Vector-space semantic maps

A data-driven approach to the study of syntactic productivity in diachrony

Florent Perek
Princeton University
fperek@princeton.edu

Syntactic productivity

- Property of a construction to attract new lexical fillers
- In diachrony:
  - The distribution of constructions may vary over time
  - e.g., verb slot in the *way*-construction (Israel 1996)
    - Verbs of physical actions attested from the 16th century

      They hacked their way through the jungle.
    - Abstract means of reaching a goal only appear in the 19th century

      She typed her way to a promotion.

- Appears to be partly arbitrary, but actually tied to prior usage

Usage-based determinants of syntactic productivity

Type frequency

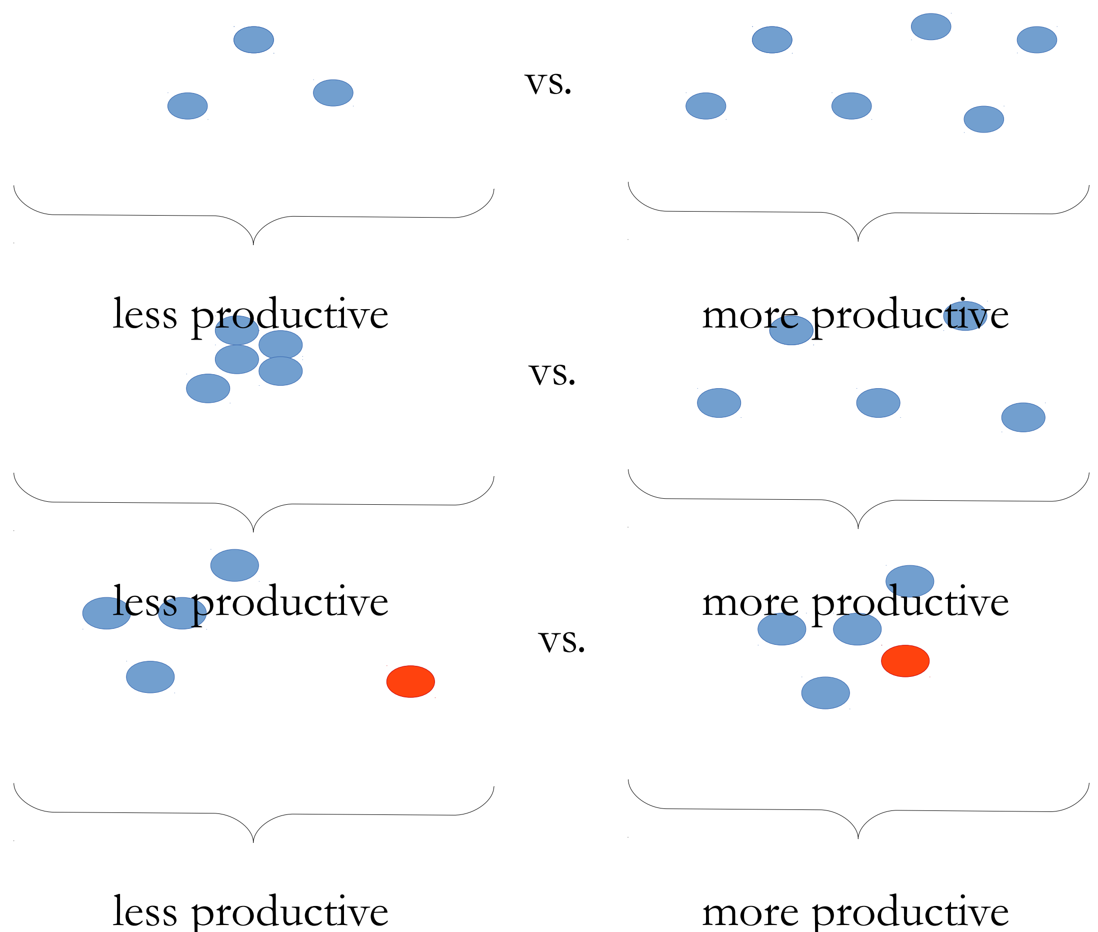
Goldberg (1995)
Bybee and Thompson (1997)
Barðdal (2008)
Wonnacott et al. (2012)

Semantic variability

Goldberg (2006)
Barðdal (2008)

Semantic similarity

Bybee and Eddington (2006)
Bybee (2010)
Suttle and Goldberg (2011)



Previous research

- Points to a strong semantic component in syntactic productivity
  - Productivity depends on the structure of the semantic space
  - Novel uses must be semantically consistent with prior usage
  - cf. the notion of coverage (Suttle and Goldberg 2011, Osherson et al. 1990)

    "the degree to which attested instances 'cover' the category determined jointly by attested instances together with the target coinage" (Suttle and Goldberg 2011: 1254)

- How to operationalize semantics?
  - In previous studies: introspection, semantic norming study (Bybee and Eddington 2006)
  - Proposal: using distributional semantics to measure semantic similarity

Distributional semantics

"You shall know a word by the company it keeps." (Firth 1957: 11)

- Words that occur in similar contexts tend to have related meanings (Miller and Charles 1991)
- Therefore, a way to characterize the meaning of words is through their distribution in large corpora
- Semantic similarity is quantified by similarity in distribution

Distributional semantics

- Vector-space model
  - Assigns an array of values (i.e., a vector) derived from distributional information to each word
  - Semantic similarity measured by similarity between vectors
  - Here, 'bag of words' approach: based on lexical co-occurrences
- Example: shared collocates of *drink*, *sip*, *eat*, and *hear*
  - *drink* and *sip*: names for beverages (*beer*, *coffee*, *tea*), containers for liquids (*glass*, *cup*, *bottle*)
  - *eat* and *drink*/*sip*: words related to dining practices (*bar*, *table*, *dinner*)
  - *hear* share very few collocates with the first three

Case study: The "*hell*-construction"

- V *the hell out of* NP, e.g., *You scared the hell out of me!*
- Intensifying function (broadly defined)
- *Scare* and *beat* most typical, but also a wide range of other verbs:

Then I [...] avoided the hell out of his presence

But you drove the hell out of it!

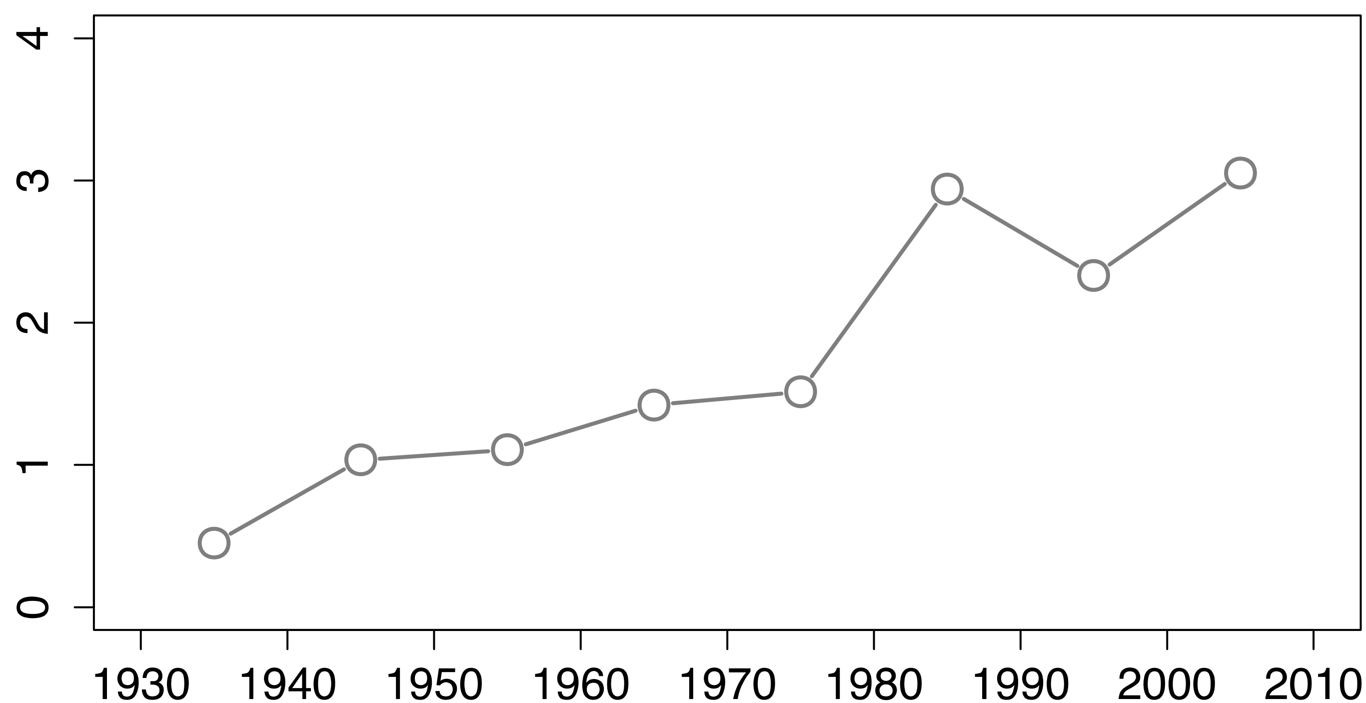
I've been listening the hell out of your tape.

I know the hell out of women!

I voice the hell out of 'b'
(Phillip Hamrick yesterday)

The *hell*-construction in diachrony

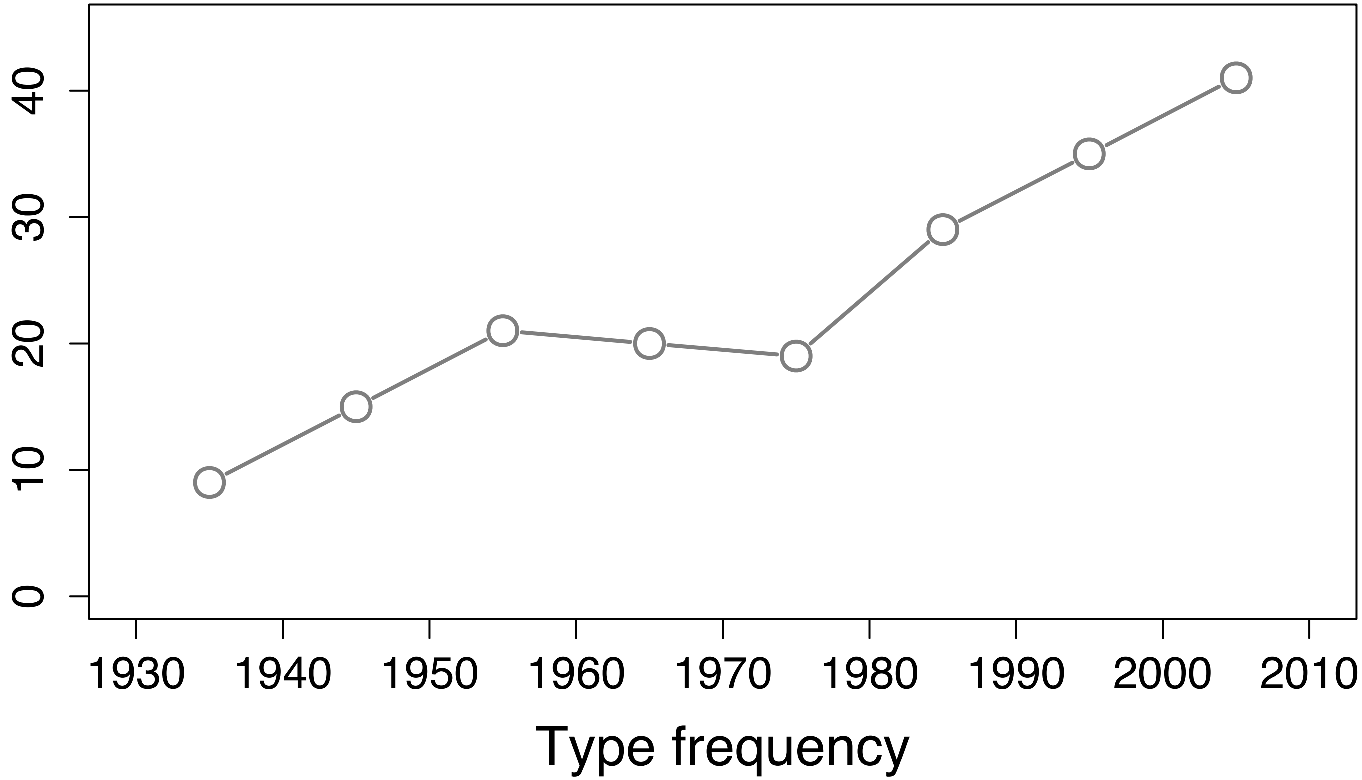
- Data from the Corpus of Historical American English (Davies 2010)
- First attestations in the 1930s
- Steady increase in token frequency since then



Normalized token frequency (per million words)

The *hell*-construction in diachrony

- More and more verb types are used in the construction



The *hell*-construction in diachrony

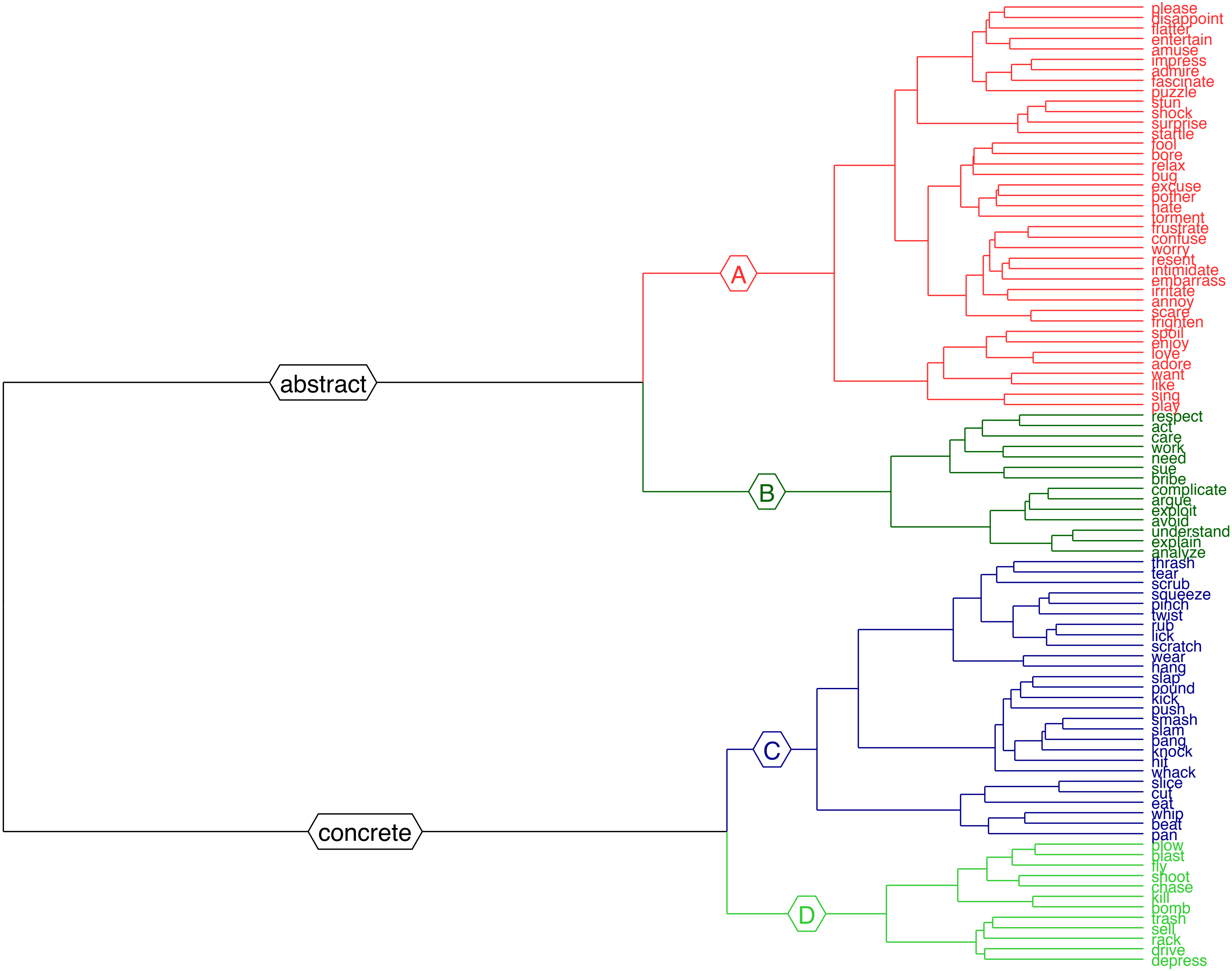
- The increase in type frequency reflects an increase in productivity
- But what kinds of verbs joined the distribution?
  - Did it become more semantically diverse?
  - Are there particular semantic domains favored by the construction?
- Proposal: track the semantic development of the construction by using distributional semantics

Method

- Vector-space model
  - Data from COCA (Davies 2008)
  - Collocates within a 5-word window, lemmatized and PoS-tagged
  - Nouns, verbs, adjectives, and adverbs from the 5,000 most frequent words
- Output of vector-space model: distance matrix
  - Pairwise semantic distances between verbs
  - Cosine distance (between 0 and 1)

Evaluation of the model

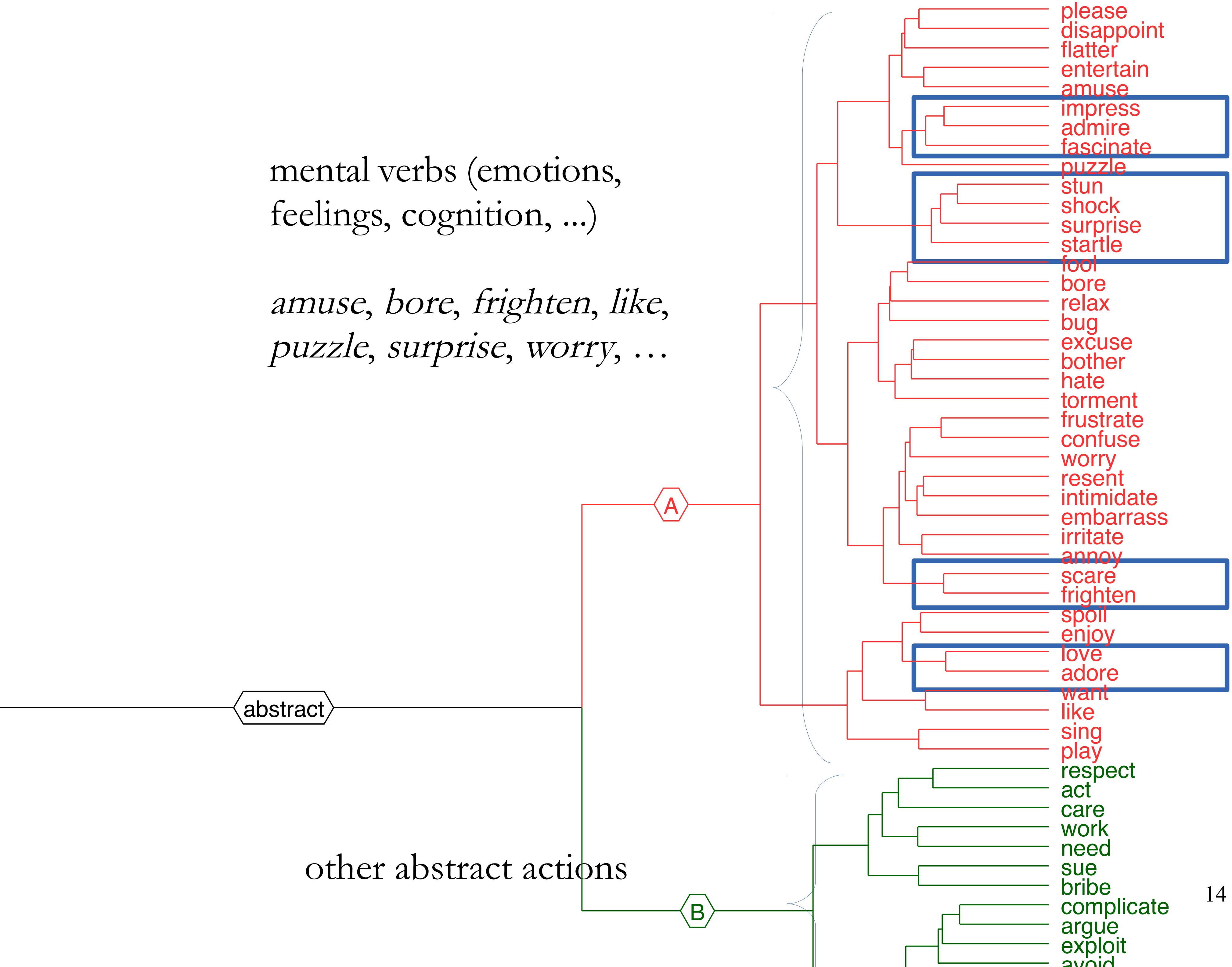
- Hierarchical cluster analysis: groups objects together in a hierarchy by recursively merging the nearest neighbors
- Output: tree diagram (dendrogram)
- Useful to visualize the major semantic distinctions in the distribution

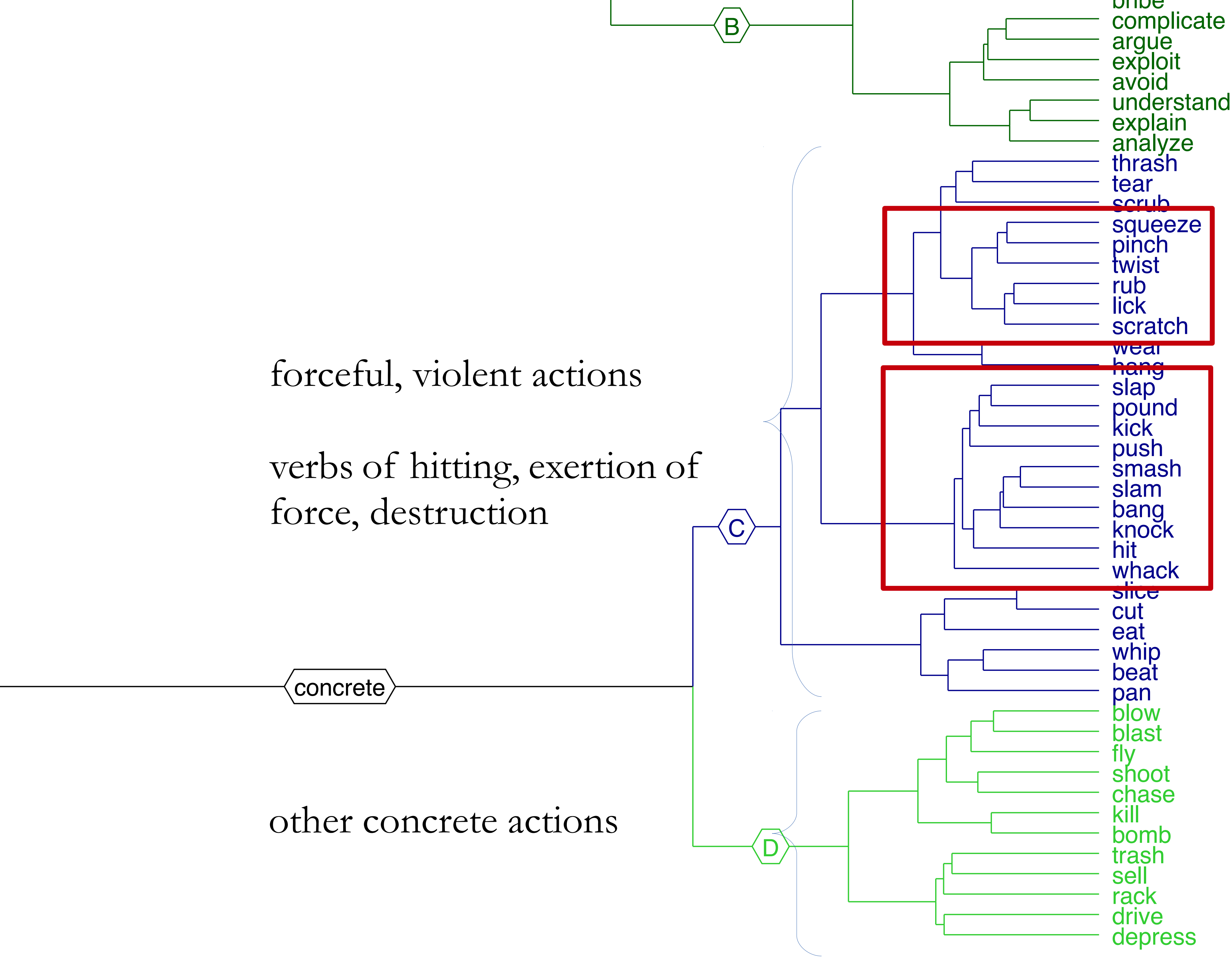

abstract
concrete
A
B
C
D
please
disappoint
flatter
entertain
amuse
impress
admire
fascinate
puzzle
stun
shock
surprise
startle
fool
bore
relax
bug
excuse
bother
hate
torment
frustrate
confuse
worry
resent
intimidate
embarrass
irritate
annoy
scare
frighten
spoil
enjoy
love
adore
want
like
sing
play
respect
act
care
work
need
sue
bribe
complicate
argue
exploit
avoid
understand
explain
analyze
thrash
tear
scrub
squeeze
pinch
twist
rub
lick
scratch
wear
hang
slap
pound
kick
push
smash
slam
bang
knock
hit
whack
slice
cut
eat
whip
beat
pan
blow
blast
fly
shoot
chase
kill
bomb
trash
sell
rack
drive
depress

mental verbs (emotions, feelings, cognition, ...)

amuse, bore, frighten, like, puzzle, surprise, worry, …

other abstract actions





forceful, violent actions
verbs of hitting, exertion of force, destruction
other concrete actions
concrete
B
C
D
complicate
argue
exploit
avoid
understand
explain
analyze
thrash
tear
scrub
squeeze
pinch
twist
rub
lick
scratch
wear
hang
slap
pound
kick
push
smash
slam
bang
knock
hit
whack
slice
cut
eat
whip
beat
pan
blow
blast
fly
shoot
chase
kill
bomb
trash
sell
rack
drive
depress

How to visualize the semantic domain of the construction?

- Can be plotted by means of multidimensional scaling (MDS)
  - Places objects in a 2-dimensional space such that the between-object distances are preserved as well as possible
  - Distance matrix converted to a set of coordinates
- Four vector-space semantic maps (1 per 20-year period)
  - 1930-1949
  - 1950-1969
  - 1970-1989
  - 1990-2009

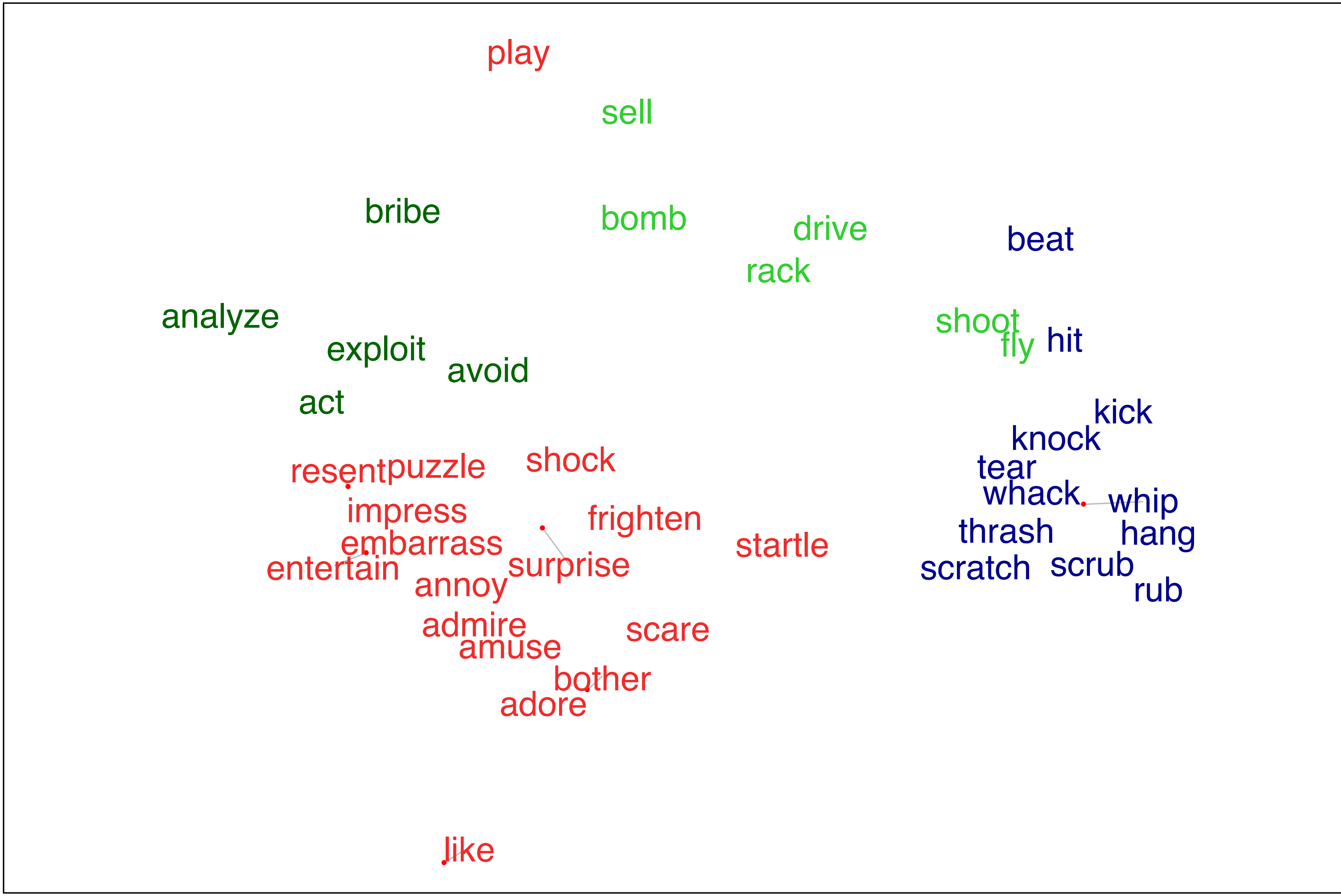
1930s – 1940s


work
beat
shoot
worry
smash
kick
knock
chase
tear
bore
whip
surprise
please
lick
scare
bother
want
eat
love

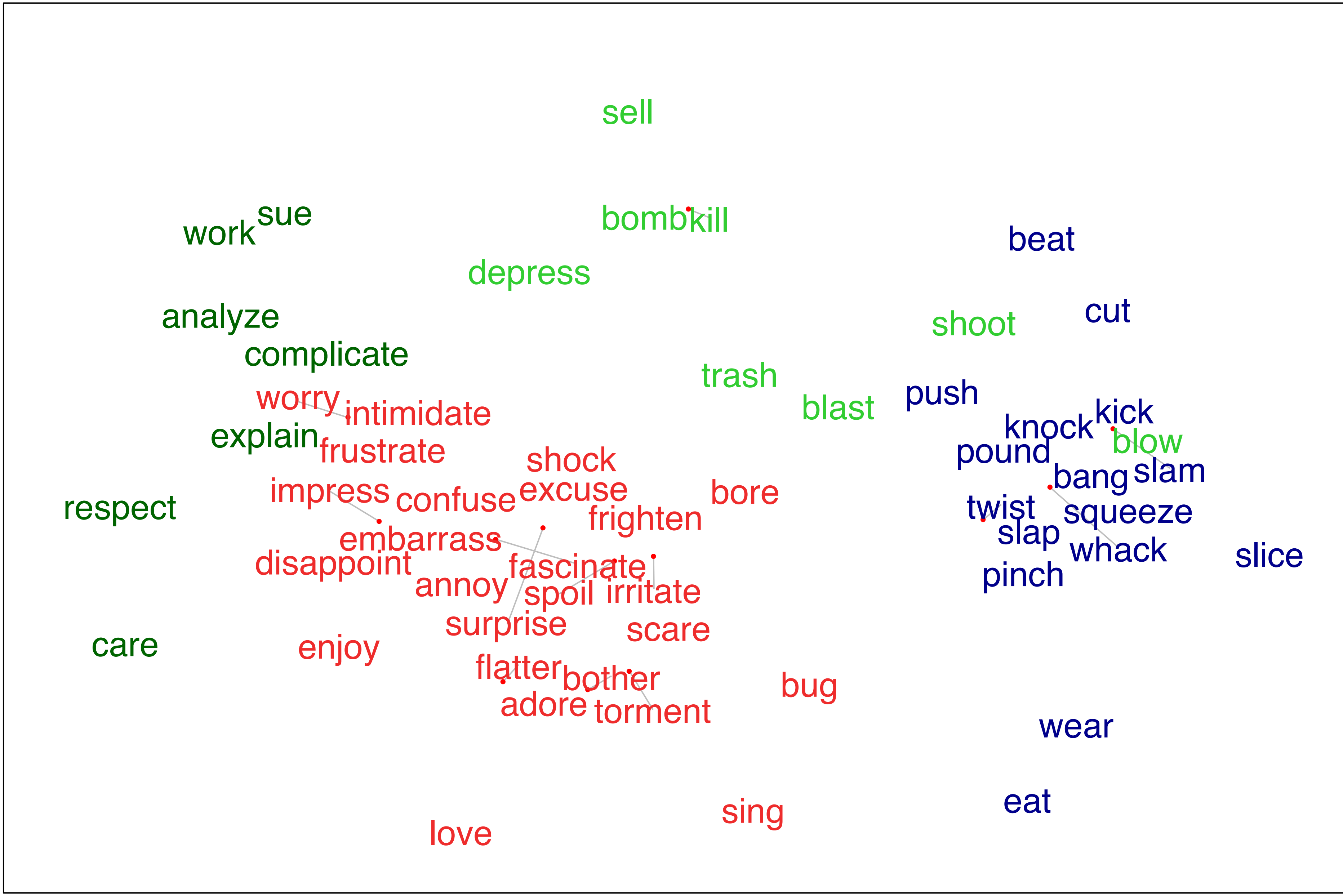
1950s – 1960s

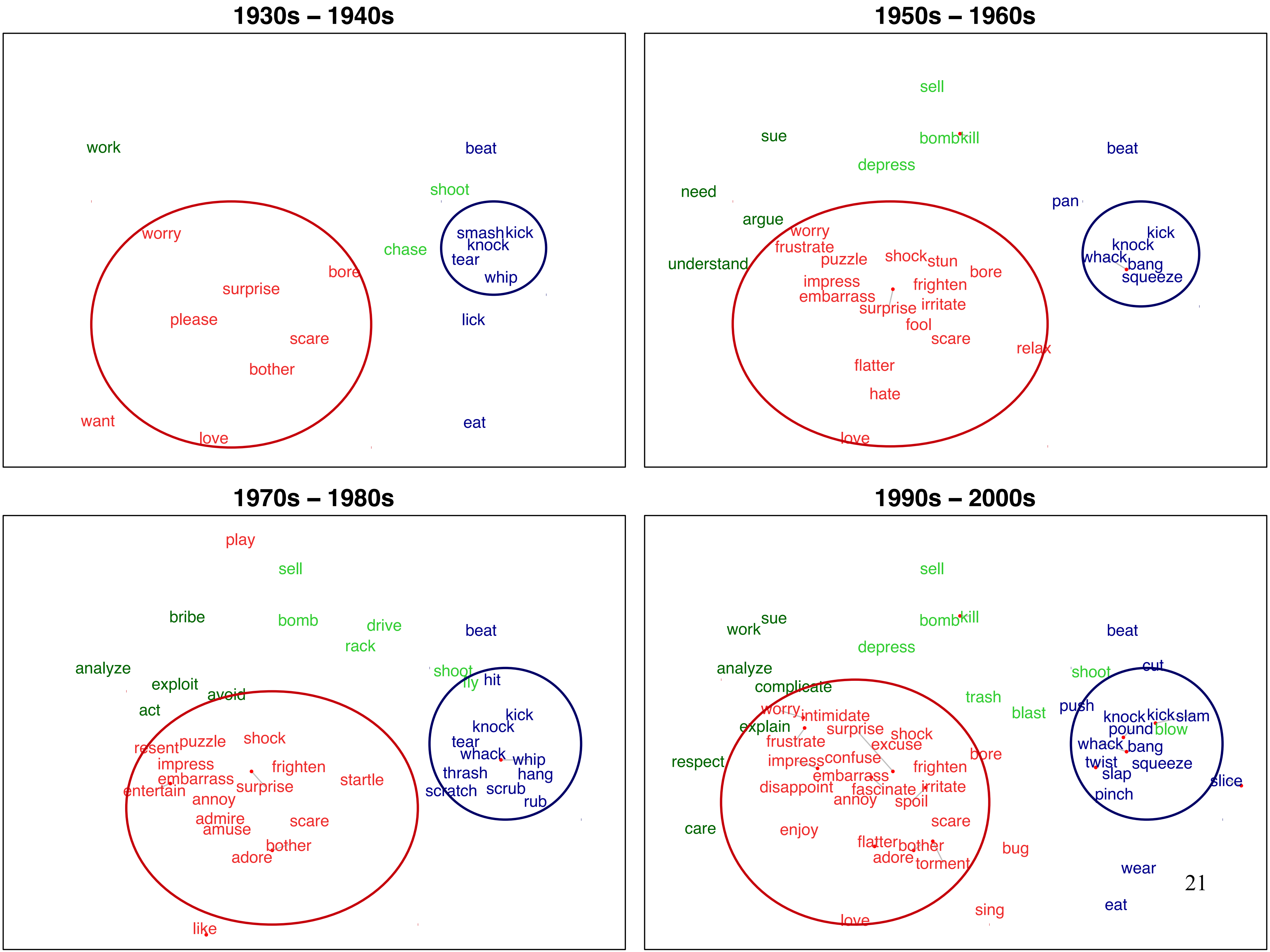

sell
sue
bombkill
beat
depress
need
pan
argue
worry
kick
frustrate
knock
shock
stun
puzzle
understand
whack
bang
surprise
bore
squeeze
impress
frighten
embarrass
irritate
fool
scare
relax
flatter
hate
love

1970s – 1980s


play
sell
bribe
bomb
drive
beat
rack
analyze
shoot
exploit
fly
hit
avoid
act
kick
knock
resent
puzzle
shock
tear
whack
whip
impress
frighten
hang
thrash
embarrass
startle
entertain
surprise
scratch
scrub
annoy
rub
admire
scare
amuse
bother
adore
like

1990s – 2000s


sell
work
sue
bomb
kill
depress
beat
analyze
shoot
cut
complicate
trash
worry
intimidate
blast
push
explain
frustrate
knock
kick
blow
pound
shock
bang
slam
impress
excuse
confuse
bore
respect
twist
squeeze
frighten
slap
embarrass
whack
slice
disappoint
fascinate
annoy
pinch
spoil
irritate
surprise
scare
care
enjoy
flatter
bother
bug
adore
torment
wear
eat
sing
love



1930s – 1940s
work
beat
shoot
worry
chase
smash
kick
knock
tear
whip
bore
surprise
please
lick
scare
bother
want
love
eat
1950s – 1960s
sell
sue
bomb
kill
beat
depress
need
pan
argue
worry
frustrate
kick
knock
whack
bang
squeeze
understand
puzzle
shock
stun
bore
impress
frighten
embarrass
irritate
surprise
fool
scare
relax
flatter
hate
love
1970s – 1980s
play
sell
bribe
bomb
drive
beat
rack
analyze
shoot
fly
hit
exploit
avoid
act
kick
knock
resent
puzzle
shock
tear
whack
whip
impress
frighten
hang
thrash
embarrass
startle
entertain
surprise
scratch
scrub
annoy
rub
admire
amuse
scare
bother
adore
like
1990s – 2000s
sell
sue
bomb
kill
work
beat
depress
analyze
shoot
cut
complicate
trash
blast
push
worry
intimidate
knock
kick
slam
explain
surprise
shock
pound
blow
frustrate
excuse
whack
bang
bore
impress
confuse
respect
frighten
twist
squeeze
embarrass
slap
disappoint
fascinate
irritate
slice
annoy
spoil
pinch
scare
care
enjoy
flatter
bother
bug
adore
torment
wear
sing
love
eat

Summary

- Densely populated regions are more likely to attract new members
- New verbs appear either close to or inside a cluster
- In line with previous accounts
- Can we derive quantitative evidence for these observations?

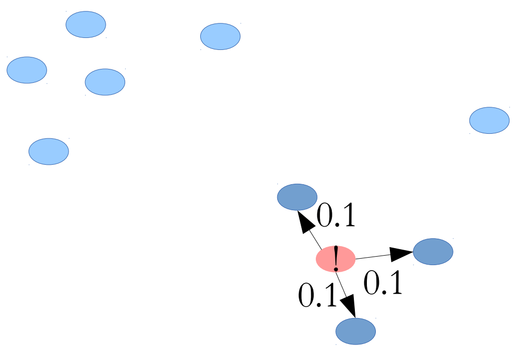
Statistical analysis

- Assumptions
  - Semantic contribution of the construction is constant
  - Hence, all verbs that ever occurred in it are equally plausible
  - Why do some verbs first occur later than others?
- Hypothesis

  The probability of occurrence of a new item is related to the density of the semantic space around this item in prior usage

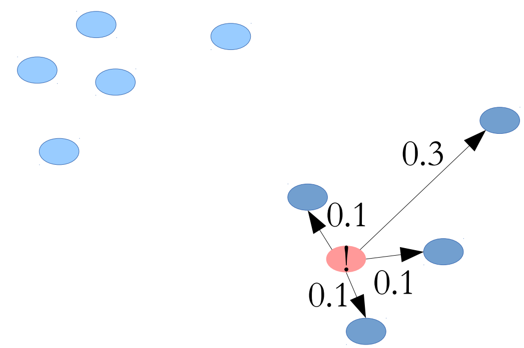
Measure of density

- Considers a subset of the nearest neighbors of a verb
- Density = 1 – mean distance to the N nearest neighbors
- N = 3 to 8



with N = 3

Density = 1 – (0.1 + 0.1 + 0.1)/3
= 0.9



with N = 4

Density = 1 – (0.3 + 0.1 + 0.1 + 0.1)/4
= 0.850

Mixed effects logistic regression

- Dependent variable: Occurrence (0/1)
- Data: Verb x Period x Occurrence triplets
- Density around the verb calculated for the preceding period
- For each verb, Occurrence = 1 on its first attestation, 0 in earlier periods
- Model:

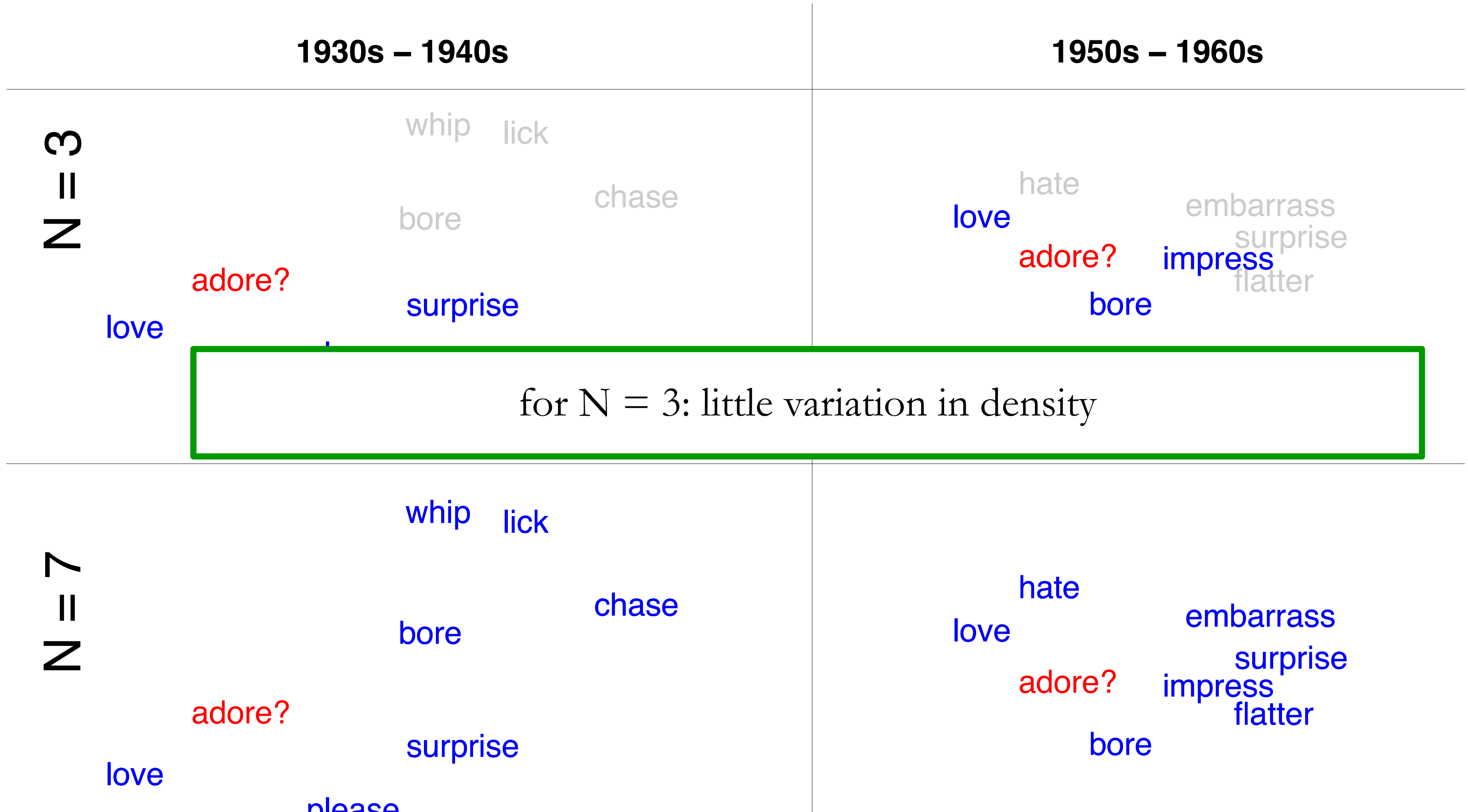
  Occurrence ~ Density + (1 + Density | Verb)

Results

Number of neighbors (N)	Effect of Density	*p*-value
3	0.7211	0.195
4	0.8836	0.135
5	1.0487	0.091 .
6	1.2367	0.056 .
7	1.4219	**0.034 ***
8	1.6625	**0.017 ***

- Positive effect of Density, significant for N = 7 and 8
  - Evidence for a relation between density and the likelihood of a new coinage
- Effect strength increases and *p*-value decreases with N
  - Considering a higher number of neighbors increases the strength *and* reliability of the measure of density as a predictor of novel uses

The role of N: example with *adore*



for N = 3: little variation in density

for N = 7: density increases

Conclusion

- Distributional semantics is appropriate for the study of syntactic productivity in diachrony; benefits:
  - Turns the informal notion of meaning into a quantified representation
  - Fully automatic and data-driven
  - Virtually no limit on the number of items to be considered
  - Enables the use of visualization techniques and statistical analysis
- Distribution-based account consistent with current views
- Promising approach to the study of syntactic productivity

I thank the hell out of you!

Barðdal, J. (2008). *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, J. & D. Eddington (2006). A usage-based approach to Spanish verbs of 'becoming'. *Language* 82 (2), 323–355.

Bybee, J. & S. Thompson (1997). Three frequency effects in syntax. *Berkeley Linguistics Society* 23, 65–85.

Davies, M. (2008). *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/

Davies, M. (2010). *The Corpus of Historical American English: 400 million words, 1810-2009*. Available online at http://corpus.byu.edu/coha/

Firth, J.R. (1957). A synopsis of linguistic theory 1930-1955. In *Studies in Linguistic Analysis*, pp. 1-32. Oxford: Philological Society.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Israel, M. (1996). The way constructions grow. In A. Goldberg (ed.), *Conceptual structure, discourse and language*. Stanford, CA: CSLI Publications, 217-230.

Miller, G. & W. Charles (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1), 1-28.

Osherson, D., W. Ormond, E. Smith, A. Lopez & E. Shafir (1990). Category based induction. *Psychological Review* 97(2), 185–200.

Suttle, L. & A. Goldberg (2011). The partial productivity of constructions as induction. *Linguistics* 49 (6): 1237–1269.

Wonnacott, E., J. Boyd, J. Thompson & A. Goldberg (2012). Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language* 66: 458–478.