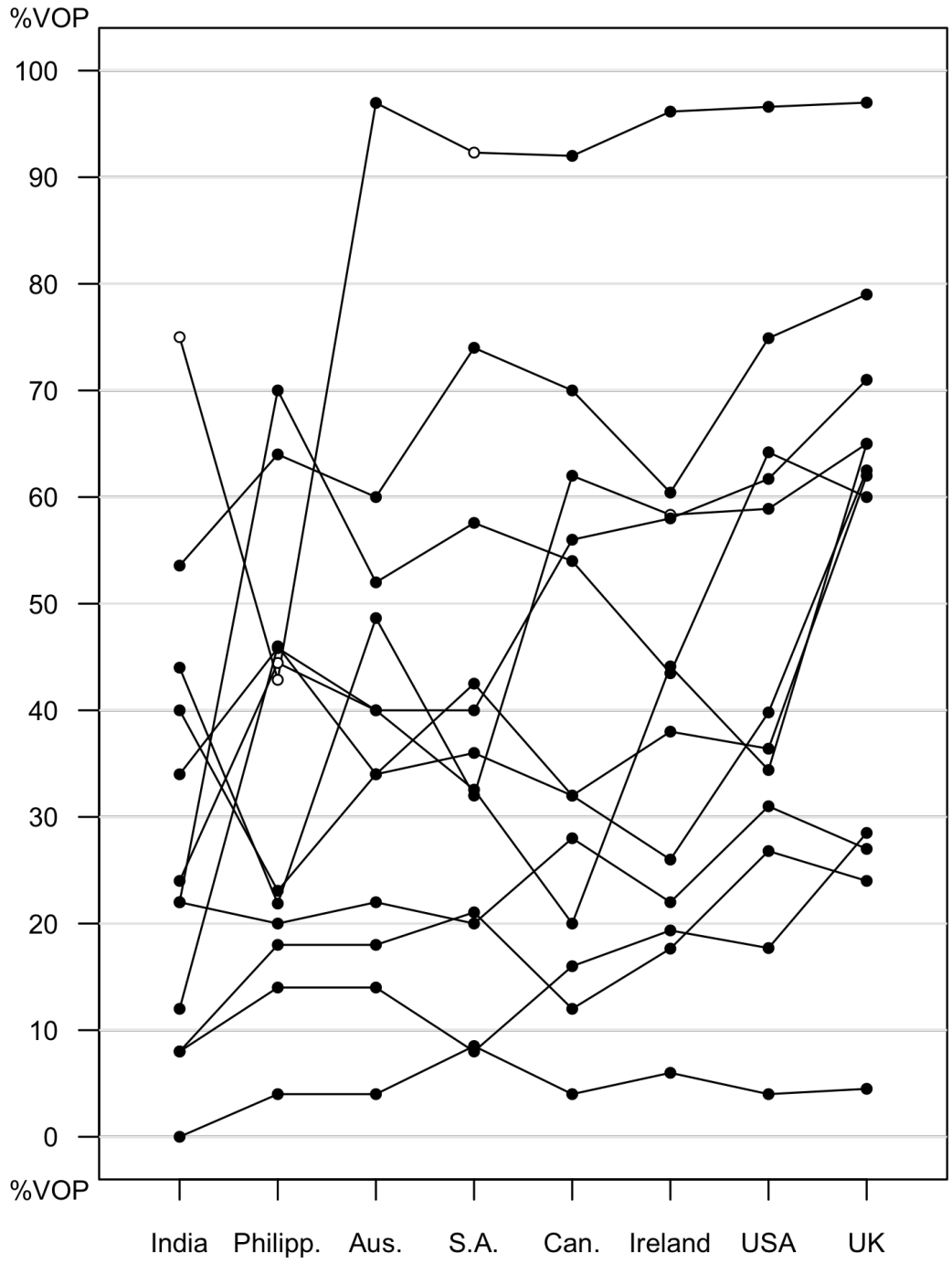
percentage of VOP order average of 12 verb-particle combinations, by country

- the two non-native varieties have the lowest levels of VOP: India lowest at 28.5% (+pronouns) Philippines next at 34.5%
- Australia, South Africa, Canada, and Ireland between 38% - 41%
- USA significantly higher at 45.5%
- UK considerably higher at 53.8%
- still, this UK-USA difference is smaller than in previous Twitter study: UK 64% vs. USA 47%
- suggests that the difference between UK and USA (and between other varieties) depends on the verb-particle combination
- colonial lag (but USA…)
- if less UK-centric…
- *the rate of VOP is much higher than in Jason's data: register*
- *less clear a split between native and non-natives: despite register*



percentage of VOP order by verb-particle combination by country

- the data speaks for itself

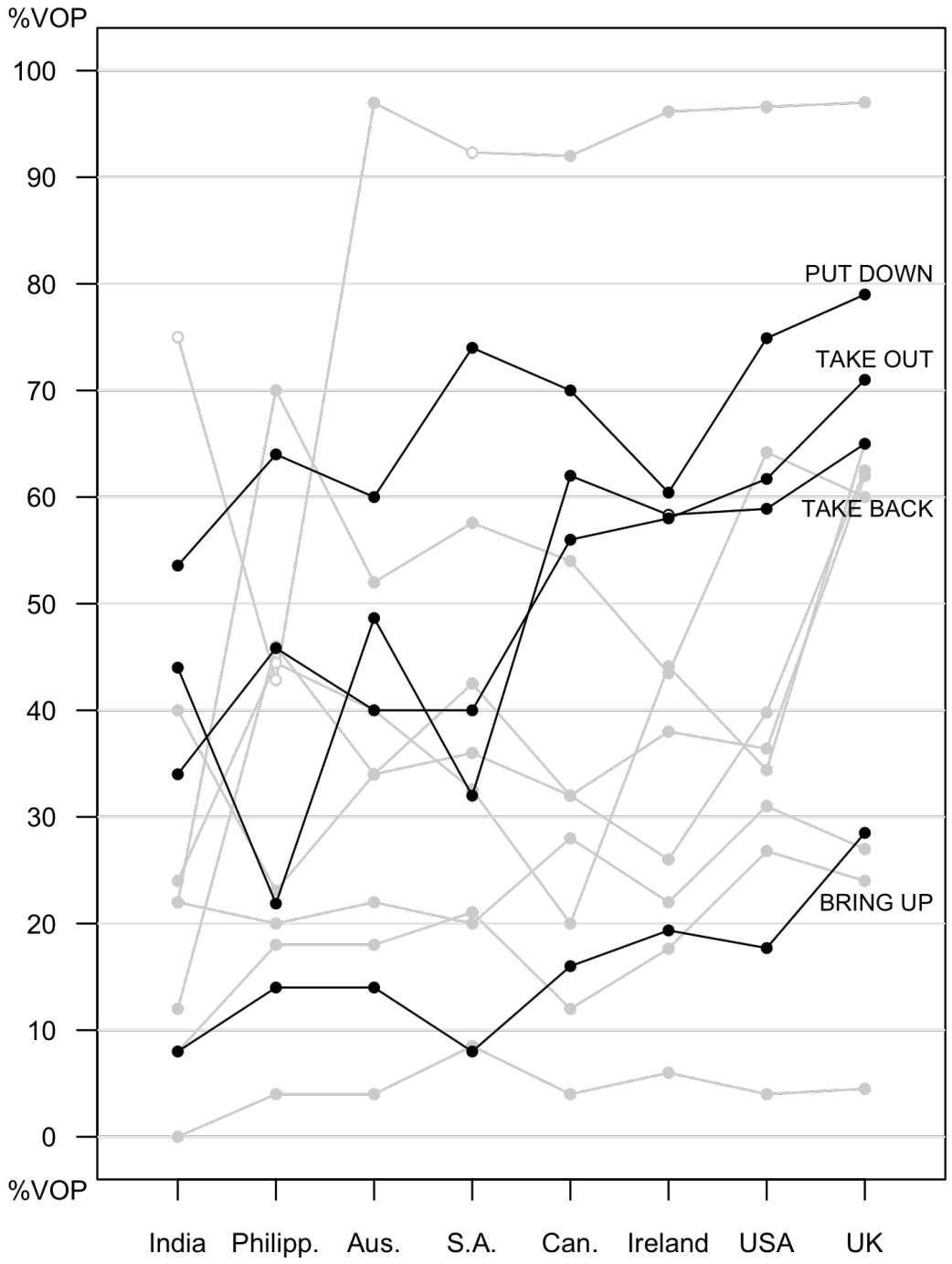


percentage of VOP order by verb-particle combination by country (normal)

- PUT DOWN, TAKE OUT, TAKE BACK, BRING UP
- large difference in rates
- similar pattern of "constraint"
- similar to average of varieties, India/Philippines/Australia low, UK always highest
- but only 5-10% above USA

- VP differences not necessarily inherent to the items themselves
- did not control for properties of object or other regularities of environment
- two types: semantic/pragmatic
  e.g. PUT DOWN [something "up"]

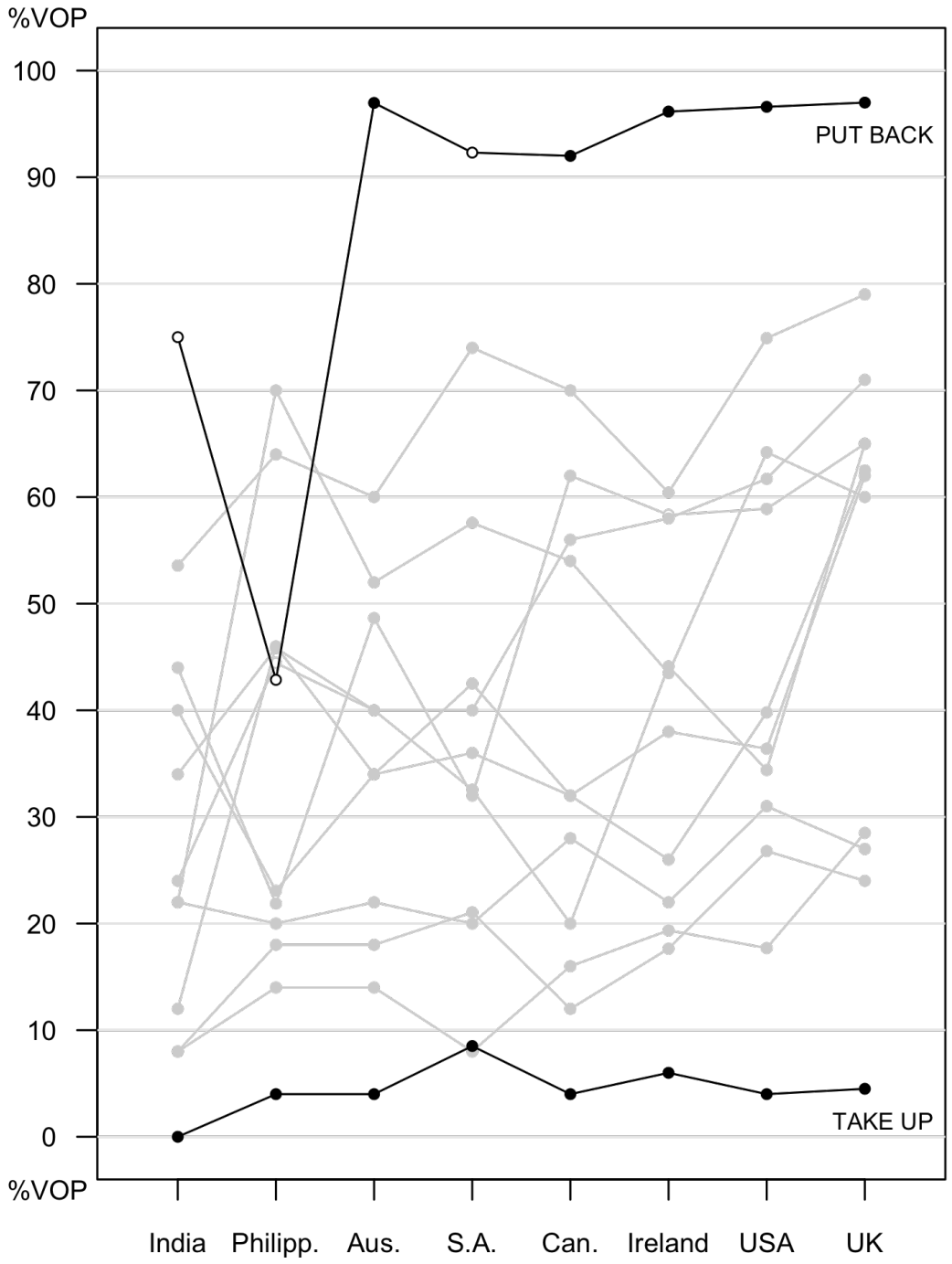
and high-frequency collocations
e.g. PUT DOWN my phone,
TAKE OUT the trash (USA) vs.
PUT OUT the bin(s) (UK)!



percentage of VOP order by verb-particle combination by country (flat)

- PUT BACK, TAKE UP
- most extreme rates (intercepts)
- no difference by variety (slopes)

- we expect more extreme intercepts to have flatter slopes in percentage terms
- this is why we use log-odds and logistic regression and shouldn't even report differences in percent
- this goes beyond that, looks like no cross-variety difference at all
- except India



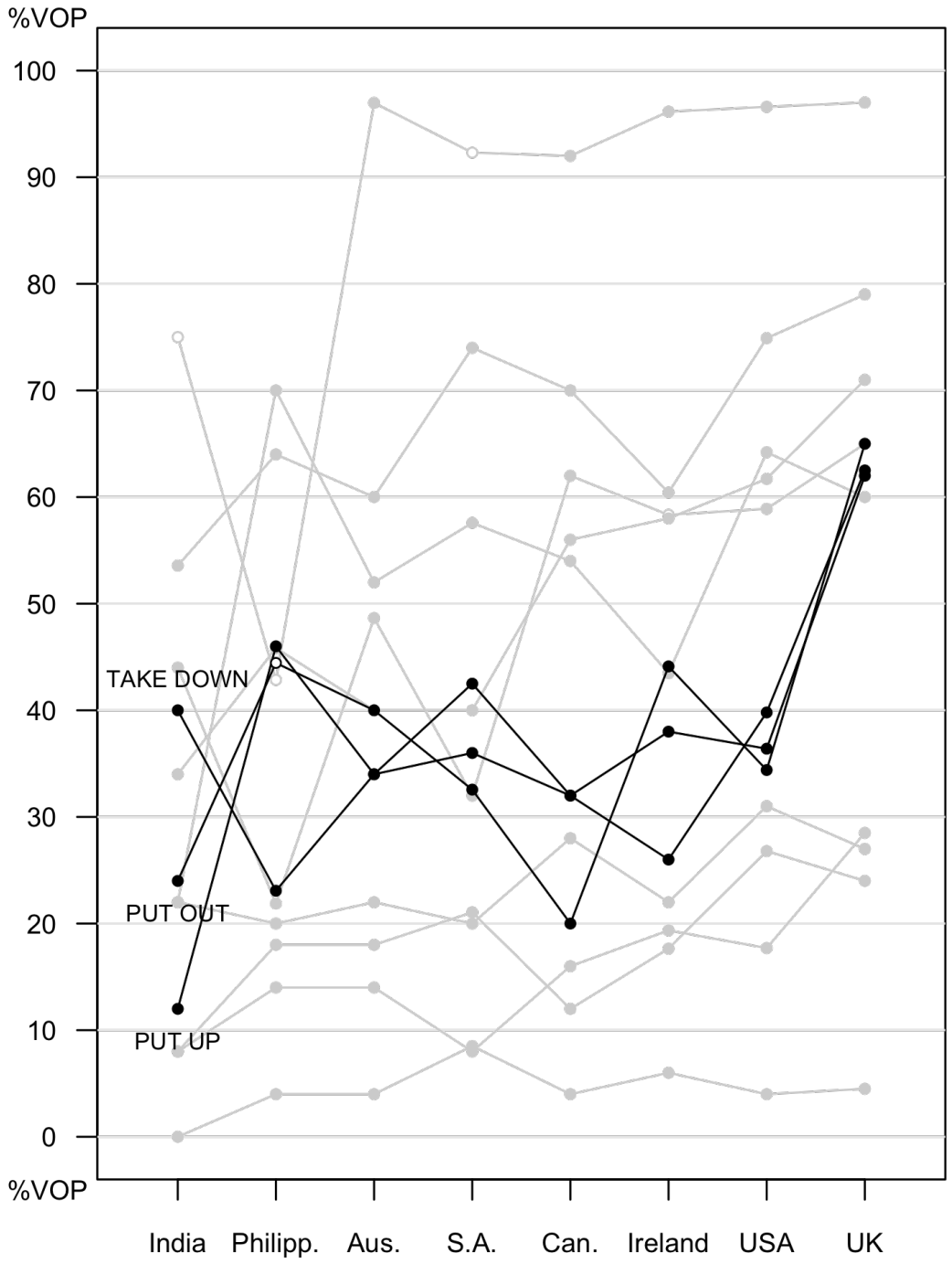
percentage of VOP order by verb-particle combination by country (UK way ahead)

- TAKE DOWN, PUT OUT, PUT UP
- a bit chaotic
- UK is more than 20% above USA
- presumably, TURN ON/OFF (the lights) in previous experiment belonged in this category

- is this real or epiphenomenal (different objects/environments)?

- TAKE DOWN … Christmas …
- UK: 60/97 = 61.9% VOP
- USA: 75/263 = 28.5% VOP

- real



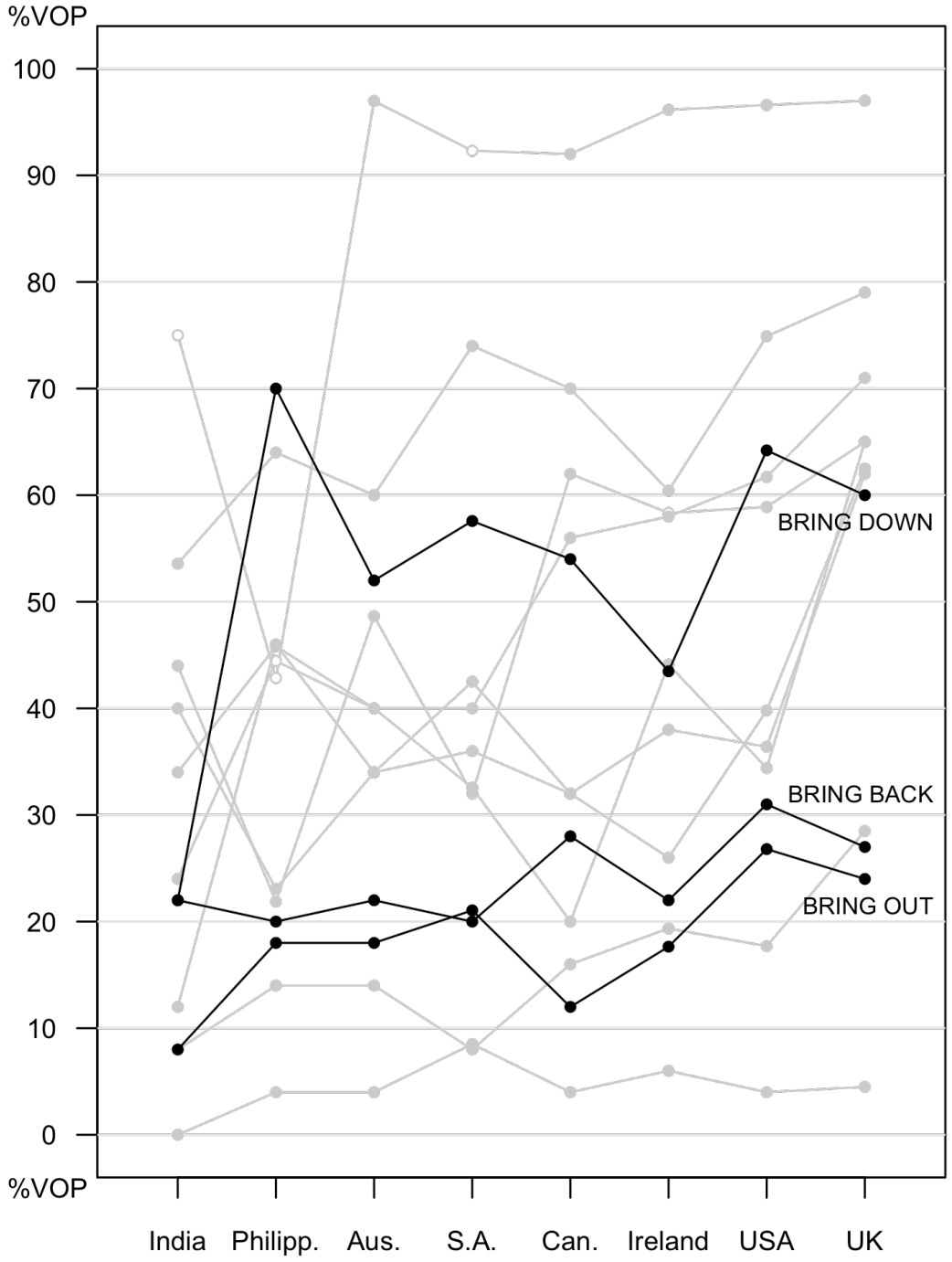
percentage of VOP order by verb-particle combination by country (USA ahead)

- BRING DOWN, BACK, OUT
- BRING UP (“normal” UK > USA)

- is this USA ahead or UK behind?
- looks like UK and Ireland behind

- no idea why (yet)

- now for more unexplained results…

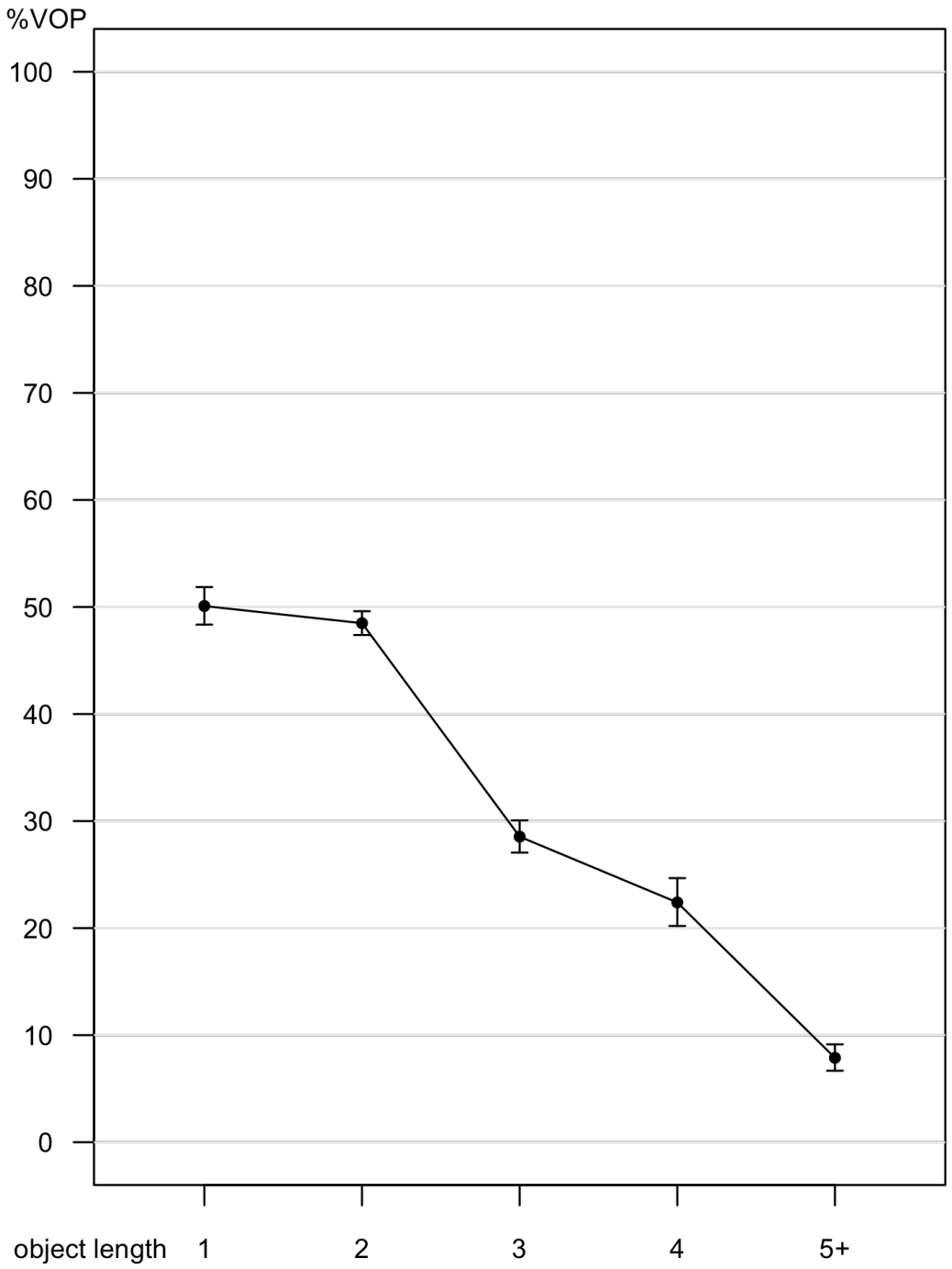


percentage of VOP order
average of eight countries
by object length

- as reported in previous work (Lohse et al. 2004)
- largest gap between 2-3 words

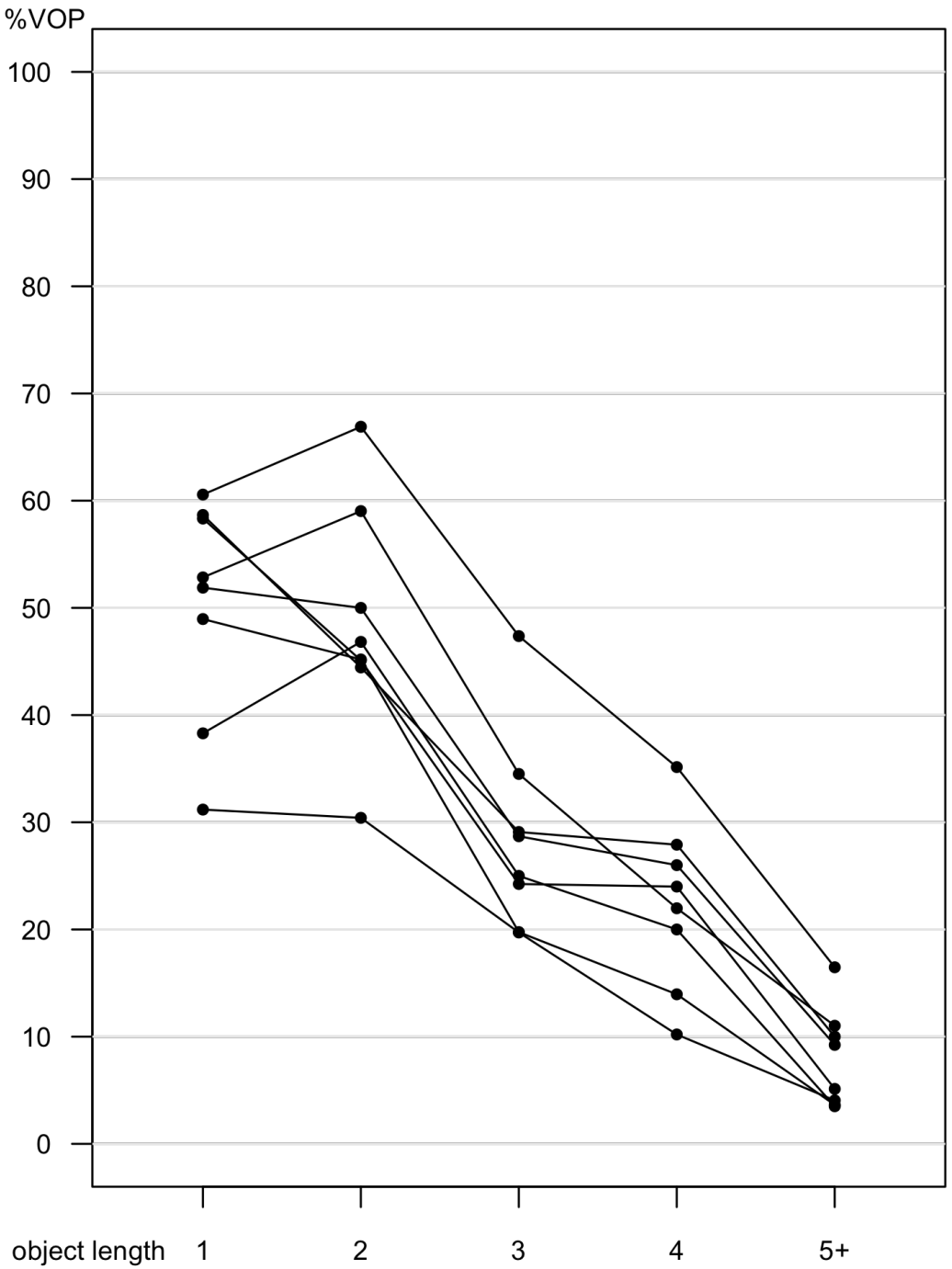
- even this pattern is not best modeled by a single linear predictor…

- *Jason's data showed a relatively higher rate of VOP for one-word objects, more of a smooth trend*



percentage of VOP order
by country
by object length

- the data speaks for itself



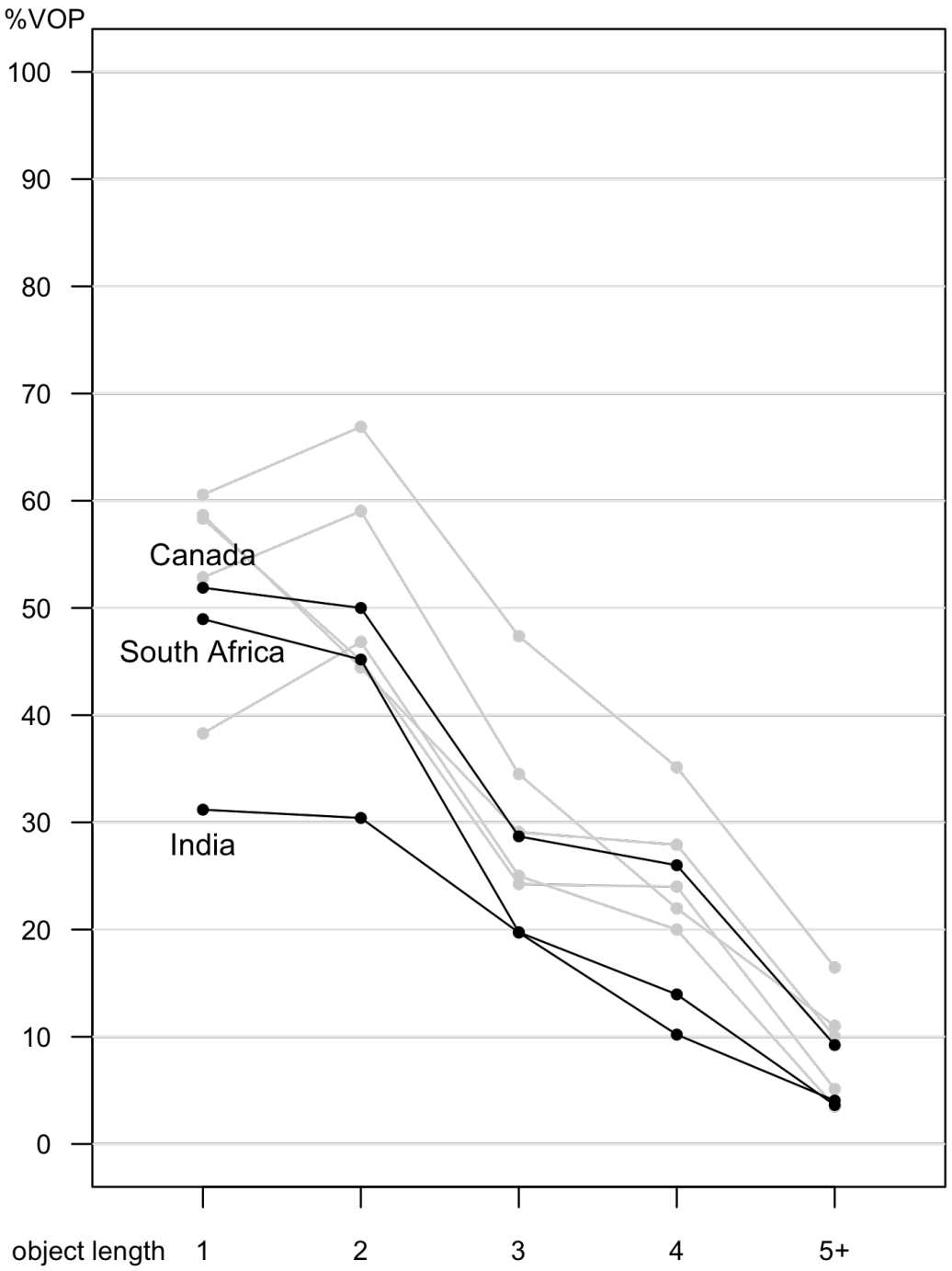
percentage of VOP order
by country
by object length (normal)

- moderate difference in rates
- similar pattern of constraint

- is India flatter (as elsewhere)?
- didn't test log-odds, maybe not
- *is India steeper? doesn't look it*

- maybe South Africa is steeper!

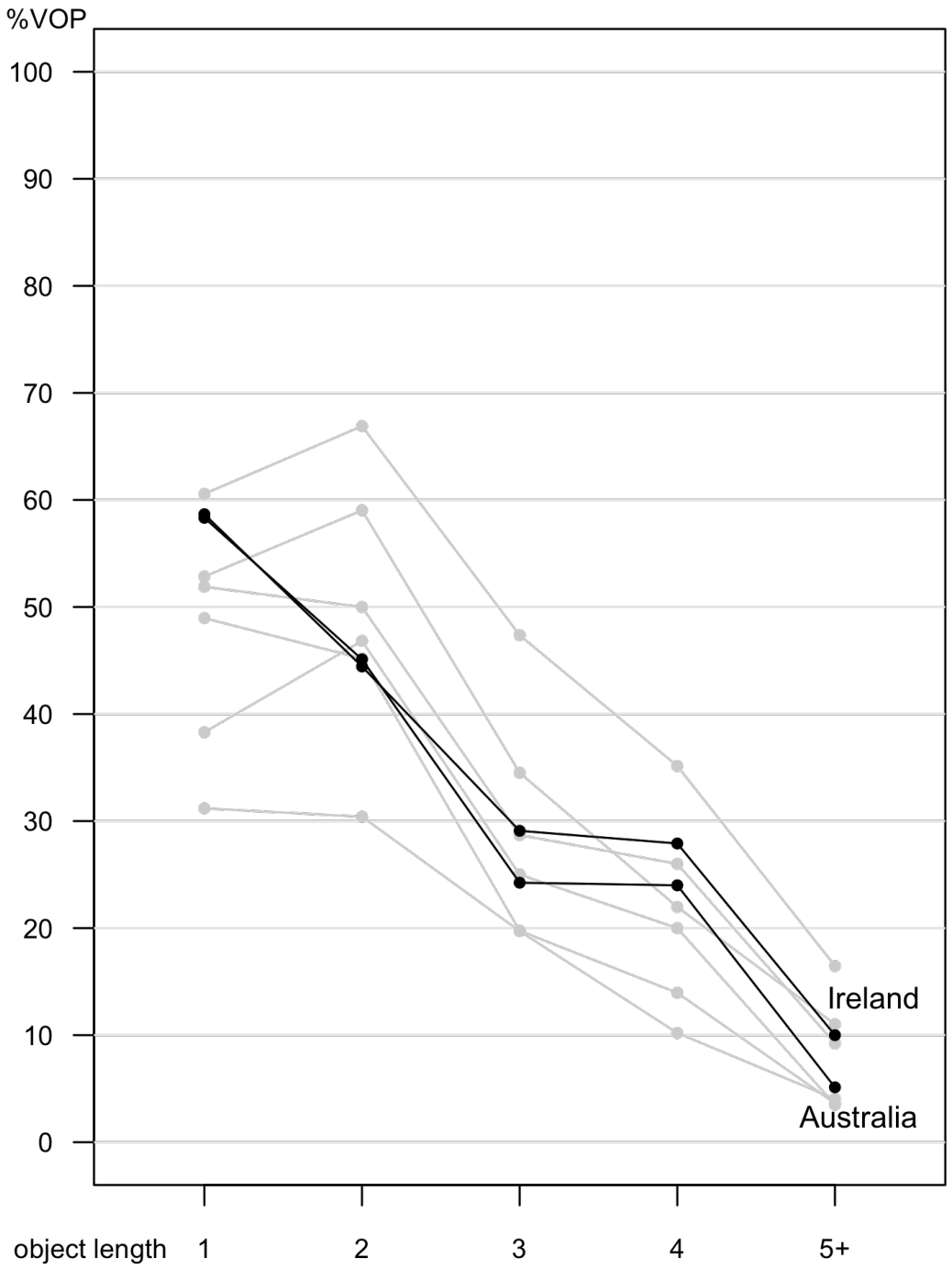
- need to adjust for other factors
- need to go beyond steep/flat?



percentage of VOP order
by country
by object length (ideal)

- while not the norm in this data, this pattern for Australia and Ireland is superficially most in line with a simple effect of end-weight

- why Australia/Ireland like this?
- or… why other countries aren't?



percentage of VOP order
by country
by object length (reversed)

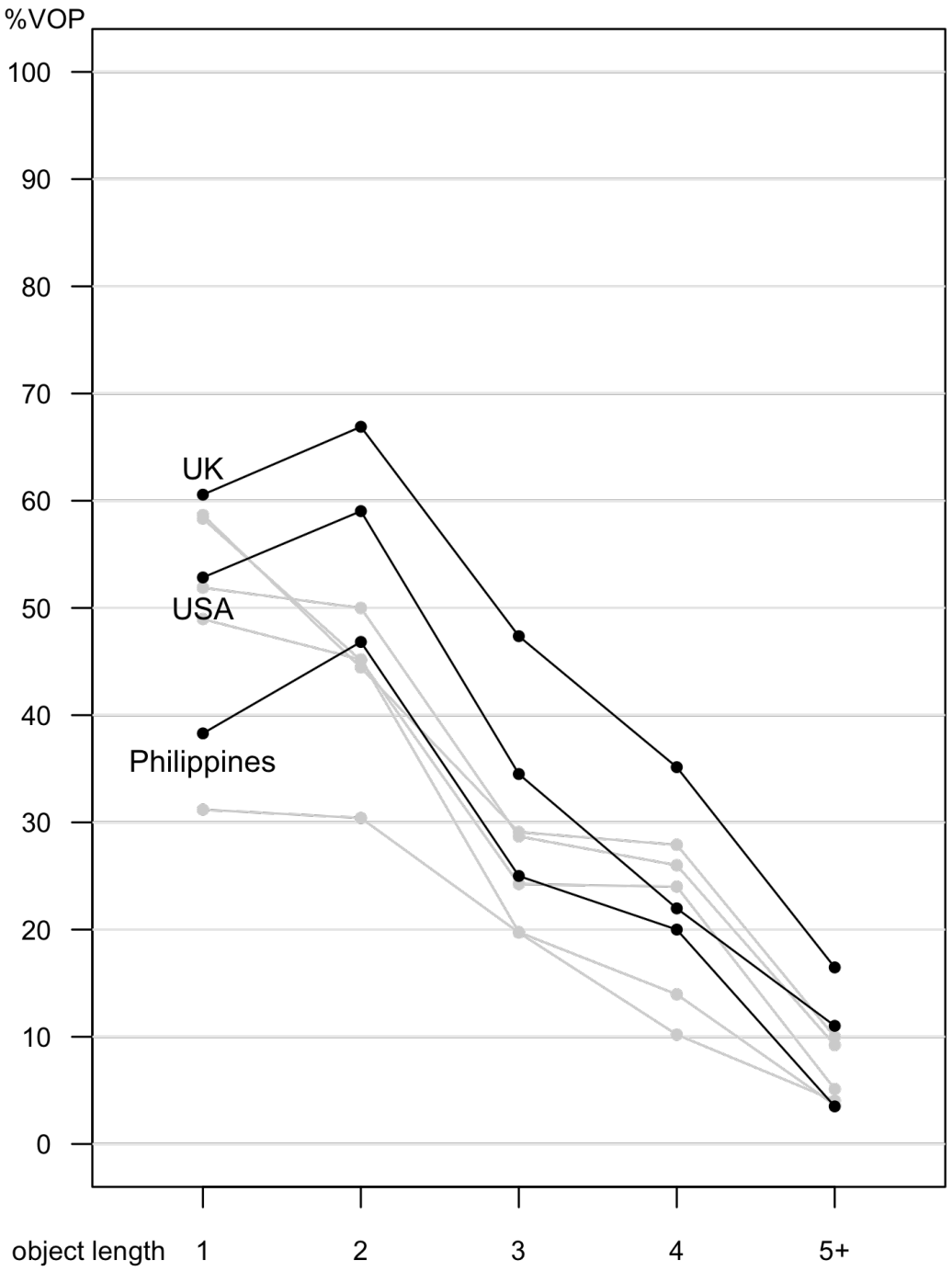
- UK, USA, Philippines together

- one-word objects are low
(not: two-word objects are high)
- not “end-weight” as understood

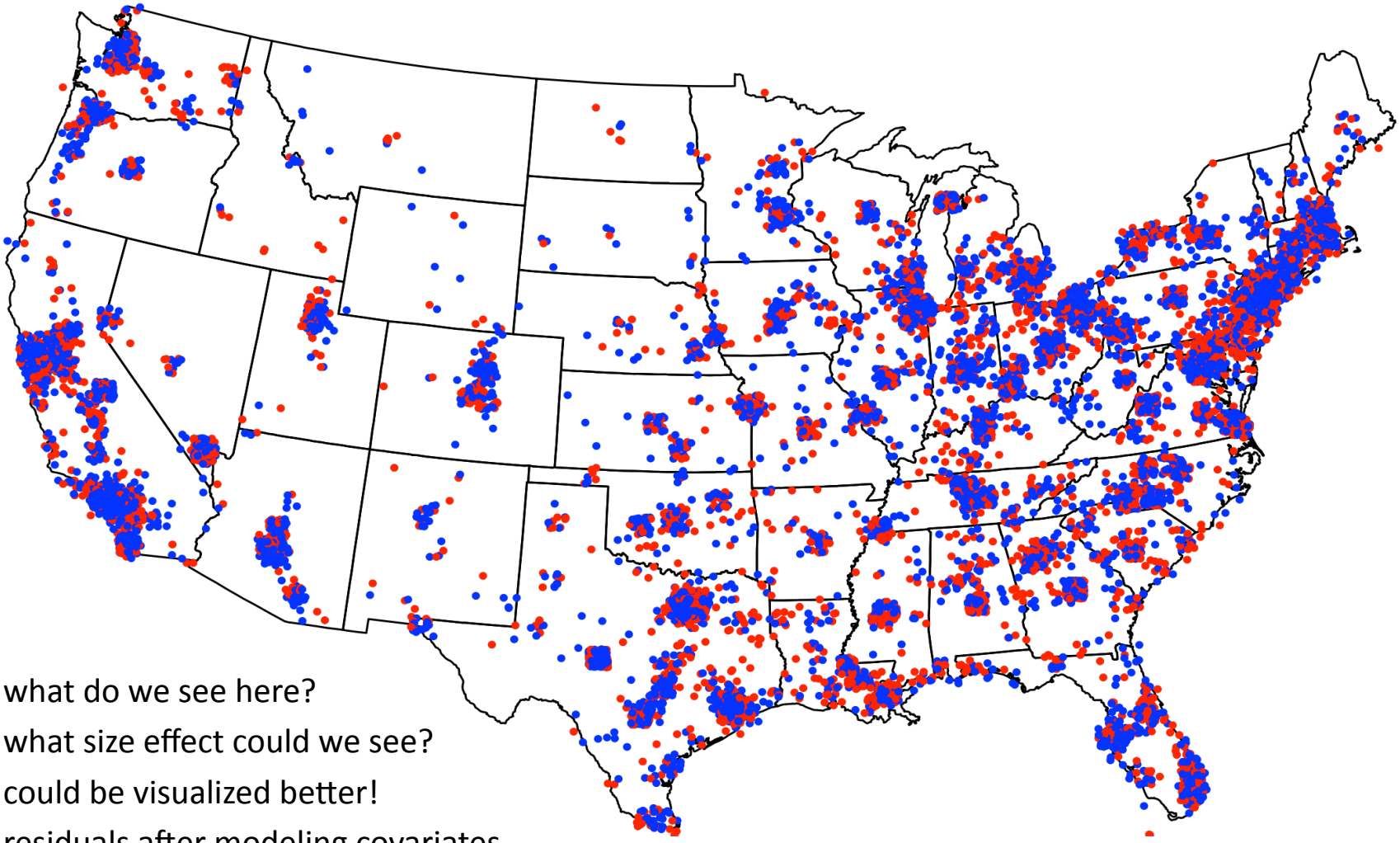
- has this pattern been reported?
- would it go away if other factors were accounted for?

- Australia, Ireland (VoP)
- Canada, South Africa, India
- UK, USA, Philippines (VPo)

- puzzling
- are the 1-word objects comparable across varieties?

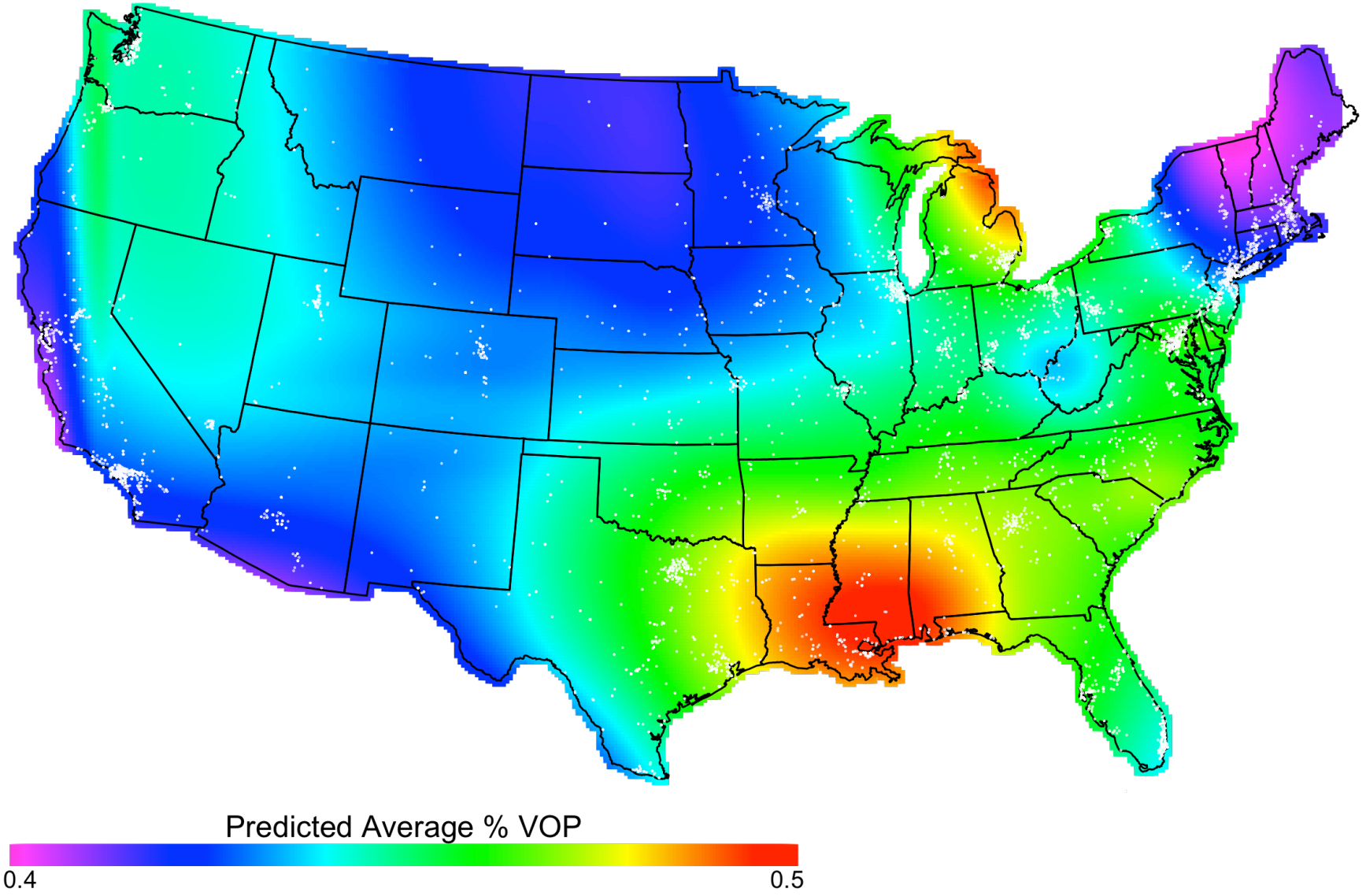


percentage of VOP order
United States (12,000 tokens)
jittered raw data (red = VOP)



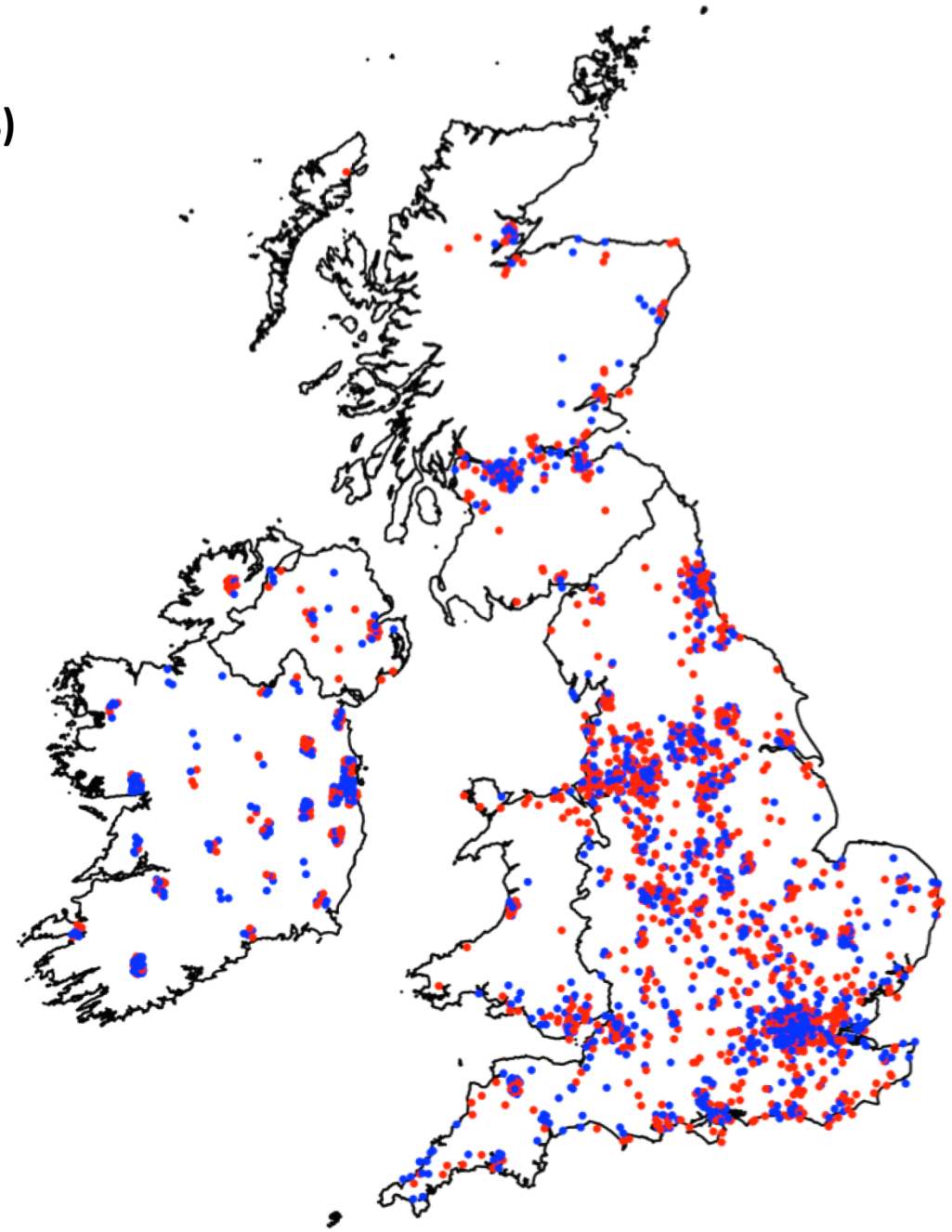
- what do we see here?
- what size effect could we see?
- could be visualized better!
- residuals after modeling covariates
- VP, O length, O type, PP vs. bare P (all very strong effects)

percentage of VOP order
United States (12,000 tokens)
GAM loess smooth (MapGAM)



percentage of VOP order
UK (2,400 tokens) and Ireland (458)
jittered raw data (red = VOP)

- more suggestive than USA map
- remember Trudgill



percentage of VOP order
UK (2,400 tokens) and Ireland (458)
GAM loess smooth (MapGAM)

- same color range but note the 2.5x wider range vs. USA map

- could not clip to data area for technical reasons, but illuminating
- what happens beyond data area?
- compare edges on USA map

- can the GAM smoothing method be salvaged, or is it better to use one of the other approaches?
- Jack Grieve's method: local spatial autocorrelation
- newer method: geographically weighted regression
- how different/better are these?

- particle variation is the best
- have only scratched the surface!