Semantic Change and Semantic Stability: Variation is Key

Claire L. Bowern
Yale University

ACL, Computational Approaches to Historical Language Change, August, 2019

Paper: https://www.aclweb.org/anthology/W19-4706

Today

- How evolutionary approaches to language change allow the modeling of cognate evolution (semantic change);
- Programmatic paper outlining some thoughts on computational/NLP approaches to semantic change;
- From the perspective of general questions of diachrony (quantitative, phylogenetics, etc, but not NLP-specific)
- Issues to consider to better realize potential of methods

REFS: Bowern (2018)

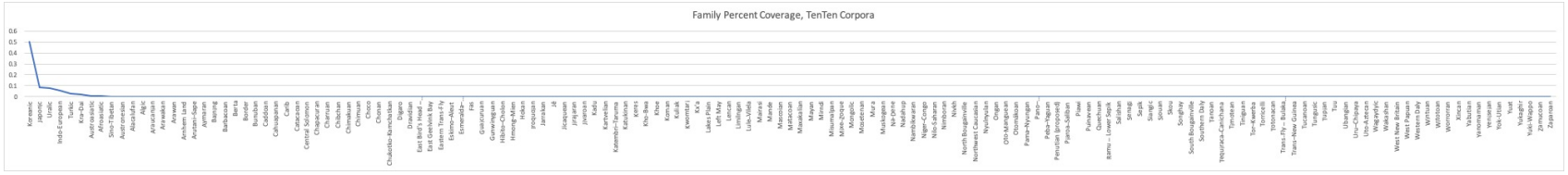
Table of Contents

Preliminaries

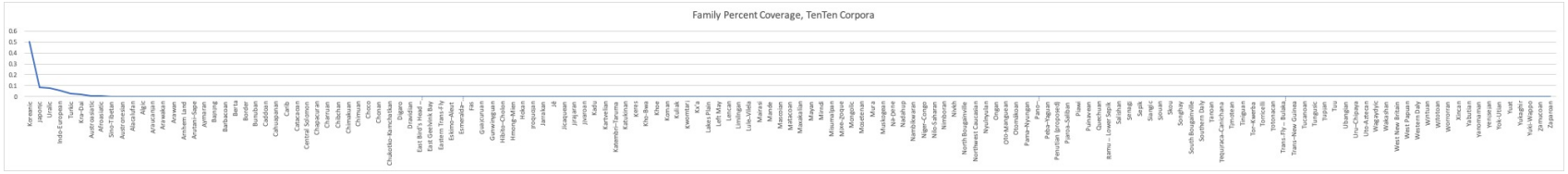
Preliminaries

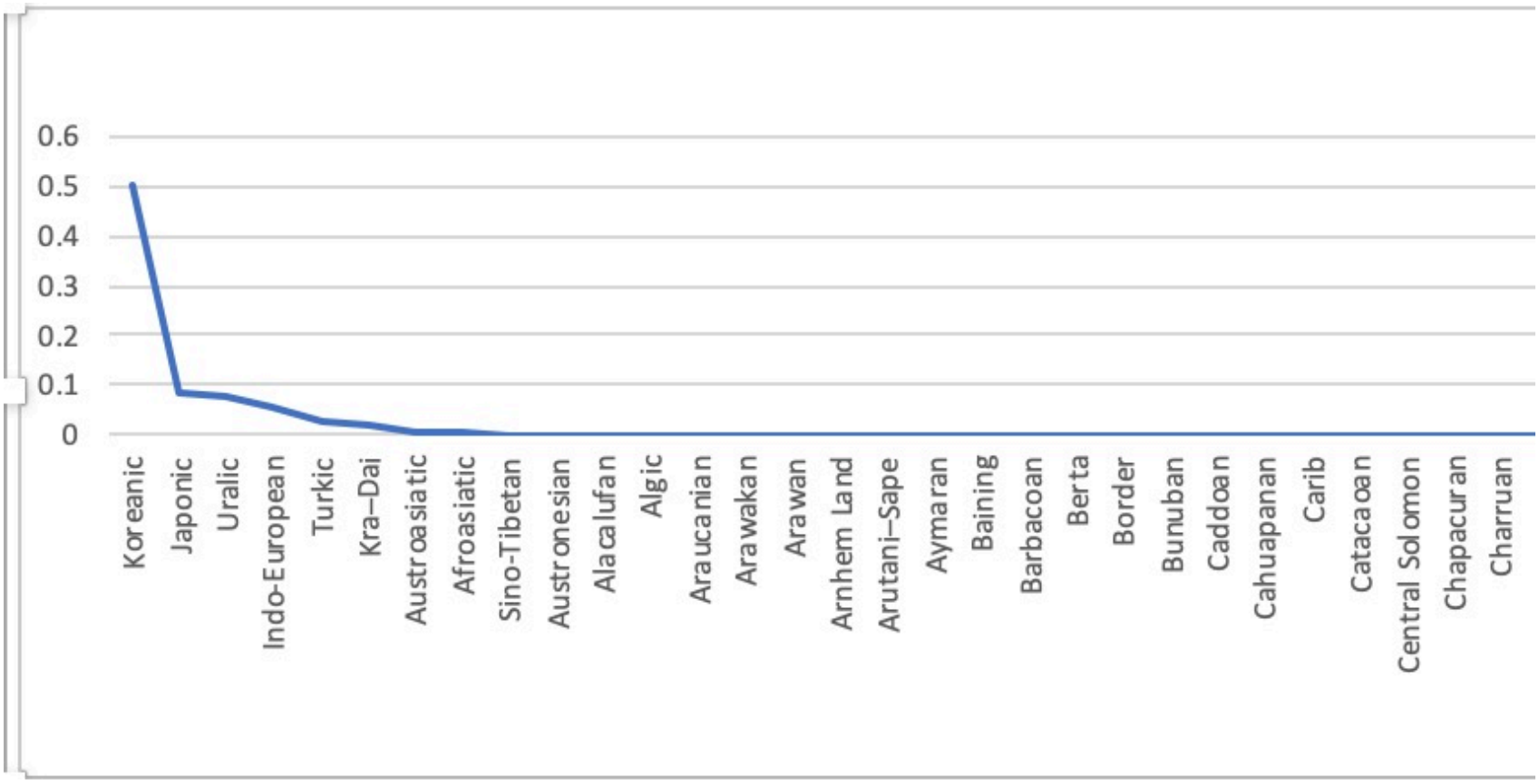
- All aspects of all languages are changing all the time.
- Most of this change is not recorded.
- The written record is sparse and shallow.

Digital Coverage



Digital Coverage





Does it matter?

Yes!!

- Languages (and families) differ in grammar;
- Cultures differ in patterns of social interaction, which shape change;
- Large industrialized languages are more homogeneous.

REFS: Bowern (2013)

Key questions

1. *What* forms have changed?
2. *How* does change work?
3. *Why* does it work the way it does?

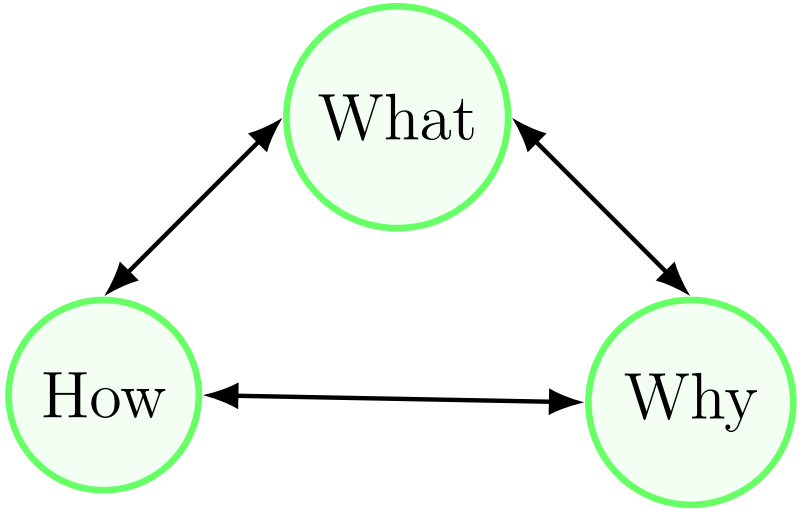


Figure 1: Key questions

REFS: Bouckaert et al. (2018); Greenhill et al. (2010); Hamilton et al. (2016b); Wedel et al. (2013)

- The *what* provides us with observations;
- The *why* provides us with a theory that explains those observations;
- The *how* provides us with a framework to structure those observations, and to predict and evaluate implications of the theory.

Change ≠ "difference"

Differences between two temporally different corpora can be caused by

- Author differences
- Dialect [West Saxon vs Northern in Anglo Saxon corpora]
- Genre
- Corpus size
- (Actual change)

How change occurs

The traditional view

- Innovation in a single language user;

How change occurs

The traditional view

- Innovation in a single language user;
- Innovation spreads through community;

How change occurs

The traditional view

- Innovation in a single language user;
- Innovation spreads through community;
- Innovations spread at different rates;

How change occurs

The traditional view

- Innovation in a single language user;
- Innovation spreads through community;
- Innovations spread at different rates;
- This forms dialects, and then languages;

How change occurs

The traditional view

- Innovation in a single language user;
- Innovation spreads through community;
- Innovations spread at different rates;
- This forms dialects, and then languages;
- (Innovations introduced through language contact)

REFS: Hock and Joseph (1996); Weinreich et al. (1968)

How change occurs

The generative view

- Innovations from children during acquisition, through errors (Paul, 1880) or induction (e.g. Lightfoot, 1991)
- Change in record as generations age

However...

- Types of errors kids make $\neq$ most common changes

However...

- Types of errors kids make $\neq$ most common changes
- Kids don't have the social networks to propagate change

However...

- Types of errors kids make $\neq$ most common changes
- Kids don't have the social networks to propagate change
- Changes proceed too fast
- Age-graded variation

However...

- Types of errors kids make $\neq$ most common changes
- Kids don't have the social networks to propagate change
- Changes proceed too fast
- Age-graded variation
- Can't reconcile with women as leaders of change in progress

REFS: Aitchison (2003); D'Arcy (2017); Hale (2007); Lightfoot (1991)

Time scales of 'change'

- Weeks: Eisenstein et al. (2014)

Time scales of 'change'

- Weeks: Eisenstein et al. (2014)
- Years: Jawahar and Seddah (today)

Time scales of 'change'

- Weeks: Eisenstein et al. (2014)
- Years: Jawahar and Seddah (today)
- Decades: Hamilton et al. (2016b); Yao et al. (2018)

Time scales of 'change'

- Weeks: Eisenstein et al. (2014)
- Years: Jawahar and Seddah (today)
- Decades: Hamilton et al. (2016b); Yao et al. (2018)
- Centuries: Zimmermann (today), Luo et al (today)

Time scales of 'change'

- Weeks: Eisenstein et al. (2014)
- Years: Jawahar and Seddah (today)
- Decades: Hamilton et al. (2016b); Yao et al. (2018)
- Centuries: Zimmermann (today), Luo et al (today)
- Millennia: Greenhill et al. (2010); Haynie and Bowern (2016) (etc)

Do these processes scale up/down?

How change occurs: Evolutionary models

How change occurs: Evolutionary models

Language is a complex evolutionary system

- Darwinian system:
  1. Variation
  2. Selection
  3. Transmission
- Unit of study: "population"

REFS: Bowern (2018); Marlowe (2005); Mesoudi (2011); Wedel (2006)

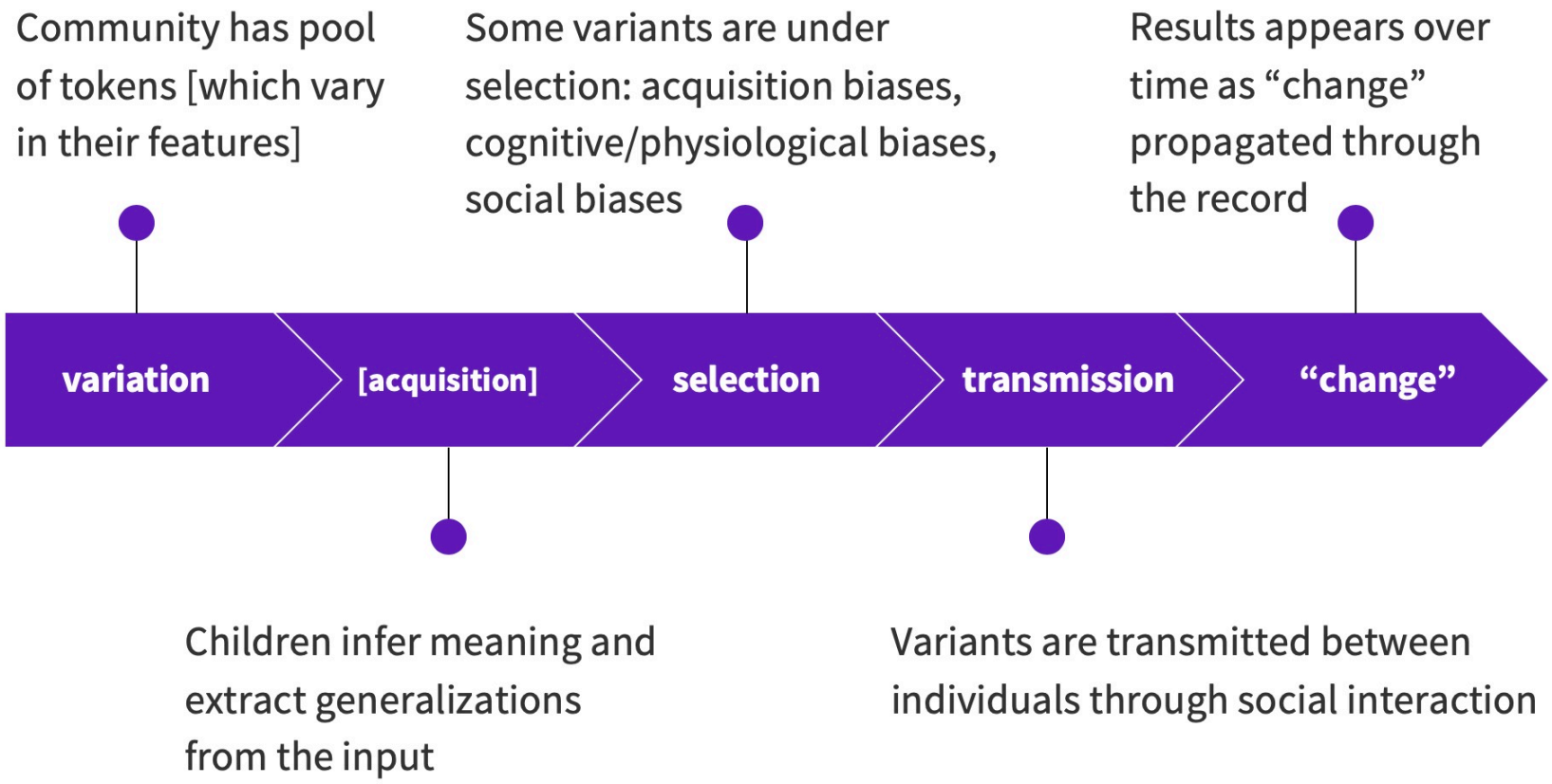


Figure 2: Schematic representation of language change in an evolutionary framework

Variation

- Variation is physiological, psychological, and social (age, gender, etc)
- Most work with embeddings treats language as monolithic (an invariant set of rules)
- But there's variation, and variation matters for studying change (cf Jawahar and Seddah (today)

Evolutionary models

Selection

- Not all variants have equal chances of spreading within a community.

Evolutionary models

Selection

- Not all variants have equal chances of spreading within a community.
- positive or negative selection, or neutral

Evolutionary models

Selection

- Not all variants have equal chances of spreading within a community.
- positive or negative selection, or neutral
- Selection can be modeled as a set of biases which inhibit or facilitate transmission. Such biases include acquisition, cognitive/physiological biases, and social biases.

Aside:

"Human language changes over time, driven by the dual needs of adapting to ongoing sociocultural and technological development in the world and facilitating efficient communication." (Intro to this workshop)

Aside:

"Human language changes over time, driven by the dual needs of adapting to ongoing sociocultural and technological development in the world and facilitating efficient communication." (Intro to this workshop)

Evolutionary view:

- Just one of several positive and negative selection biases;
- No drive for efficiency in communication

Consequences for studying change

- Not generalizing across a system
- Looking within a system for difference
- Not discrete and uniform entities at each time point.
- Must be able to distinguish between relative shifts in frequency of use among subsenses, and the spread of genuine innovations.

Transmission

Transmission

- Not just intergenerational.
- Transmission from any other language user.
- poorly formalized!
- Difficult to reconcile with models that take slices at time points to study change in progress.

Taking an evolutionary view of language change does *not* entail that it be studied with direct and concrete analogues to biological replication and speciation.

Summary

- Provides a framework against which we can test hypotheses for the how, what, and why of change,
- beyond the identification of differences

Table of Contents

The Lexicon

The Lexicon as Mappings

- set of mappings between forms, meanings, and the world.

CAT ≠ /kæt/ ≠ 

Like other parts of language, the lexicon is also constantly changing.

REFS: Traugott and Dasher (2002); Urban (2014)

Words vs concepts

As Bender (2019) has noted, because of the heavy emphasis on English, the distinction between 'words' and 'concepts' is sometimes obscured in NLP (because of focus on English).

REFS: Bender (2019)

Lexical Replacement

1. Semantic change: that is, change in mappings between a lexical item, concepts, and world
2. Borrowing from other languages;
3. Creation of words *de novo*;
4. (Loss)

Types of semantic change

- bleaching, amelioration, meronymic, extension

Types of semantic change

- bleaching, amelioration, meronymic, extension
- Changes in connotation [e.g. presuppositional content]
- Changes in denotation

Detection of types of change

eg: Hamilton et al. (2016a)

- Nouns are more likely to undergo irregular cultural shifts (e.g. expansion due to technological innovations);
- Verbs are more likely to show regular processes of change, such as drift

Issue: Typology matters for the lexicon

BUT:

- verb numbers differ across languages (2–open class)
- this affects functional load, polysemy, and lexicalization patterns
- and these are the very factors that Hamilton et al showed were important for assessing the likelihood of change.
- Results tested with English, German, and French

Relative frequency of types of changes

Technological innovation

- How much semantic change is driven by technological innovation?
- Salient to tech professionals, but small part of semantic change overall, compared to euphemism, metaphorical extension, and bleaching.

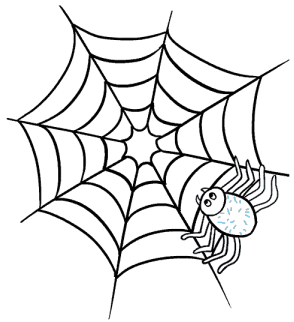
Modeling semantic change in an evolutionary framework

- variation
- selection
- transmission

Variation:

Spiderweb vs Cobweb

- The two words are synonymous;
- Spiderwebs are spiral or wheel-shaped, cobwebs are collapsed;
- Spiderwebs have spiders in them;
- Spiderwebs have spiders, while cobwebs = dust bunnies;



Speakers are unaware of these distinctions, and the variants do not pattern by age, gender, etc.
Such variation is not under selection and is below the level of consciousness. It's hard to detect as researchers don't know to look for it.

Selection skewing change

- bias against novel meanings?

Selection skewing change

- bias against novel meanings?
- bias towards lexicalization of salience

Selection skewing change

- bias against novel meanings?
- bias towards lexicalization of salience
- psychological bias towards discounting interlocutors (Ahern and Clark, 2017)

Selection skewing change

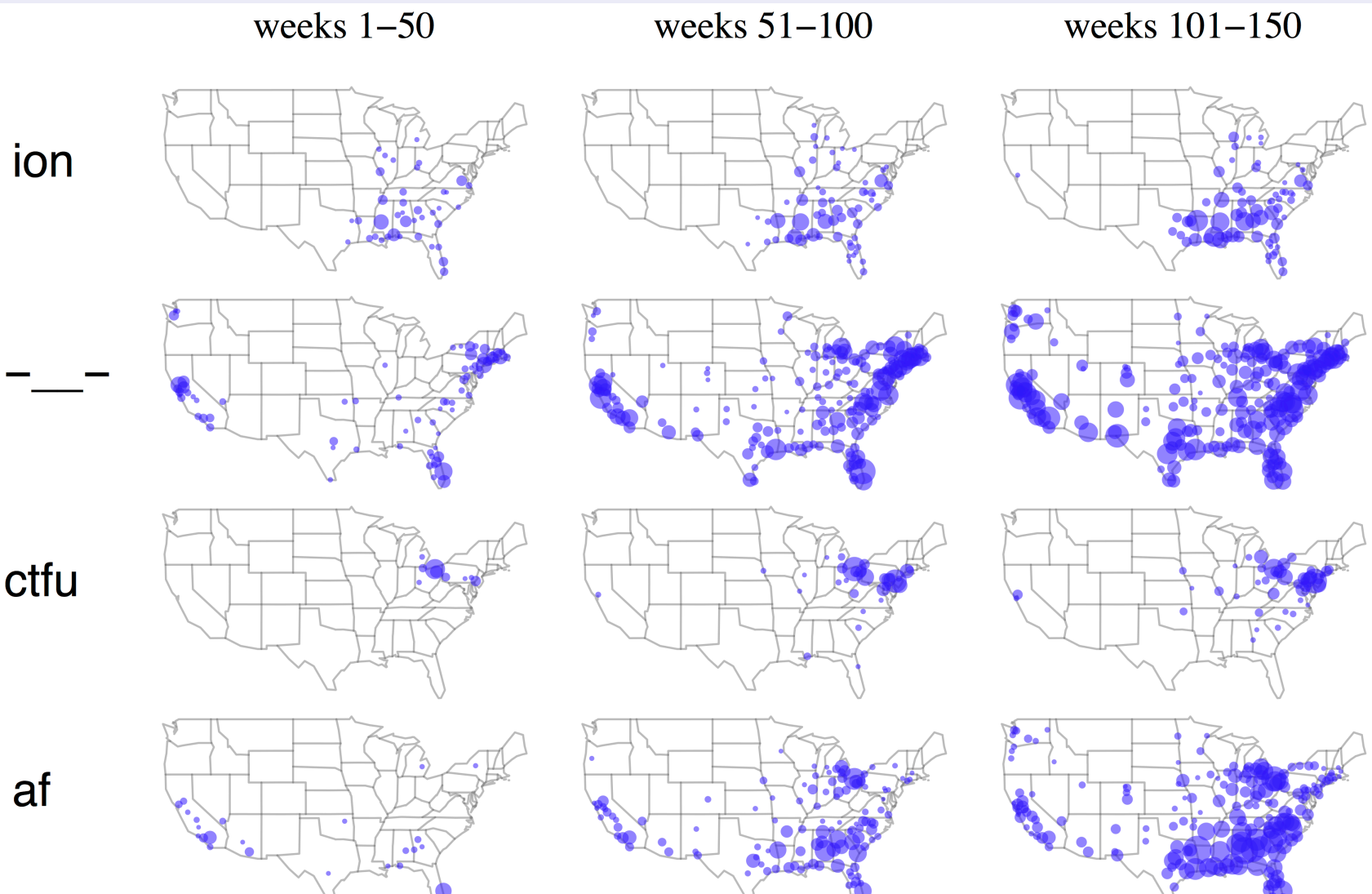
- bias against novel meanings?
- bias towards lexicalization of salience
- psychological bias towards discounting interlocutors (Ahern and Clark, 2017)

Much need for formalizing, examining further how cognitive biases influence semantic change

REFS: Ahern and Clark (2017)

Transmission

Eisenstein et al. (2014)



Selection and Transmission

From the study of changes we can make inferences about what speakers are doing, how they learn

But they are easiest to detect when tracking novel items (no reason why change in existing words should work the same way)

These processes also differ crosslinguistically

Color in Pama-Nyungan and Indo-European

Sources of Color Terms in Indo-European

- Other color terms
- Other visual terms (e.g. 'bright')
- Extensions from the natural world (e.g. orange, purple)
- modifiers (e.g. light, dark)

Sources of Color Terms in Pama-Nyungan

- Extensions from the natural world (e.g. 'ashes')
- Other visual terms (e.g. 'bright')
- no modifiers
- no other color terms

REFS: Haynie and Bowern (2016)

Summary

- Meanings vary, have positive or negative selectional biases, and are transmitted through language use.
- This varies crosslinguistically

Aside: Hamilton et al. (2016b)

(Further example of the how, why, and what for transmission and change)

Polysemy and Frequency

- Studied English, German, French, Chinese
- Low frequency words more likely to change
- High polysemy words more likely to change

Aside: Hamilton et al. (2016b)

Why

- Low frequency → speakers have less information about meaning :: words more vulnerable to reinterpretation or replacement (further eroding their frequency).
- High polysemy leads to change? perhaps because they are both more ambiguous and more likely to be further extended.
- BUT polysemy and frequency are corelated (Pagel and Calude 2016)
- cf. Dubossarsky (today): noise is important

Table of Contents

Implications for embeddings

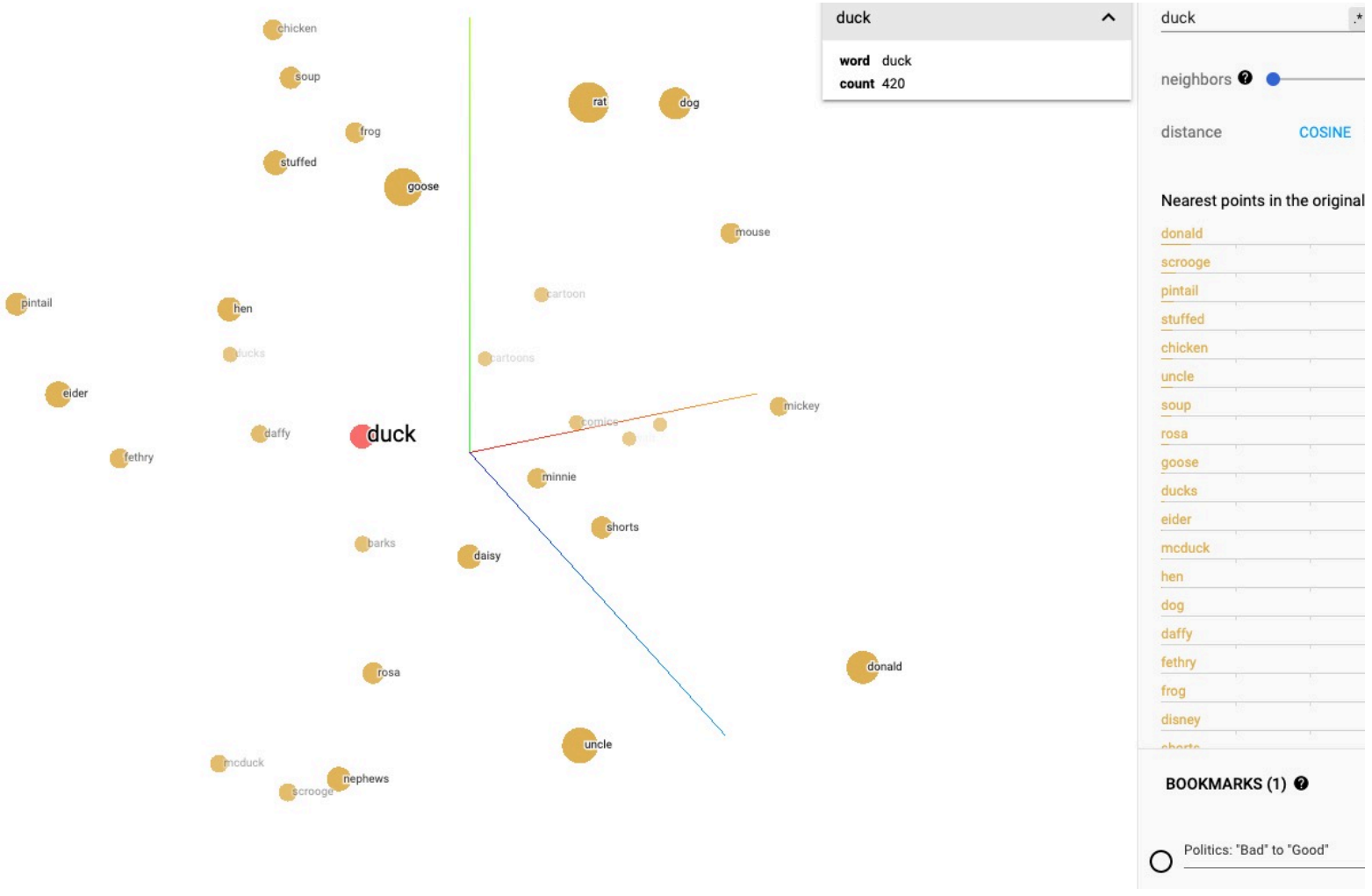
Implications for embeddings

word embeddings

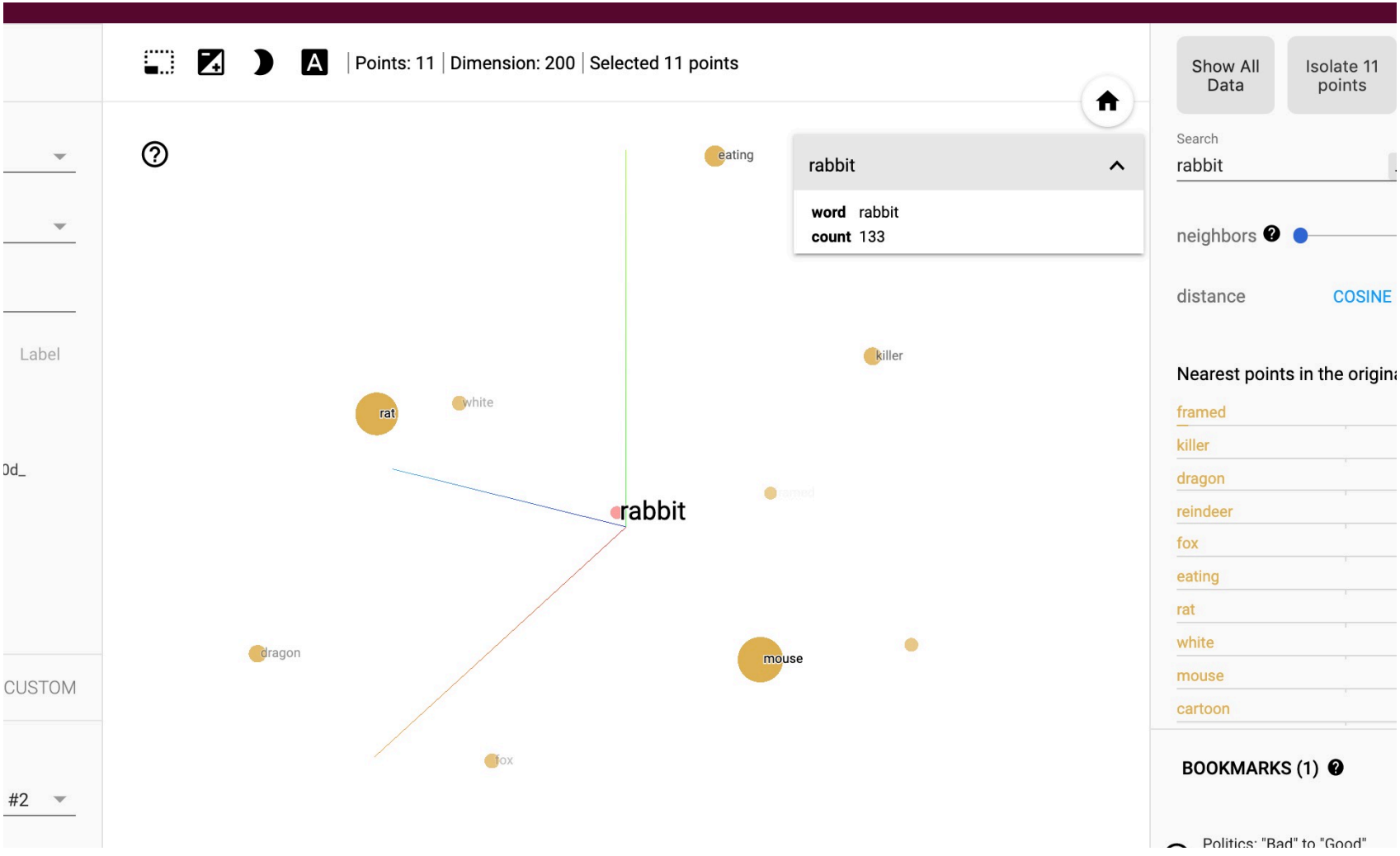
- "you can know a word by the company it keeps" (Firth, 1957, 11)
- (cf Dubossarsky (today))

REFS: Kulkarni et al. (2015); Turney and Pantel (2010)

But can we?


duck
word duck
count 420
duck
neighbors
distance COSINE
Nearest points in the original
donald
scrooge
pintail
stuffed
chicken
uncle
soup
rosa
goose
ducks
eider
mcduck
hen
dog
daffy
fethry
frog
disney
BOOKMARKS (1)
Politics: "Bad" to "Good"
chicken
soup
frog
stuffed
goose
rat
dog
mouse
cartoon
cartoons
pintail
hen
ducks
eider
daffy
duck
fethry
comics
mickey
minnie
shorts
barks
daisy
rosa
donald
uncle
mcduck
nephews
scrooge

But can we?


Points: 11 | Dimension: 200 | Selected 11 points
Show All Data
Isolate 11 points
Search
rabbit
neighbors
distance
COSINE
Label
CUSTOM
#2
eating
rabbit
word rabbit
count 133
killer
rat
white
rabbit
dragon
mouse
fox
Nearest points in the origina
framed
killer
dragon
reindeer
fox
eating
rat
white
mouse
cartoon
BOOKMARKS (1)

Critiques

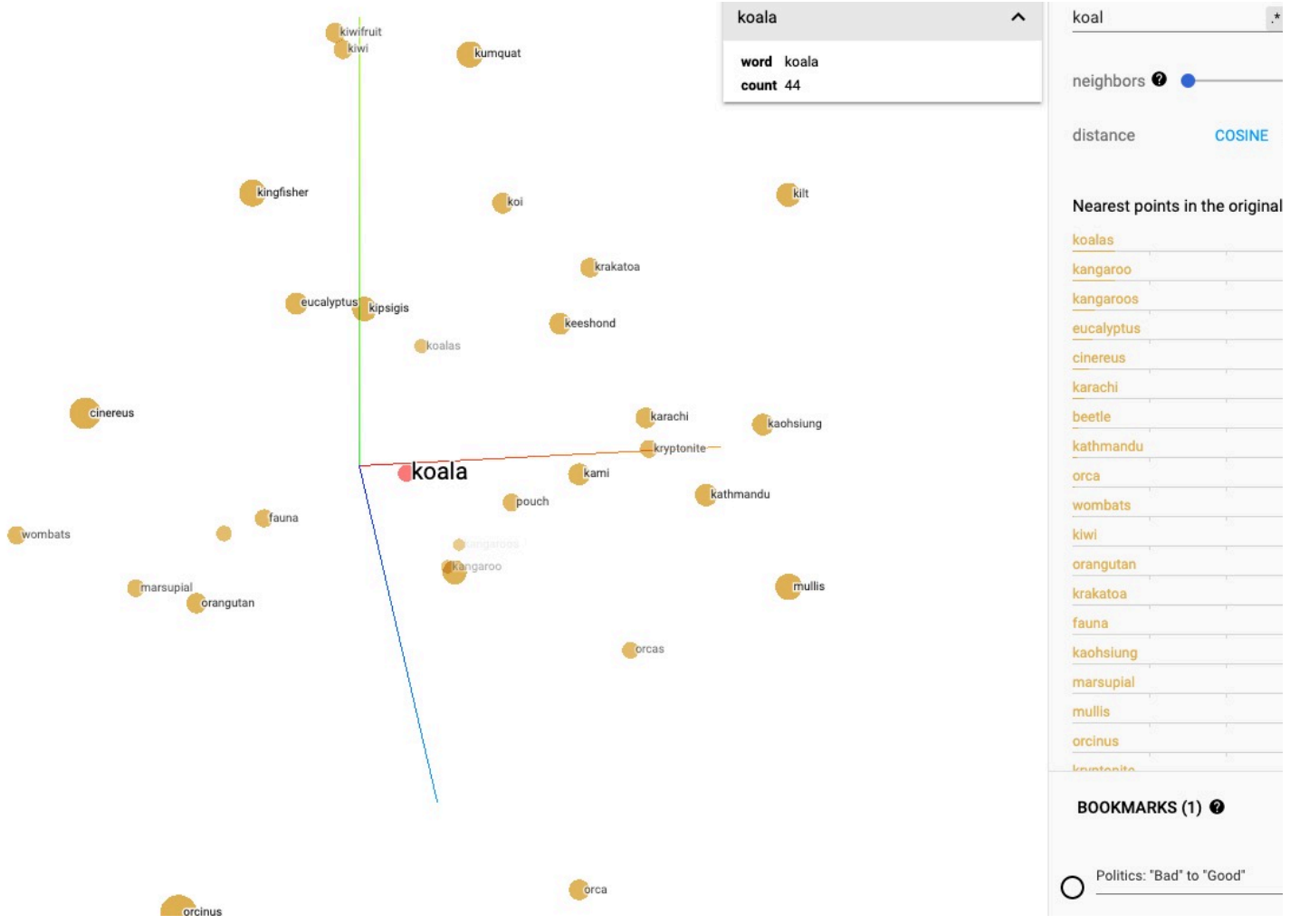
- Need large corpora
- Results are brittle
- Can't study senses independently → overlooking the variation that is the input to change

REFS: Dubossarsky et al. (2017); Tahmasebi et al. (2018) cf. also Zimmermann (today)

- Embeddings across massive corpora assume that all speakers have the same knowledge of the vocabulary of their language.
- conflates linguistic knowledge with real-world knowledge (e.g. by using Wikipedia)

For example:

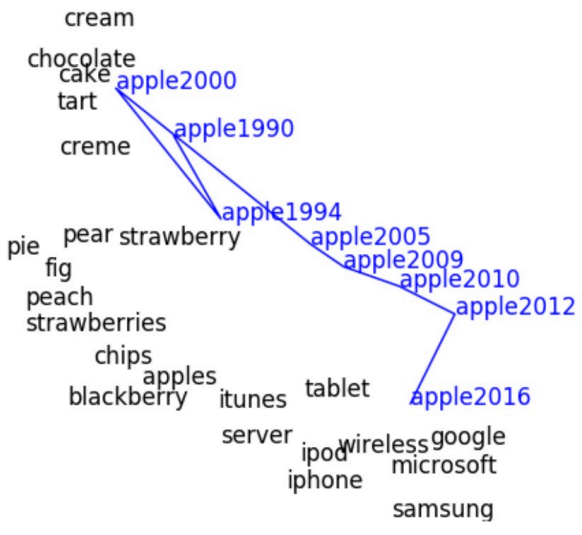
- Is membership of the genus Phascolarctos part of the meaning of 'koala'?
- (corpora skewed towards such information)



projector.tensorflow.org

Change in frequency ≠ change in meaning

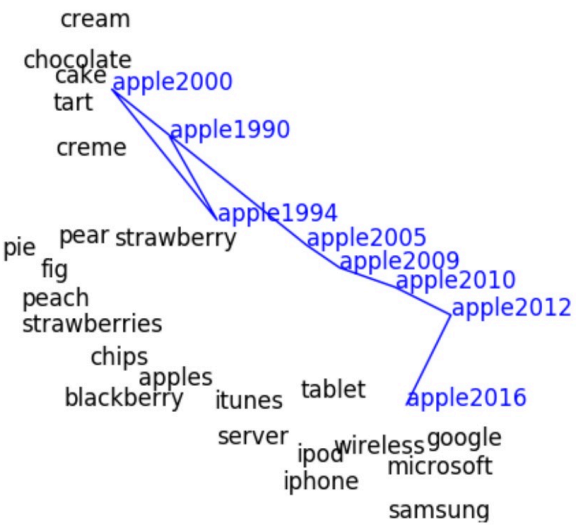
Change in frequency



(a) apple

Change in frequency ≠ change in meaning

Change in frequency



(a) apple

Diagnostic of change?

- Yao et al. (2018):


→


- Change in frequency is precursor to change, not the change itself
- Frequency is very genre-dependent
- Variation leads to change, but not all variation is change.

REFS: Kulkarni et al. (2015); Kutuzov et al. (2018); Yao et al. (2018)

Lexical replacement and phylogenetics

Lexical replacement and phylogenetics

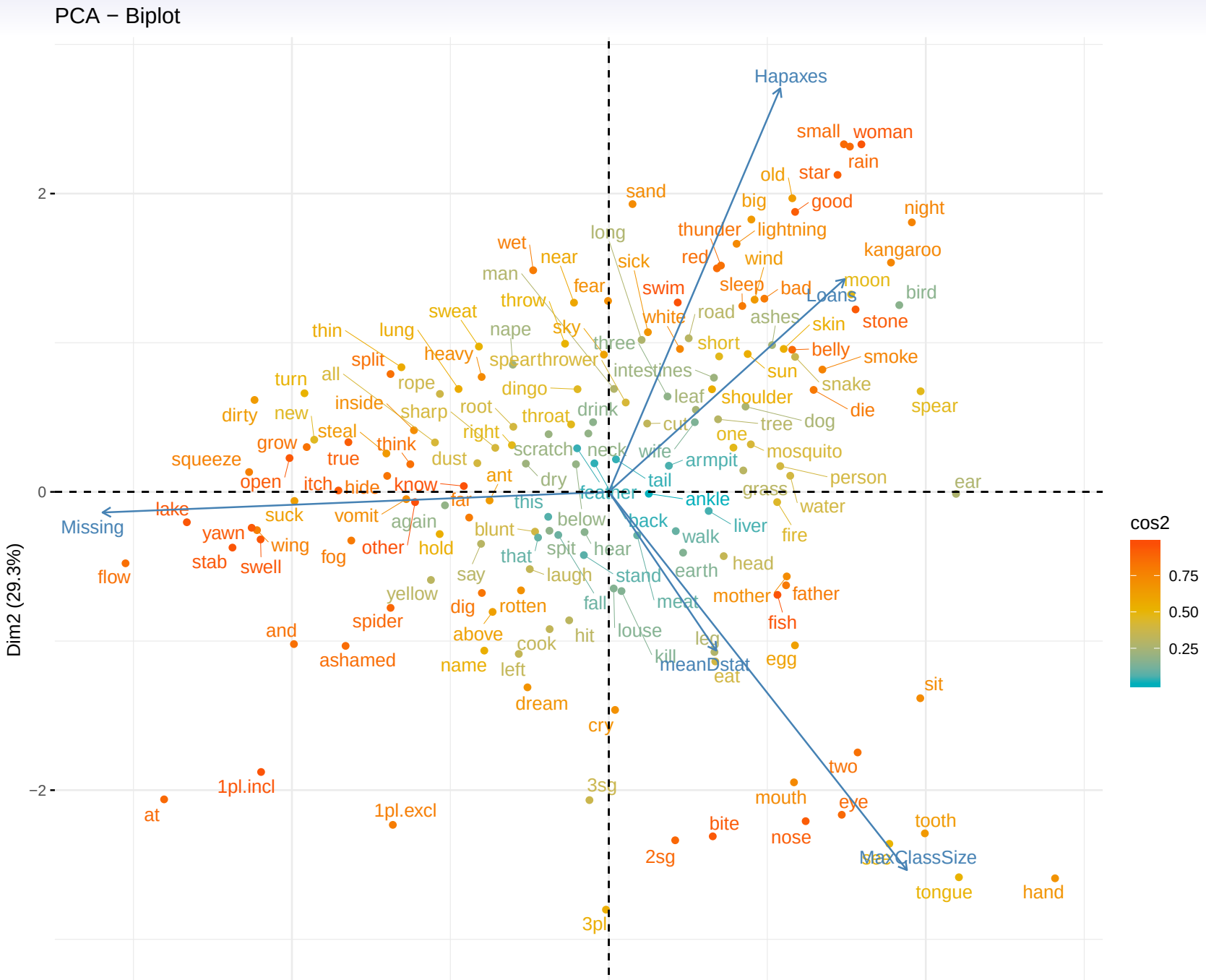
legacy problem

- Phylogenetics models cognate evolution through lexical replacement
- Swadesh lists: sample of convenience
- Little reason to assume these are the most appropriate words for modeling deep relations
- Can we do better? (test for Pama-Nyungan, other families in progress; coding botteneck for other terms)

REFS: McMahon and McMahon (2006); Swadesh (1952, 1955); **?**

PCA

- Max class size (stability)
- Number of cognate classes (proxy for replacement rate)
- D statistic (measure of conformity to treelike evolution)
- Amount of missing data
- Number of loans


PCA − Biplot
Dim2 (29.3%)
2
0
−2
Hapaxes
Loans
meanDstat
MaxClassSize
Missing
cos2
0.75
0.50
0.25
small
woman
rain
old
star
sand
big
good
night
thunder
lightning
kangaroo
wet
near
long
sick
red
wind
man
fear
swim
sleep
bad
moon
bird
throw
white
road
ashes
skin
stone
sweat
thin
lung
nape
sky
three
short
belly
smoke
split
heavy
spearthrower
intestines
sun
snake
turn
all
rope
dingo
leaf
shoulder
die
spear
dirty
new
inside
sharp
root
throat
drink
cut
tree
dog
steal
right
one
grow
think
scratch
neck
wife
armpit
mosquito
squeeze
true
dust
dry
tail
person
open
itch
hide
know
ant
feather
grass
ear
lake
suck
vomit
far
this
ankle
water
again
below
back
liver
Missing
yawn
blunt
walk
fire
wing
other
hold
that
spit
hear
stab
fog
swell
say
laugh
stand
earth
head
flow
yellow
dig
rotten
fall
meat
mother
father
spider
above
hit
louse
leg
fish
and
cook
kill
egg
ashamed
name
left
eat
sit
dream
cry
two
1pl.incl
3sg
mouth
eye
at
1pl.excl
bite
tooth
nose
2sg
tongue
hand
3pl

Choosing stable words

- Massive bottleneck for investigation: manual coding

Choosing stable words

- Massive bottleneck for investigation: manual coding

Could add:

- Most stable vectors across diachronic corpora
- Lowest error translations from parallel corpora

Table of Contents

Conclusions

- Word embeddings obscure critical data on change
- At present, they look at a very small slice of language
- This matters for accuracy and coverage

Conclusions

- Word embeddings obscure critical data on change
- At present, they look at a very small slice of language
- This matters for accuracy and coverage
- This is in part because of a lack of clarity about models of change.
- Evolutionary models of change bring that clarity and allow us to better realize the potential for these methods.
- Evolutionary models are applicable to semantic change
- It would be fruitful to further investigate how we can study stability and innovation through these models.

References

Christopher Ahern and Robin Clark. 2017. Conflict, cheap talk, and Jespersen's cycle. *Semantics and Pragmatics*, 10.

Jean Aitchison. 2003. Psycholinguistic perspectives on language change. In *The Handbook of Historical Linguistics*, chapter 25, pages 736–743. Wiley Blackwell.

Emily M. Bender. 2019. English isn't general for language, despite what NLP papers might lead you to believe. Sumposium on Data Science and Statistics, Bellevue, WA.

Remco R Bouckaert, Claire Bowern, and Quentin D Atkinson. 2018. The origin and expansion of Pama–Nyungan languages across Australia. *Nature ecology & evolution*, 2:741–749.

Claire Bowern. 2013. Relatedness as a factor in language contact. *Journal of Language Contact*, 6(2):411–432.

Claire Bowern. 2018. Computational phylogenetics. *Annual Review of Linguistics*, 4:281–296.

Alexandra D'Arcy. 2017. *Discourse-pragmatic variation in context: Eight hundred years of like*. John Benjamins Publishing Company.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2017.

Thanks

- Funding: NSF grant BCS-1423711
- Workshop organizers
- Workshop presenters
- Yale Histling lab