

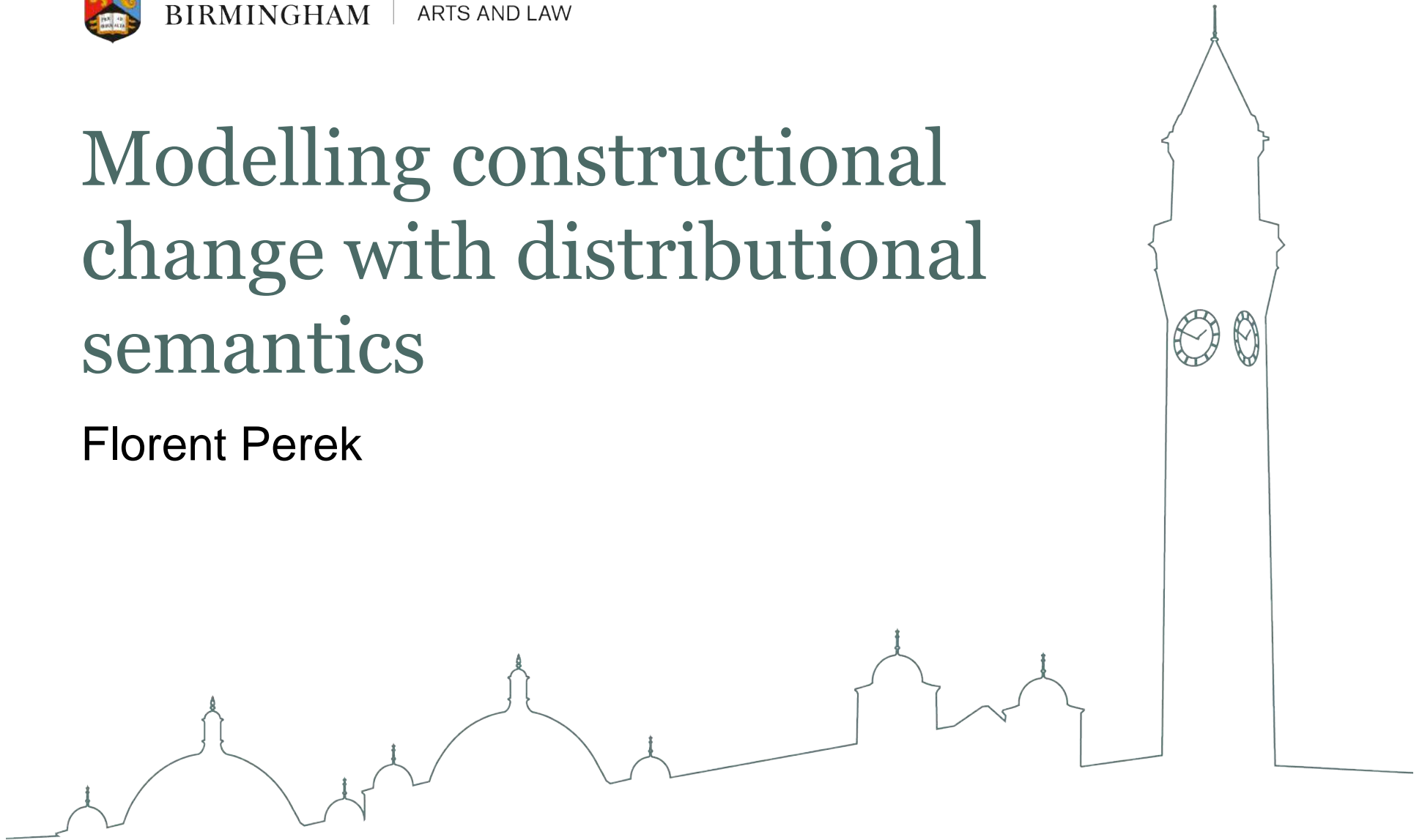


UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

Modelling constructional change with distributional semantics

Florent Perek



Overview

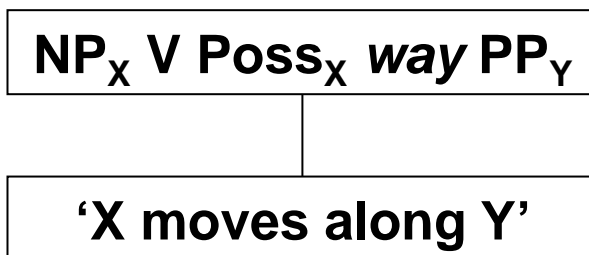
- Applying distributional semantics to diachronic studies
- Introduction: diachronic construction grammar
- Problem: productivity and schematicity in corpus data
- Two methods drawing on distributional semantics
- Case studies

Diachronic construction grammar

- New approach to language change (Traugott & Trousdale 2013)
- Grammar seen as inventory of form-meaning pairs, aka constructions (Goldberg 1995)
- E.g., the *way*-construction

They hacked their way through the jungle

We pushed our way into the bar



Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

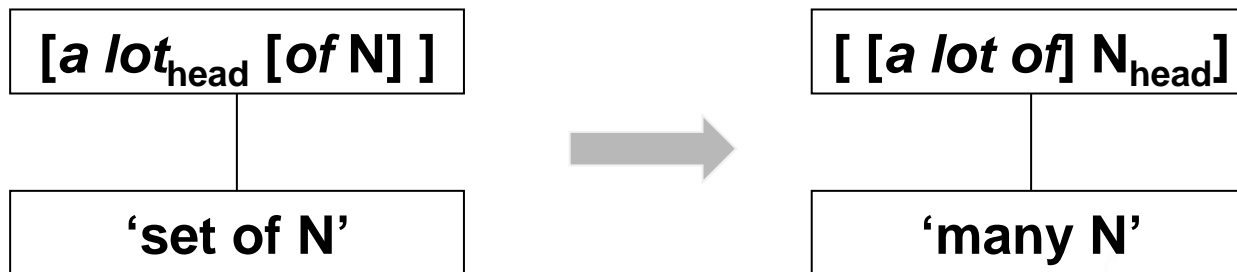
Traugott, E. & G. Trousdale (2013). *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Constructions

- Constructions come in all shapes and sizes
- Words: *freckle, yellow, bespectacled, anyone*
- Partly-filled words: N-s, *un-Adj, V-ment*
- Idioms: *throw in the towel, think out of the box*
- Word order patterns: NP V NP NP (ditransitive),
NP BE V-ed (*by NP*) (passive)

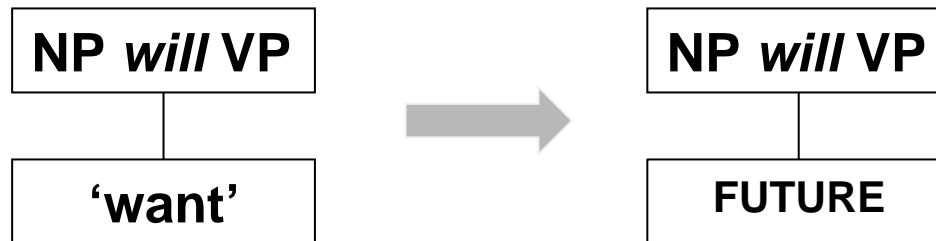
Two types of change

- Two types of change in DCxG: constructionalisation and constructional change
- Constructionalisation
 - Creation of a new form-meaning
 - Usually from instances of existing constructions
 - E.g.: *a lot of N* (binominal quantifier)



Constructional change

- Change in the form or meaning of existing constructions
- E.g., *will*



The study of constructional change

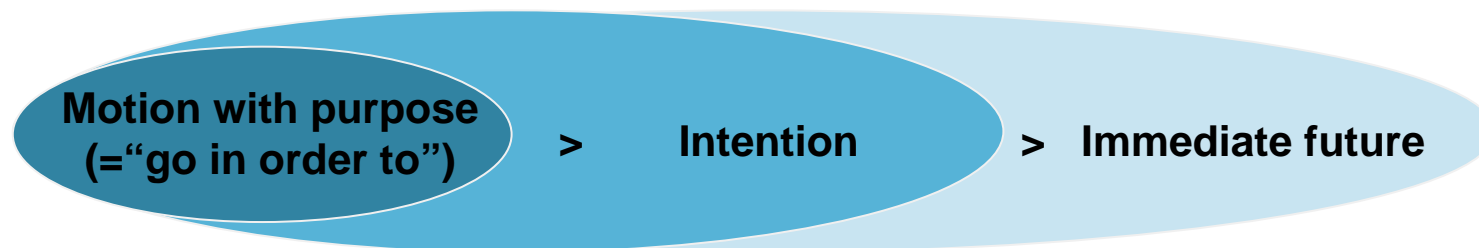
- DCxG = usage-based theory
 - Important aspects of grammatical representations are shaped by natural language use
 - Constructional change can be characterized by examining usage data, i.e., from corpora
- Two aspects of constructions are commonly described:
 1. Productivity
 2. Schematicity

Productivity

- The range of lexical items that can be used in the slots of a construction
- E.g., verbs in the *way*-construction (Israel 1996)
 - Verbs of physical actions attested from the 16th century
They hacked their way through the jungle.
 - Abstract means only appear in the 19th century
She talked her way into the club.

Schematicity

- Increase/decrease in schematicity = the meaning of the construction becomes more general/more specific
- Example: the *be going to* future



They are going (outside)
to harvest the crop.

I'm going to be an
architect.

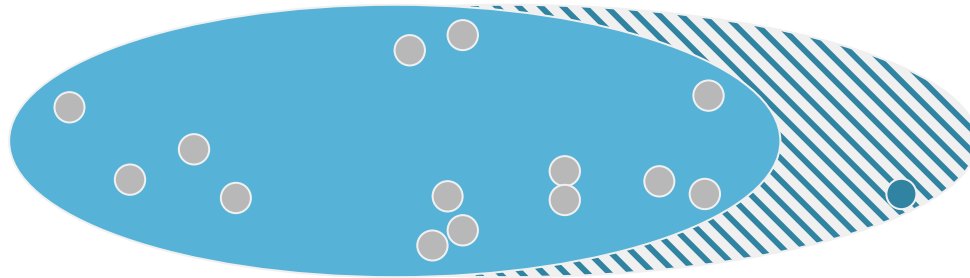
It's going to rain
today.

Productivity and schematicity

- Commonly thought to be interrelated (Barðdal 2008)
- A more schematic meaning can be applied to a wider range of situations
- Hence, more items are compatible with the schema
- Example: the *be going to* future
 - Stative verbs are incompatible with an intentional reading: *like, know, want, see, hear, feel, etc.*
 - The futurity meaning makes them compatible with the construction

Productivity and schematicity

- Conversely, the occurrence of new types may contribute to schema extension
- If a new type is not covered by the schema, the latter must be implicitly adjusted

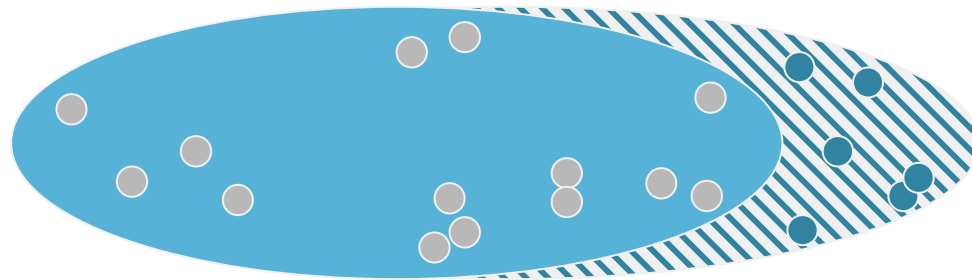


● : attested type

● : new type

Productivity and schematicity

- If repeated, creative uses that once sounded 'deviant' can become conventional through schema extension

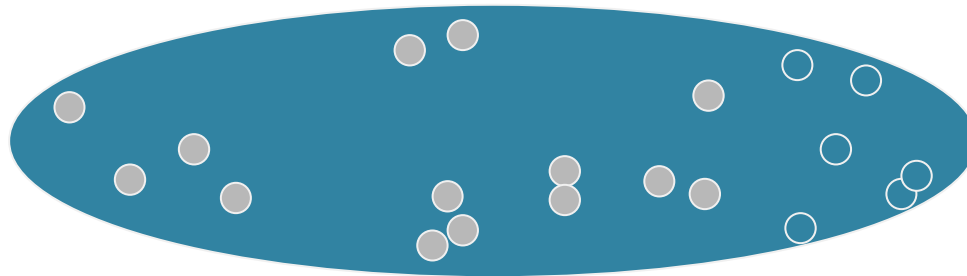


● : attested type

● : new type

Productivity and schematicity

- If repeated, creative uses that once sounded 'deviant' can become conventional through schema extension



● : attested type

● : new type

Productivity and schematicity

- Two types of schema extension
 - Change in the constructional meaning
 - Change in the semantic restrictions on the slots of the construction (host-class expansion, Himmelmann 2004)
 - e.g., quantifier *a lot of* N: gradual expansion from concrete entities to increasingly abstract ones
- Depends on how new types are related to attested types (Suttle & Goldberg 2011) and to the construction
- Conclusion: interpreting changes in productivity requires an assessment of the meaning of new types

Himmelmann, N. (2004). Lexicalization and grammaticization: Opposite or orthogonal? In Bisang, W., Himmelmann, N. P., & Wiemer, B. (eds.), *What Makes Grammaticalization: A look from its components and its fringes* (pp. 21–42). Berlin: Mouton de Gruyter.

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

Operationalizing meaning

- Semantic intuitions
 - Manual identification of semantic trends in the data
 - Potentially subjective and limited by one's introspection
 - Does not lend itself to precise quantification
- Semantic norming (Bybee & Eddington 2006)
 - Similarity judgments provided by a group of speakers
 - Also time-consuming and constraining
 - Limited in terms of the number of lexical items considered

Distributional semantics

- A third alternative: distributional semantics
- Widely used in computational linguistics and NLP
- “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 & Charles 1991)
 - Distributional Semantic Models (DSMs) capture the meaning of words through their distribution in large corpora

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

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

Distributional semantics

- Offers a solution to these problems:
 - Data-driven: more objective, no manual intervention needed
 - No limits on the number of lexical items
 - Precise quantification
- Robust, adequately reflects semantic intuitions
 - Correlates with human performance (e.g., Landauer et al. 1998, Lund et al. 1995)
 - Evidence for some psychological adequacy (Andrews & Vigliocco 2008)

Andrews, Mark, Gabriella Vigliocco & David P. Vinson. 2009. Integrating Experiential and Distributional Data to Learn Semantic Representations. *Psychological Review* 116(3). 463–498.

Landauer, Thomas K., Peter W. Foltz & Darrell Laham. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes* 25. 259–284.

Lund, Kevin, Curt Burgess & Ruth A. Atchley. 1995. Semantic and associative priming in a high-dimensional semantic space. In *Cognitive Science Proceedings (LEA)*, 660–665.

Two methods

- **Distributional semantic plots**

 - To visualize the semantic development of lexical slots of constructions

- **Distributional period clustering**

 - To partition this development into stages

Distributional semantic plots

- Visual representation of the semantic spectrum of a construction
- Semantic distance can be derived from DSMs
 - Semantic similarity is quantified by similarity in distribution
 - Capture how words are related to each others
 - Can be interpreted as distance in a semantic space

Distributional semantic plots

1. Determine the lexical distribution of a construction at different points in time
2. Create a DSM containing (at least) all lexical items ever attested in the construction
3. Compute pairwise distances between all items from the DSM
4. Use the set of distances to locate each item with respect to the others
5. Plot the distribution at different points in time

Distributional semantic maps

- Pairwise distances converted to set of coordinates
- Achieved with, e.g, multidimensional scaling (MDS)
- Here, t -Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten & Hinton 2008)
 - Places objects in a 2-dimensional space such that the between-object distances are preserved as well as possible
 - Superior to MDS for dense spaces with many dimensions
 - Proven solution for visualizing DSMs

Corpus and DSM

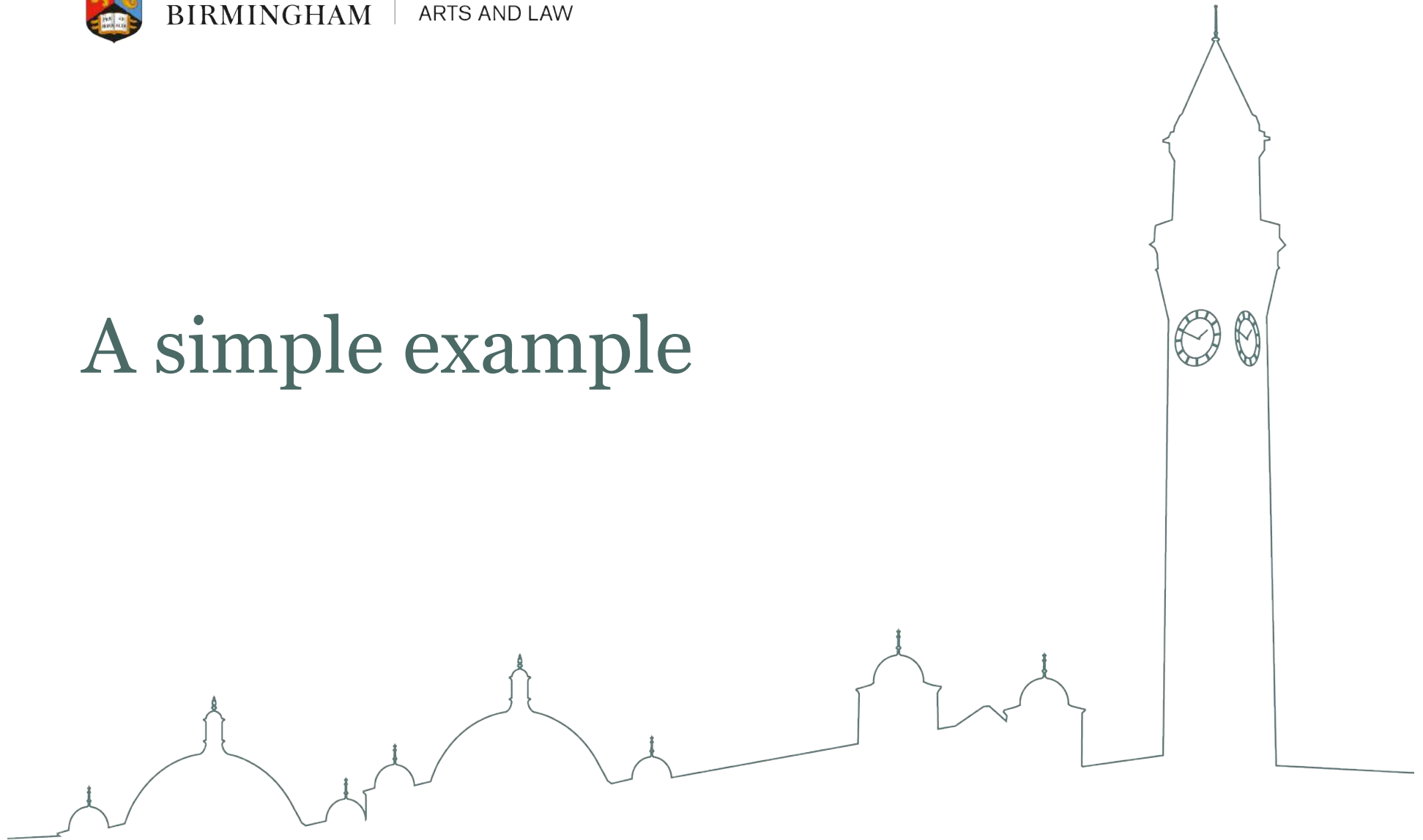
- Distributional data extracted from the Corpus of Historical American English (COHA; Davies 2010)
 - 400 MW from 1810 to 2009
 - Balanced by decade and genre (fiction, mag, news, non-fict)
- “Bag of words” approach: collocates in a 2-word window
- Restricted to the 10,000 most frequent nouns, verbs, adjectives and adverbs
- PPMI weighting, reduced to 300 dimensions with SVD
- Two models: all verbs, all nouns (both with $F > 1000$)



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

A simple example



The *hell*-construction

- Verb *the hell out of* NP (Perek 2014, 2016)
- “Intensifying” function

You scared the hell out of me!

I enjoyed the hell out of that show!

But you drove the hell out of it!

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

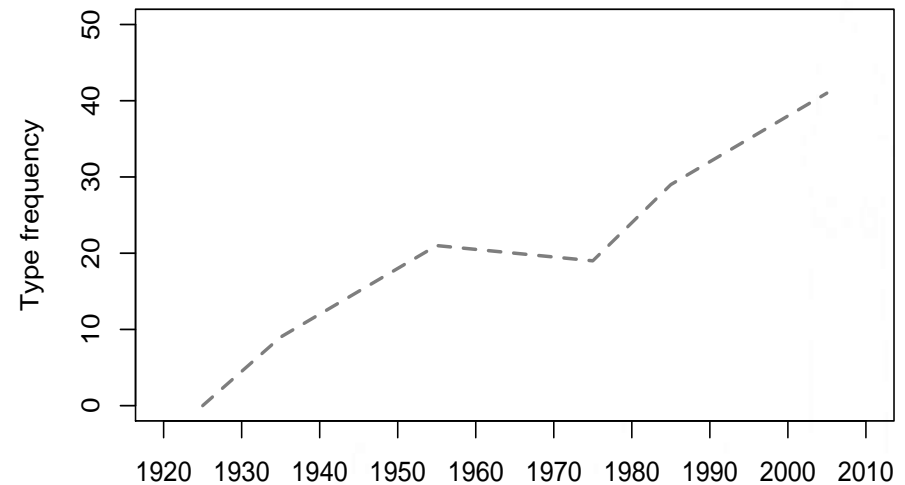
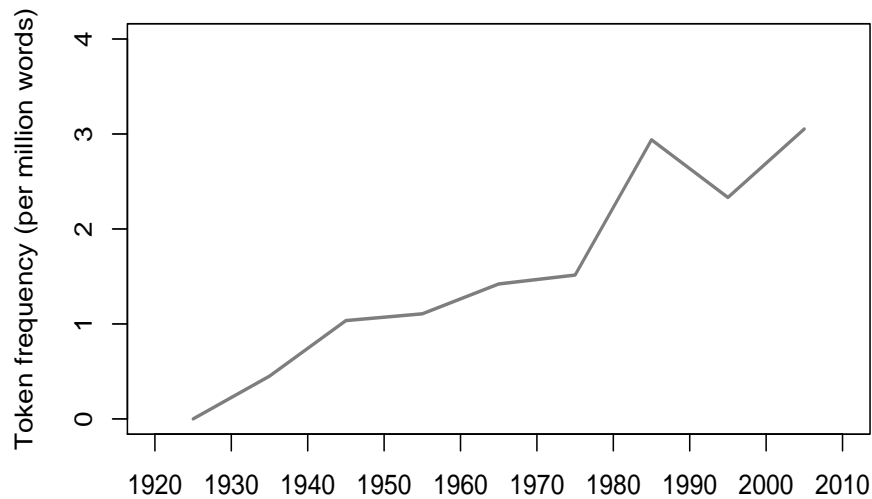
I voiced the hell out of 'b' (heard at GURT 2014, Georgetown)

Perek, F. (2014). Vector spaces for historical linguistics: Using distributional semantics to study syntactic productivity in diachrony. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland USA, June 23-25 2014* (pp. 309-314).

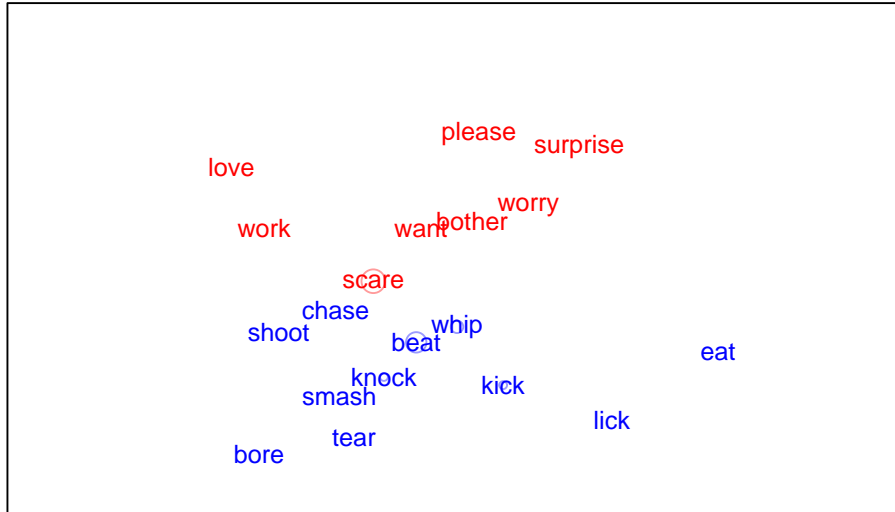
Perek, F. (2016). Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics*, 54(1), 149–188.

The *hell*-construction in the COHA

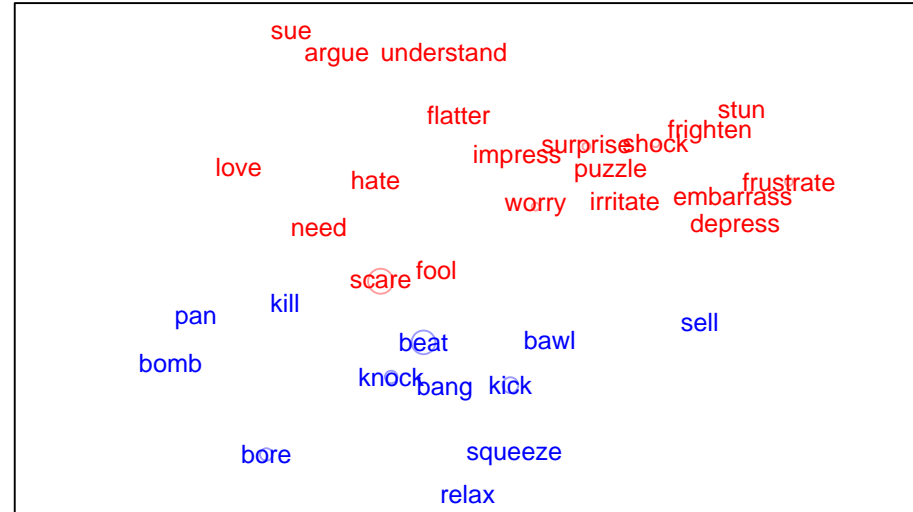
- Recent construction: first instances in the 1930s
- Increasingly popular
- More and more verbs in the construction



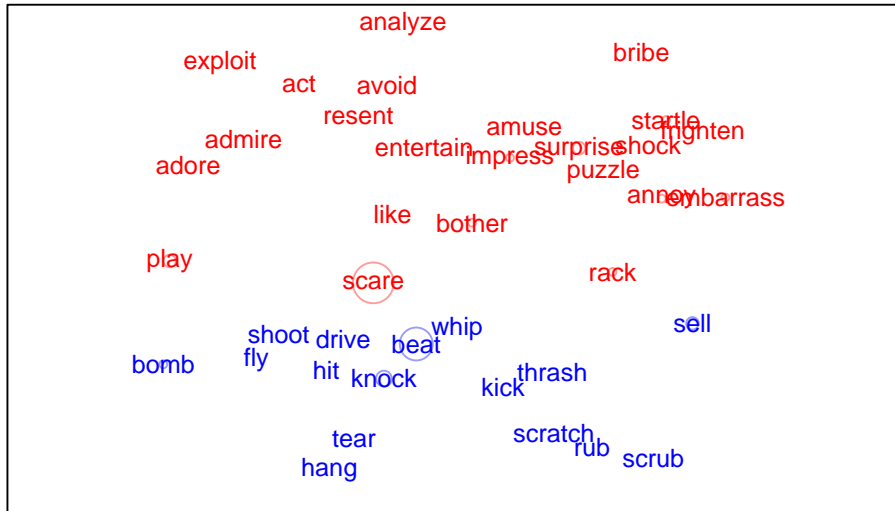
1930-1949



1950-1969



1970-1989



1990-2009



Red: emotions, feelings, thoughts, mental activities

Blue: violent contact, exertion of force

Two domains of predilection

- Cognition verbs

bother, disappoint, shock, startle, worry

adore, enjoy, impress, love, want

analyze, explain, understand

- Verbs of hitting and other forceful actions

beat, knock, hit, kick, slap

push, squeeze, twist

blast, kill, shoot

Change in the *hell*-construction

- Schema centered on these two classes
- Few members outside of them: e.g., *drive*, *sell*, *sing*, *wear*
- Too sporadic to cause schema extension
- Increase in productivity, little to no increase in schematicity



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

A more complex example

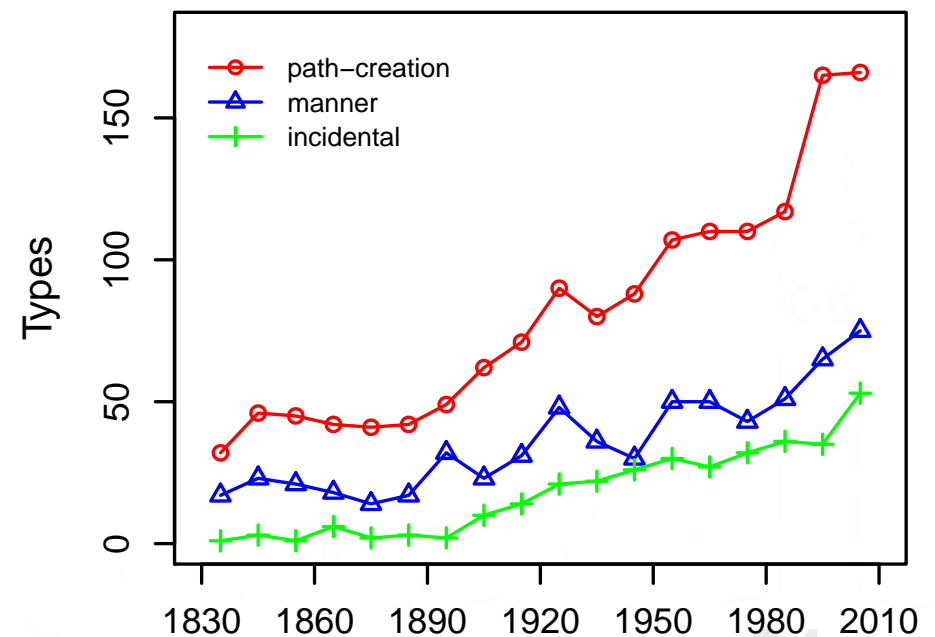
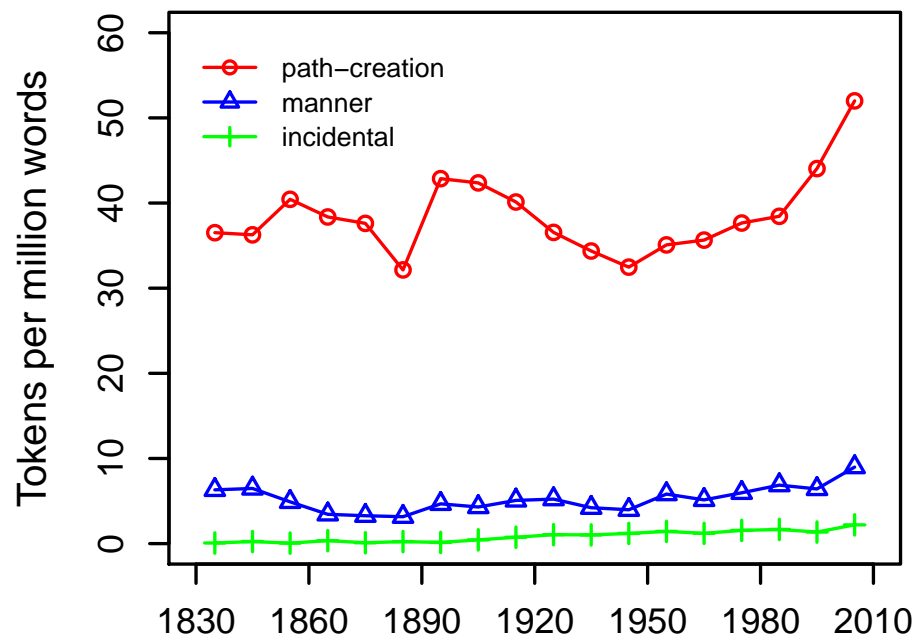


The *way*-construction

- Verb *one's way* PP (Perek, submitted)
- Describes motion of the subject referent
- Two senses of the construction:
 - Path-creation: the verb describes what enables motion
They hacked their way through the jungle.
 - Manner: the verb describes the manner of motion
They trudged their way through the snow
 - A third sense, incidental-action (not discussed here): the verb refers to some co-occurring action unrelated to motion
He whistled his way across the room

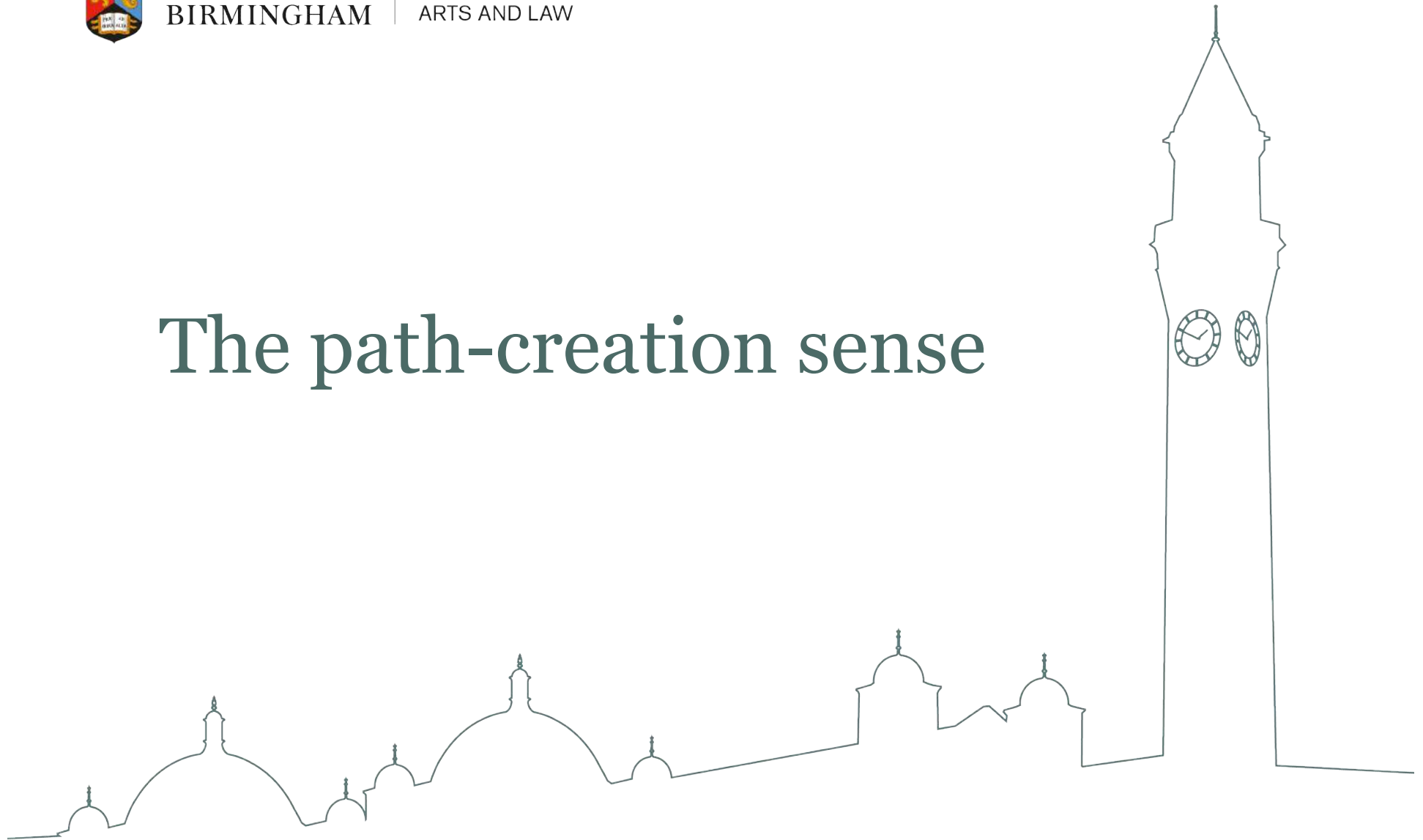
Data

- All tokens of “V Poss way Prep” from 1830 to 2009
- Manually filtered, annotated for constructional meaning: path-creation, manner, incidental-action

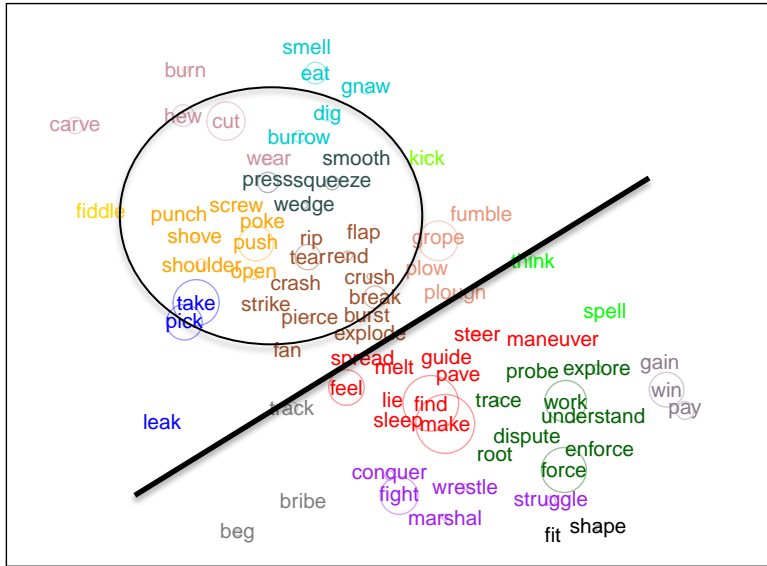




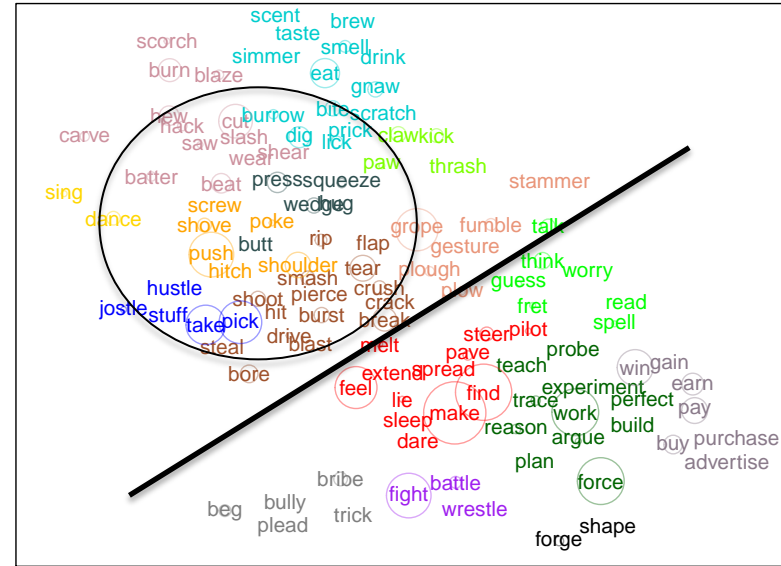
The path-creation sense



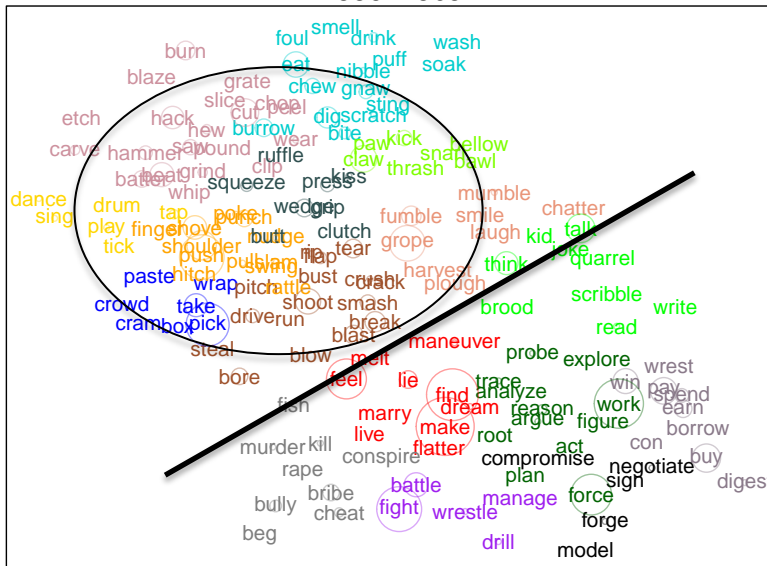
1830-1879



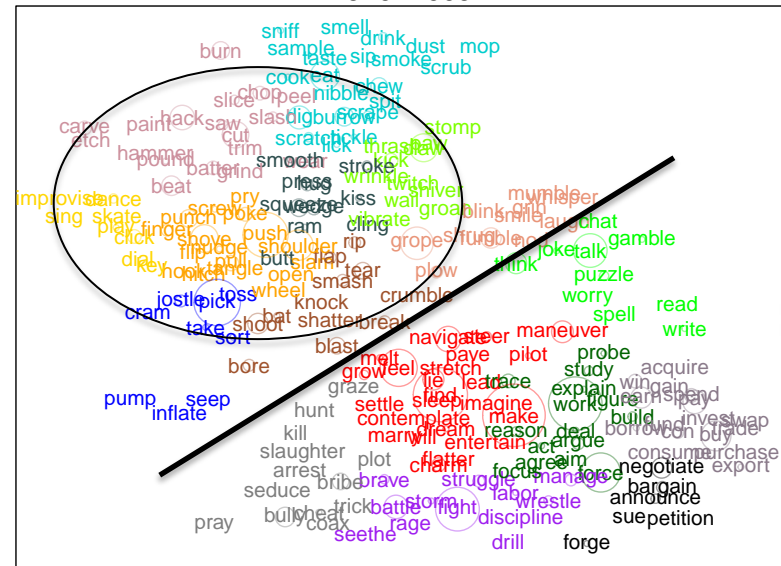
1880-1929



1930-1969



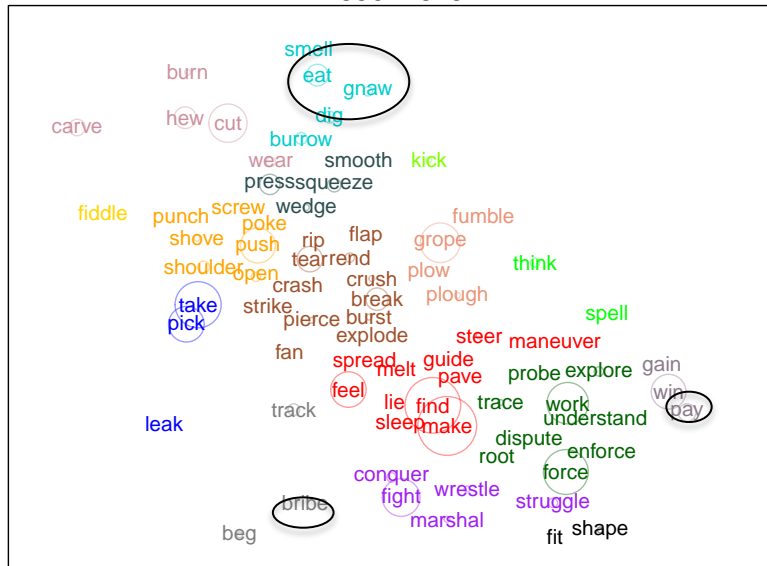
1970-2009



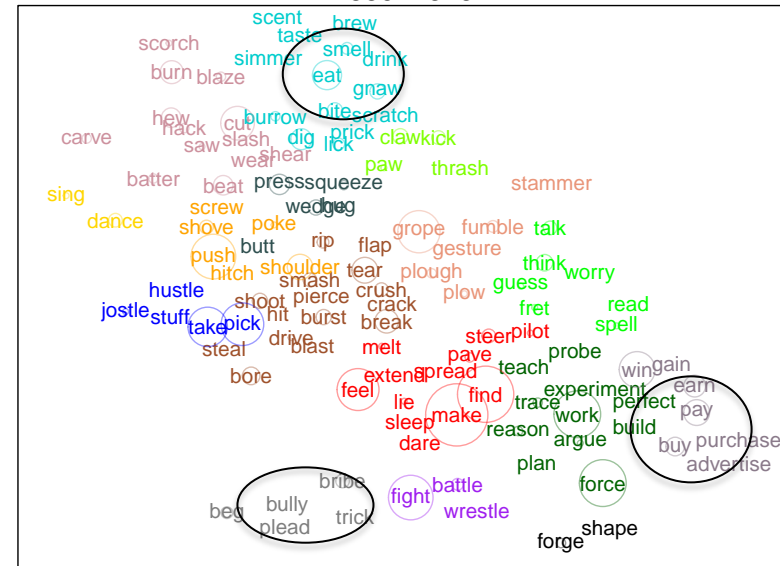
Clear concrete/abstract divide in the distributional semantic plot

Higher density of verbs describing forceful actions (*cut, push, kick, ..*), especially in earlier periods

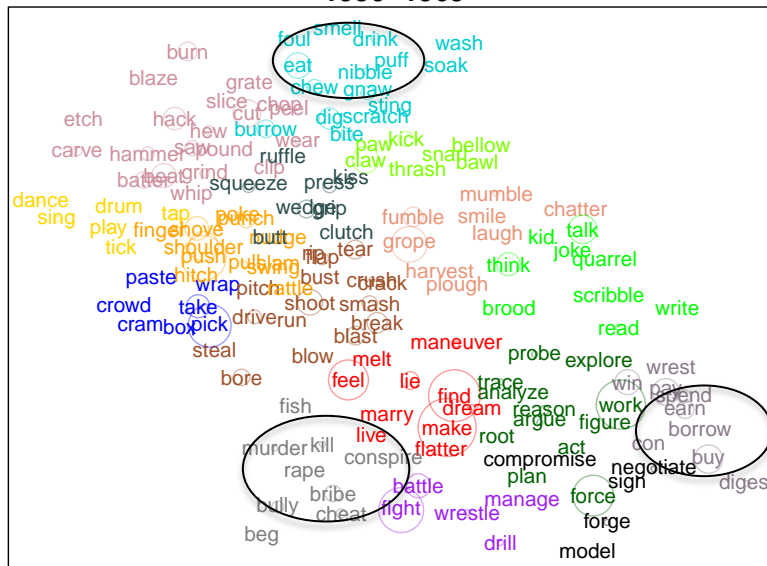
1830-1879



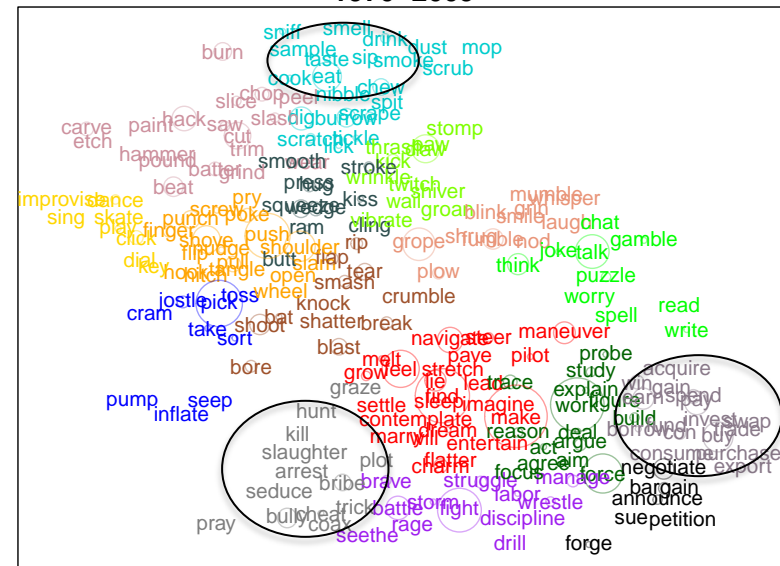
1880-1929



1930-1969

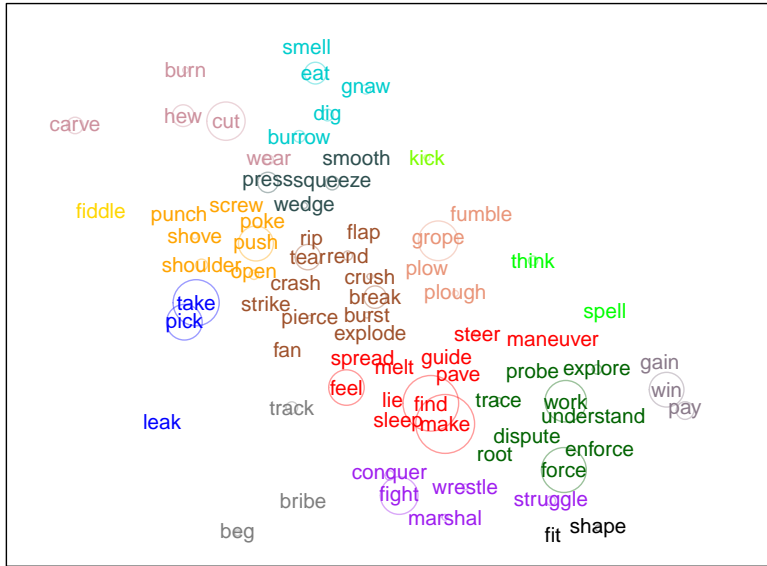


1970-2009

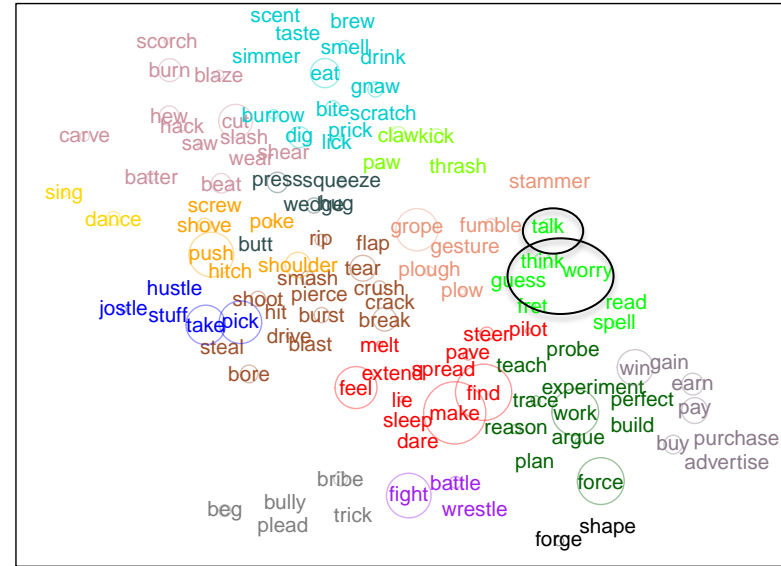


From period 2 onwards: ingestion (*eat, drink, nibble, puff, sip, smoke, ..*), commerce & finance (*buy, export, fund, invest, pay, spend, ..*), misconduct (*bribe, bully, cheat, conspire, kill, murder, plot, rape, trick, ..*)

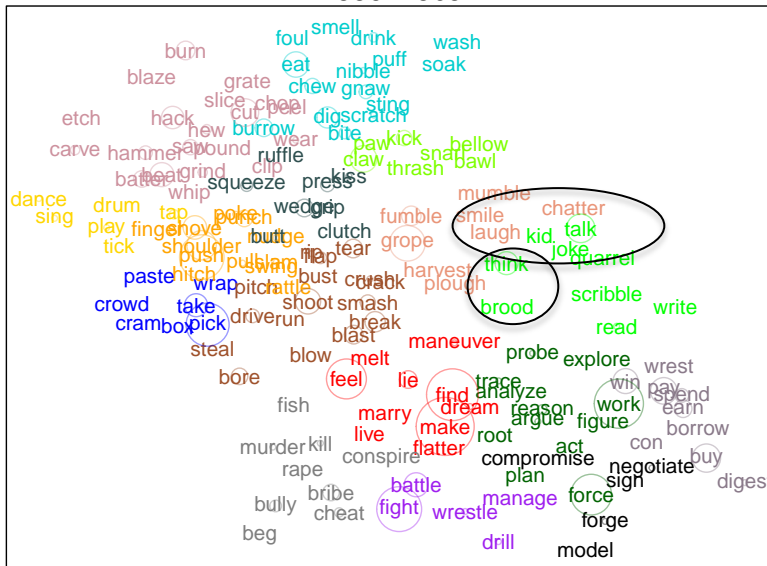
1830–1879



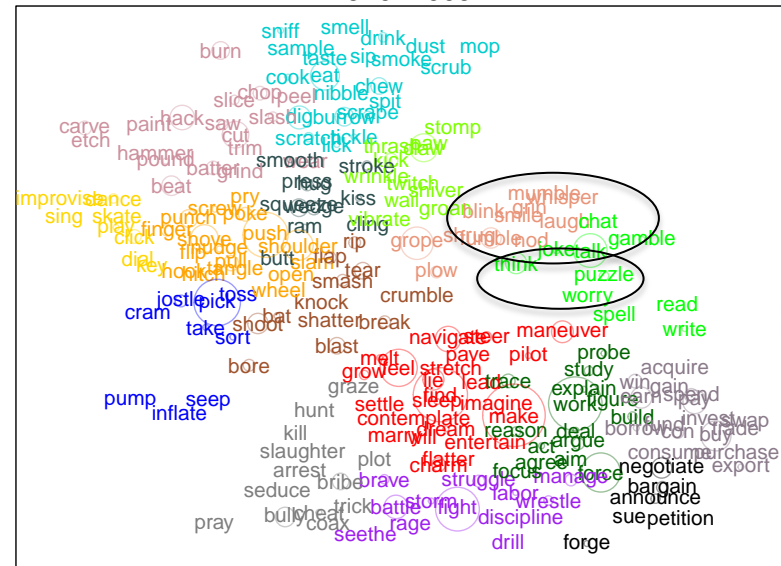
1880–1929



1930–1969



1970–2009



From period 3 onwards: social interaction (*chat, chatter, joke, kid, nod, quarrel, talk*), emotion (*grin, laugh, smile, shrug, laugh*), cognition (*brood, fret, puzzle, think, worry*)

The path-creation sense

- Many new verb classes refer to unusual ways to cause motion: interaction, commerce, cognition, etc.
- Most uses involve abstract, metaphorical motion, e.g.:

[T]hey talk about Uncle Paul having **bought his way into the Senate!**

By the time he was four he could **spell his way through his book** with only occasional pauses for breath.

I sit and watch [...], **grazing my way through a muffuletta.**

I saw Wallace Shawn [...] **lisping his way through a mournful monologue.**

The path-creation sense

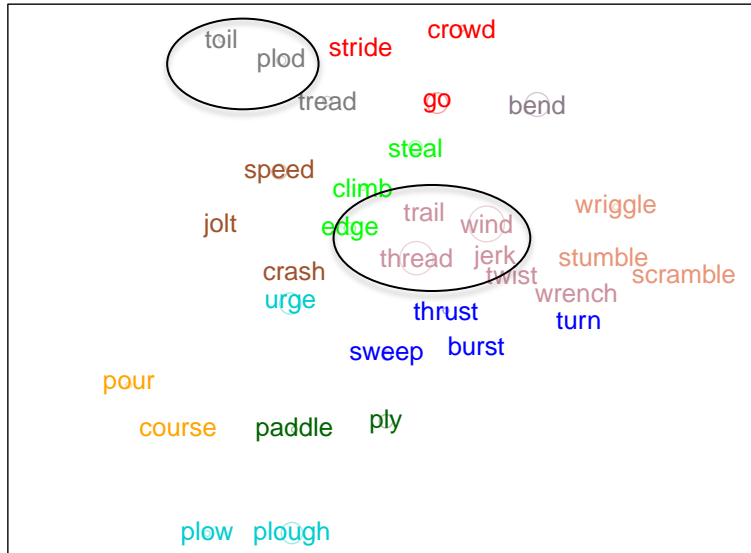
- The inclusion of classes of abstract verbs is likely to contribute to schema extension
 - The verb slot is more open
 - The motion component becomes more general
- Increase in both productivity and schematicity



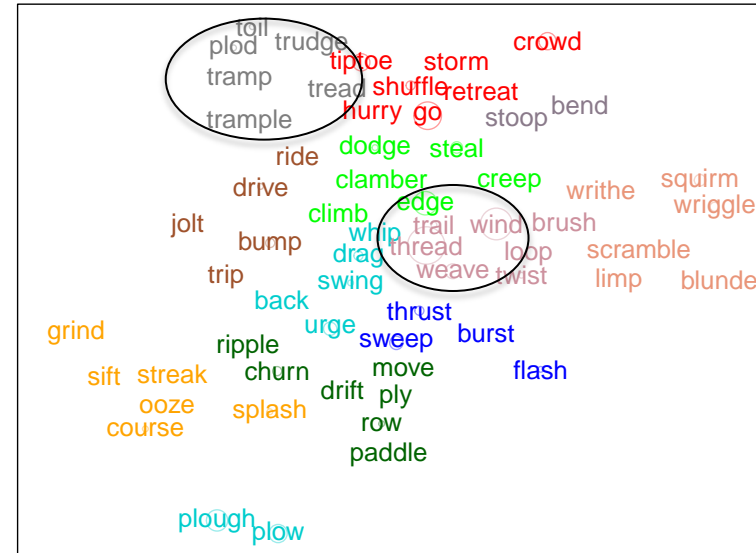
The manner sense



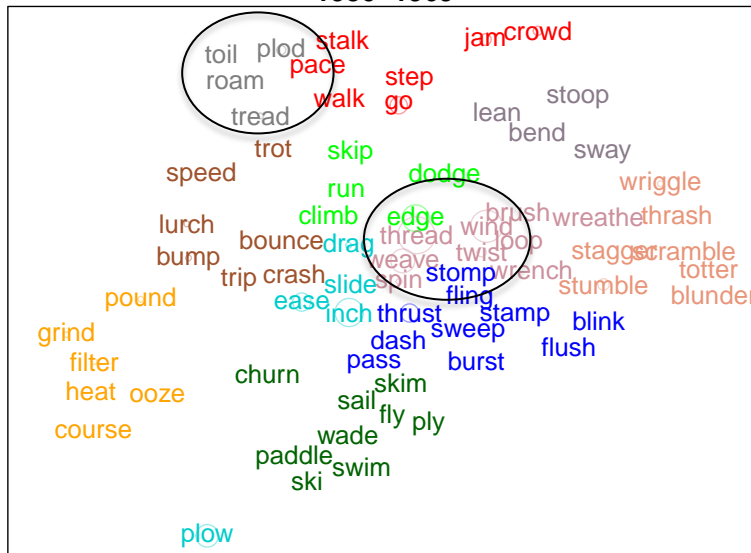
1830–1879



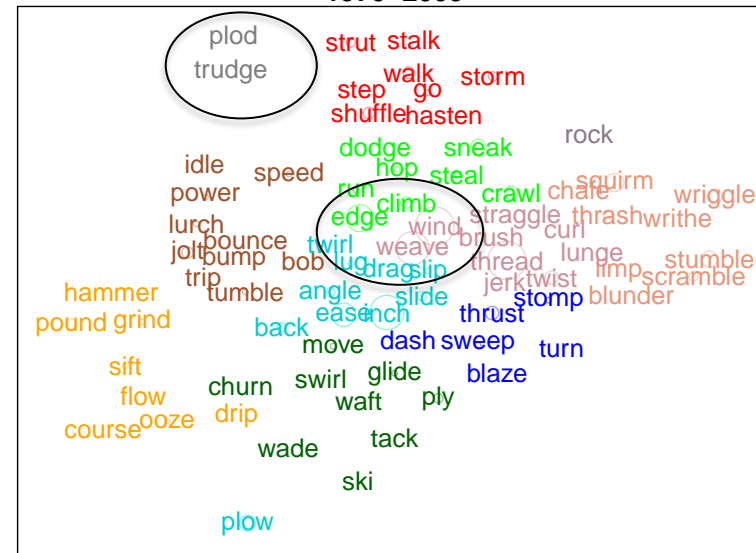
1880–1929



1930–1969

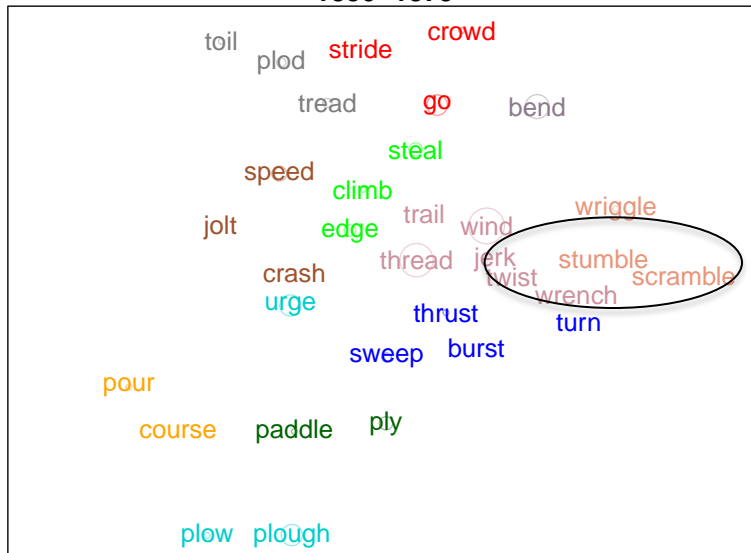


1970–2009

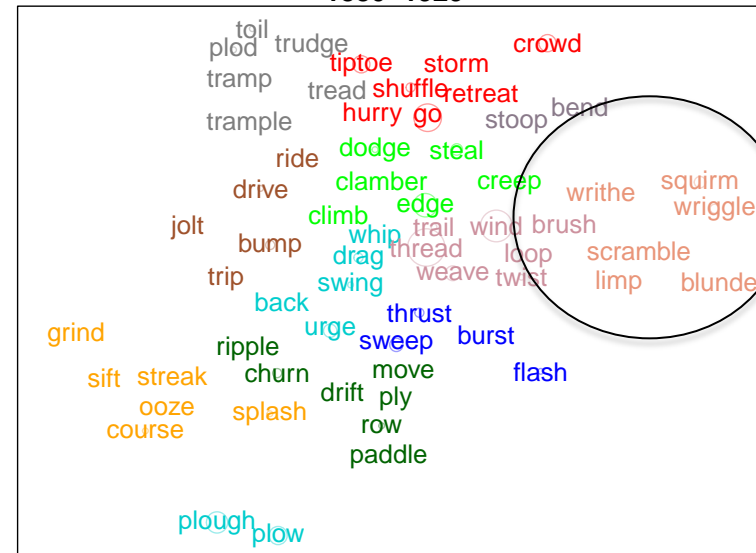


Verbs describing slow, indirect, or difficult motion: *thread*, *trial*, *weave*, *wind*, *plod*, *toil*, *tramp*, *trudge*.

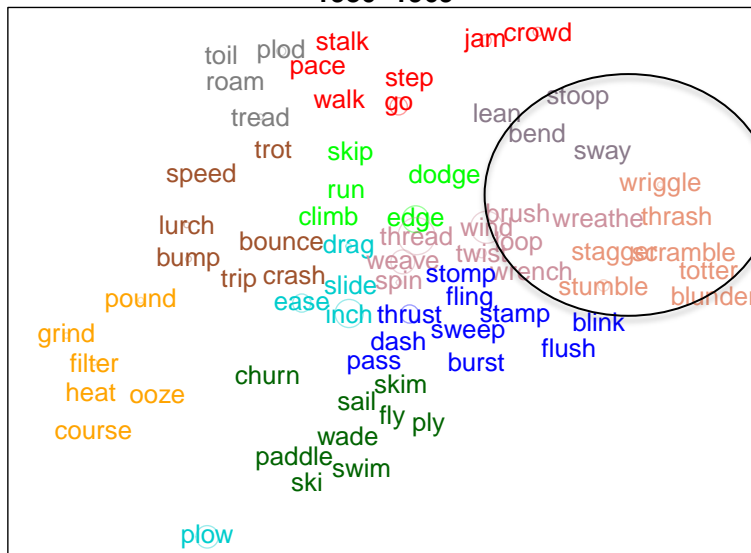
1830–1879



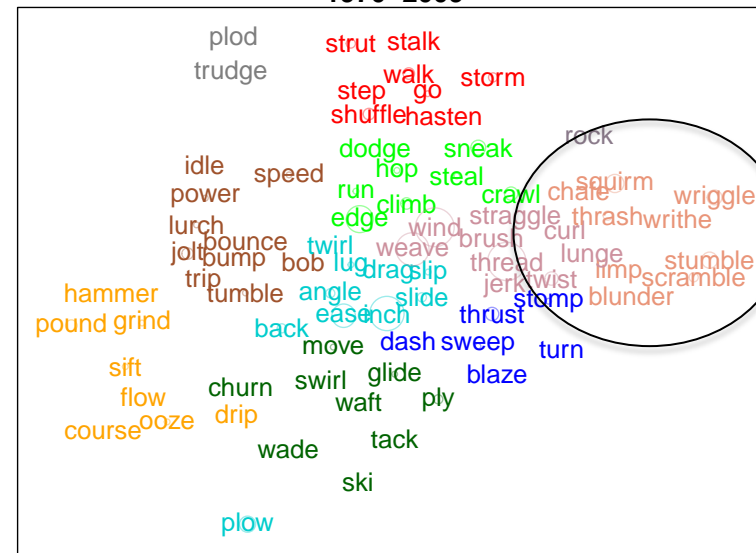
1880–1929



1930–1969



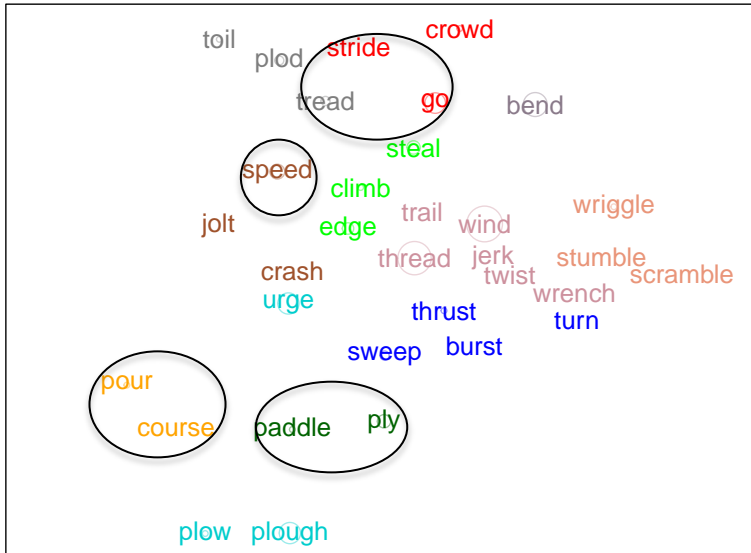
1970–2009



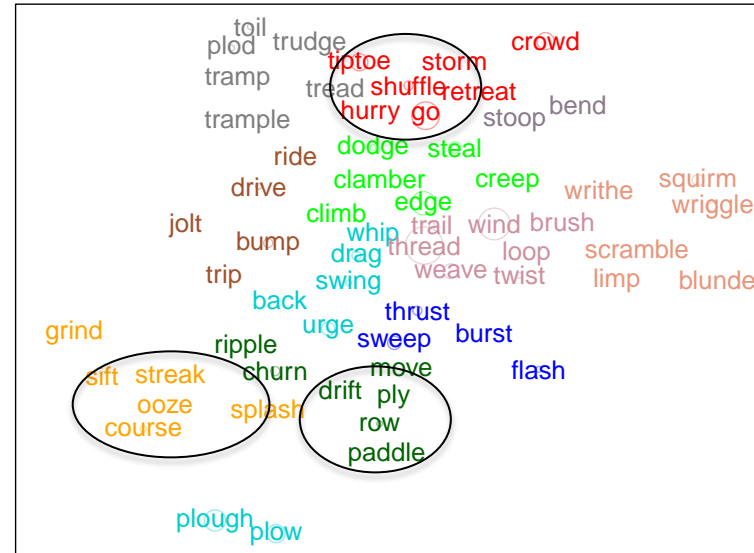
Clumsy or unsteady motion: *blunder*, *limp*, *scramble*, *stagger*, *stumble*, *totter*

Surrounded by verbs that encode body movements to facilitate motion: *bend*, *jerk*, *lean*, *lunge*, *stoop*, *thrash*, *twist*, *wrench*, *wriggle*, *writhe*

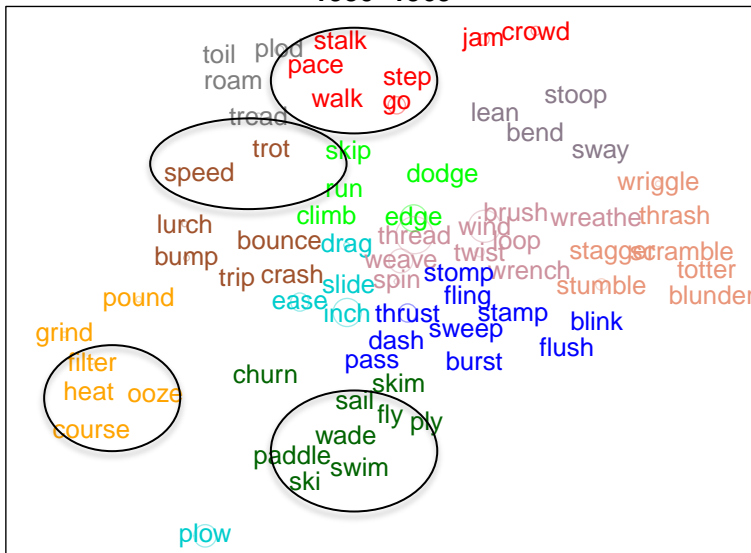
1830–1879



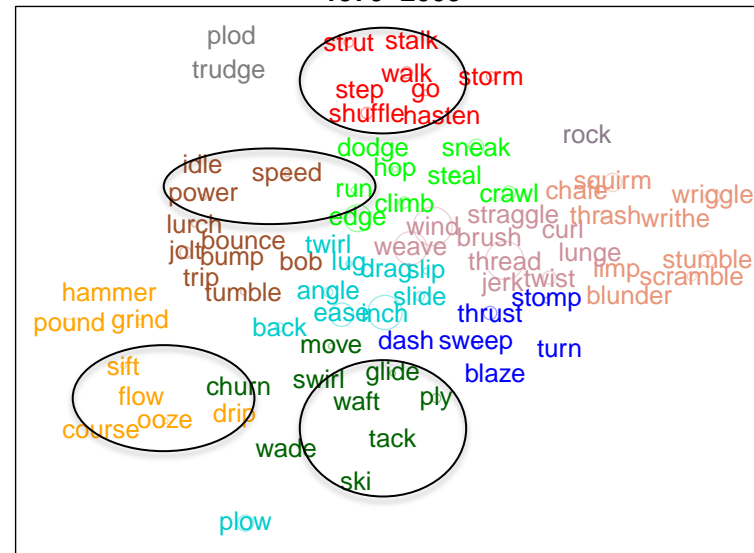
1880–1929



1930–1969



1970–2009



More ‘neutral’ manners of motion: walking (*stride, strut, tiptoe, walk, ..*), rapid motion (*power, run, speed, ..*), liquid motion (*course, drip, sift, ooze, ..*), vehicle/ theme (*fly, paddle, ply, sail, ski, ..*)

The manner sense

- Difficult motion = semantic 'core' of the construction (Goldberg 1995)
- Increase in diversity in later periods
- Non-difficult motion becomes more prominent
- Likely interpretation: increase in schematicity of the verb slot, from difficult motion to general manner of motion



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

One last example



The *many a* Noun construction

- A nominal construction: *many a* N (Hilpert & Perek 2015)
- Conveys plurality (=‘many Ns’)

Many a sailor has suffered from scurvy.

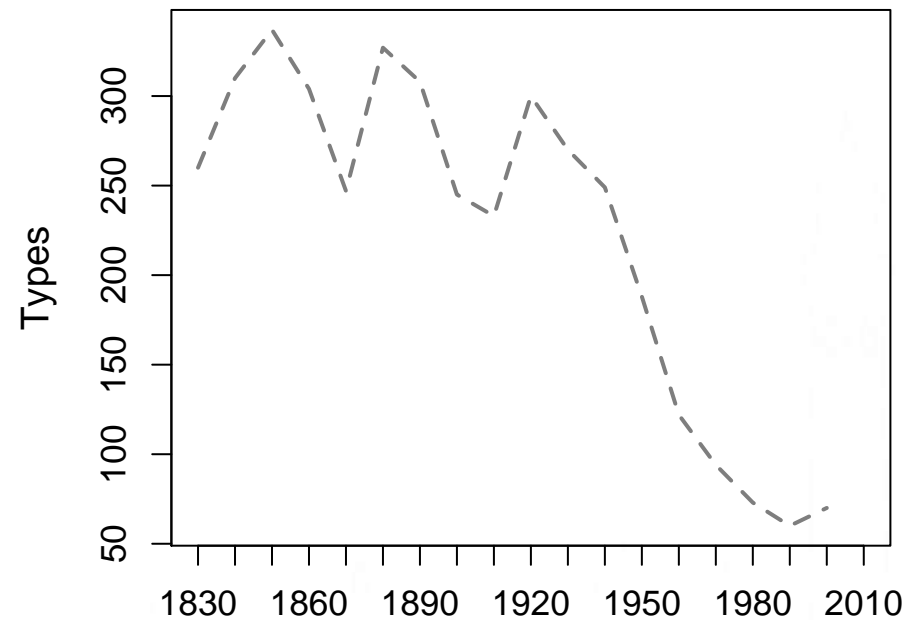
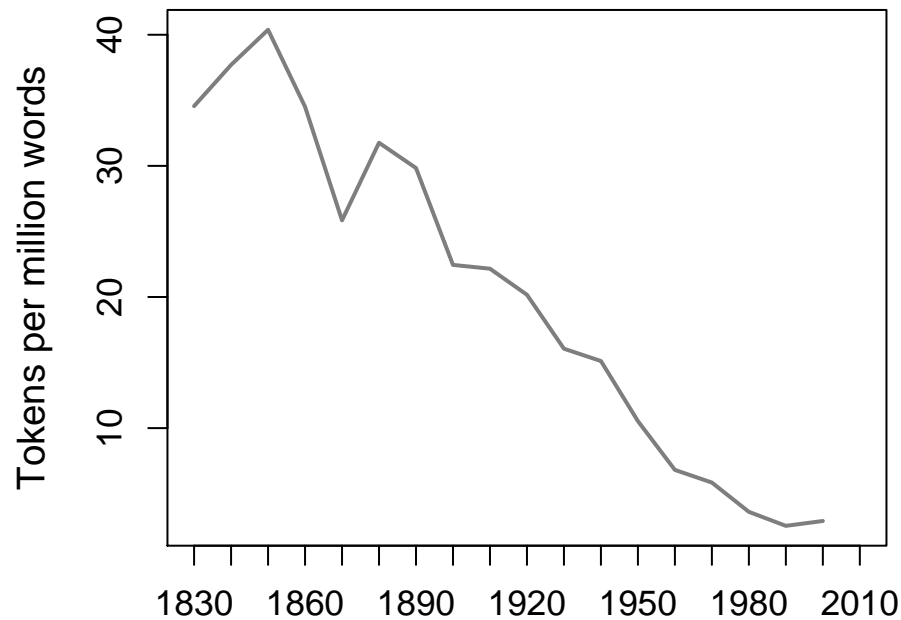
*[T]he volumes offer favorable contrast with **many a book** published in recent years.*

*For **many a day** the flowers have spread.*

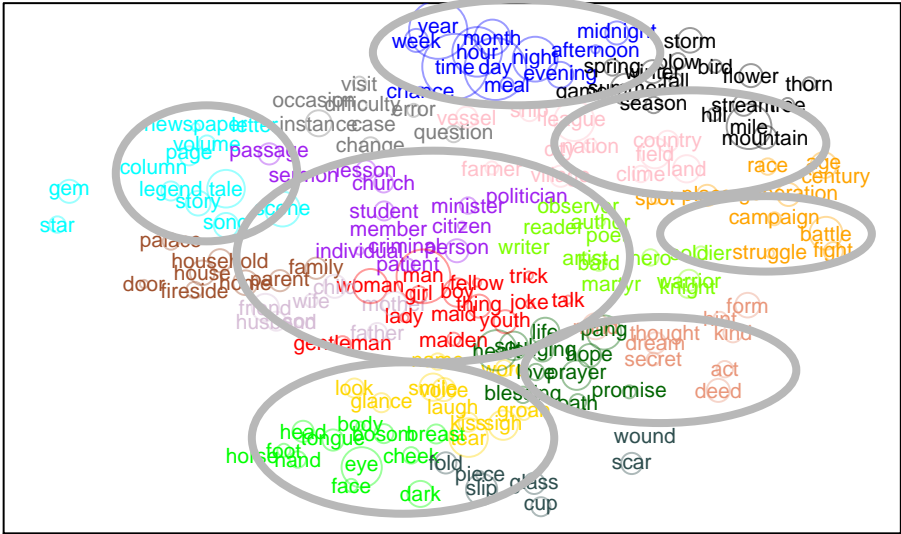
*The old meeting house has stood **many a storm**.*

The *many a* Noun construction

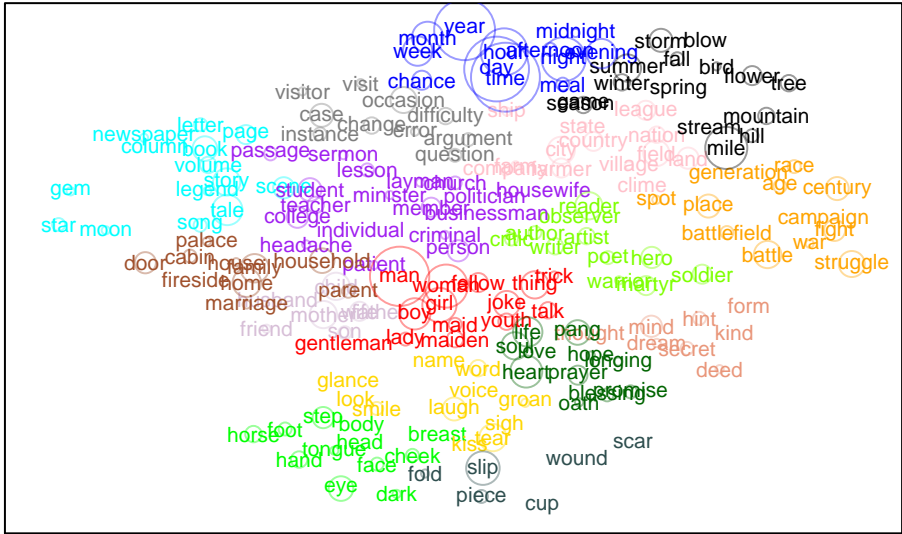
- An obsolescent construction, instead of a rising one
- 2015 different nouns: study limited to the top 200



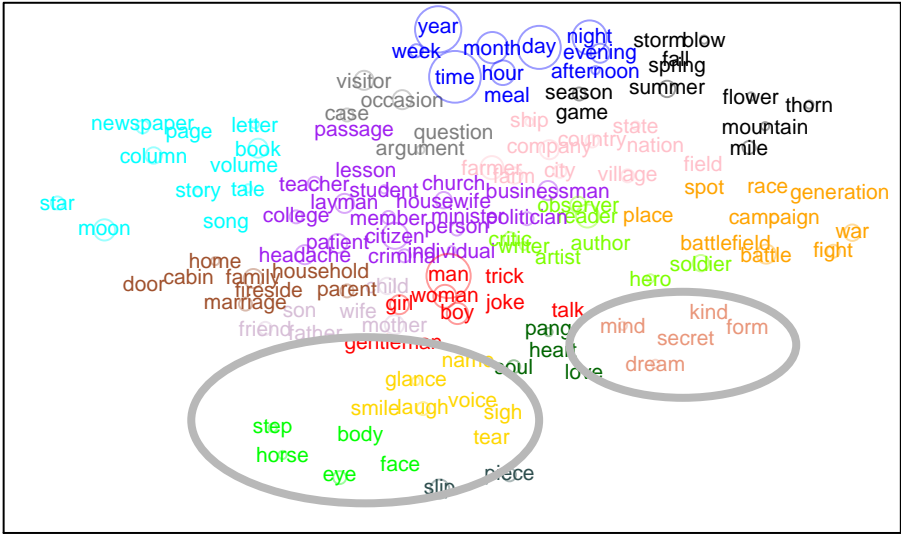
1830-1879



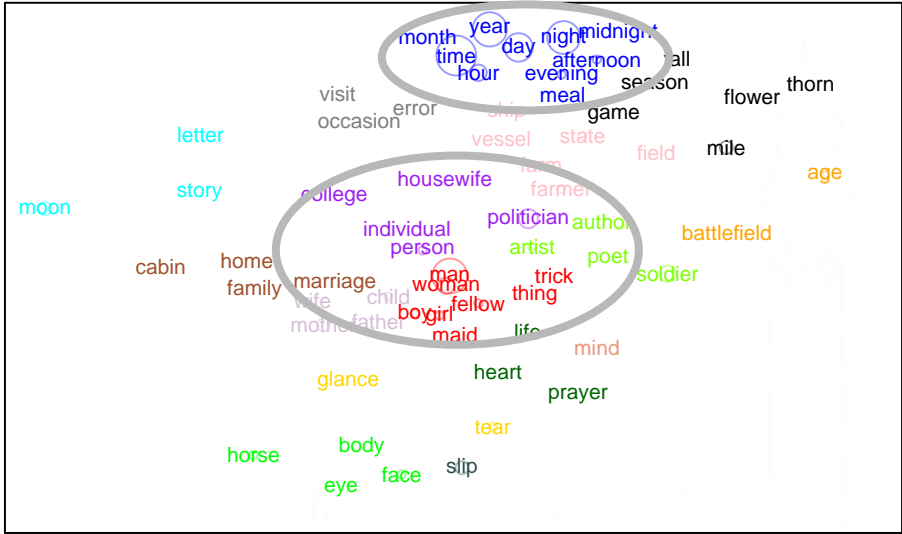
1880-1929



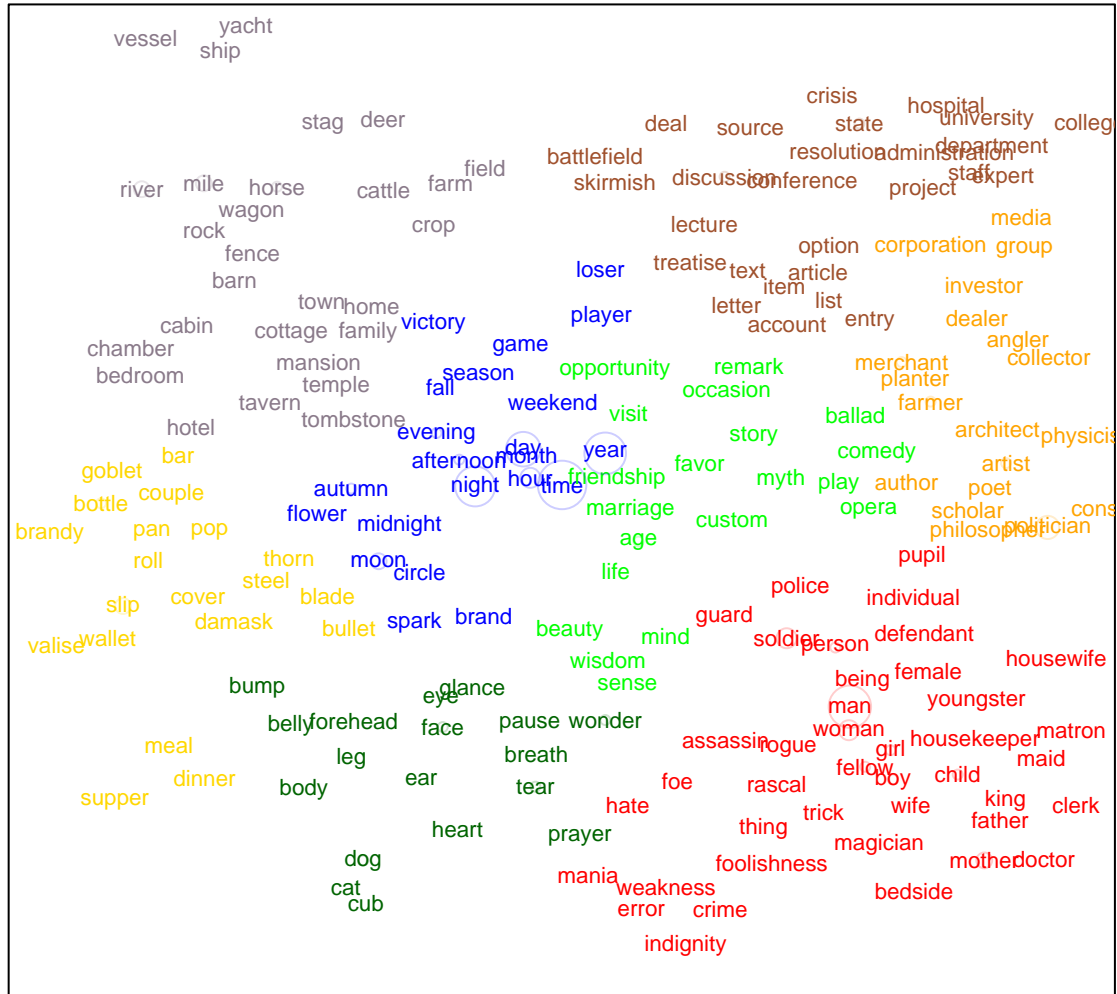
1930-1969



1970-2009



1970-2009, all types



The *many a* Noun construction

- Wide distribution, with a few domains of predilection
- Stable throughout the 19th century and early 20th
- Most groups recede in the mid-20th century
- Decrease in schematicity? Hard to tell
 - The remaining types are very spread out (openness)
 - The heyday of the construction is still recent: “legacy” effect?



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

Distributional period clustering



Periodization

- Distributional semantic plots are a useful tool to observe the development of constructions
- However, it is limited by the arbitrary division of the data
 - Periods of same length
 - Might not be consistent with regards to semantics
- Changes are assessed impressionistically rather than inferred quantitatively

Periodization

- The problem of periodization was first exposed by Gries & Hilpert (2008)
- They describe “variability-based neighbour clustering” (VNC) as a method for automatic periodization
- Variant of agglomerative clustering algorithm
 - Periods are grouped according to their similarity, following some pre-defined criteria
 - **Only time-adjacent period can be merged**

The VNC algorithm

- Starting point: data partitioned into “natural” time periods (years, decades, etc.)
 1. Look at all pairs of adjacent periods (e.g, 1830s-1840s, 1840s-1850s, etc.). Measure their similarity according to some quantifiable property/ies.
 2. Merge the two periods that are the most similar.
 3. Calculate the properties of the merger as the mean values of its constituent periods.
- Repeat until all periods have been merged.

VNC: an example

- VNC with one variable: frequency (Hilpert 2013: 36)



Hilpert, M. (2013). *Constructional Change in English. Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press

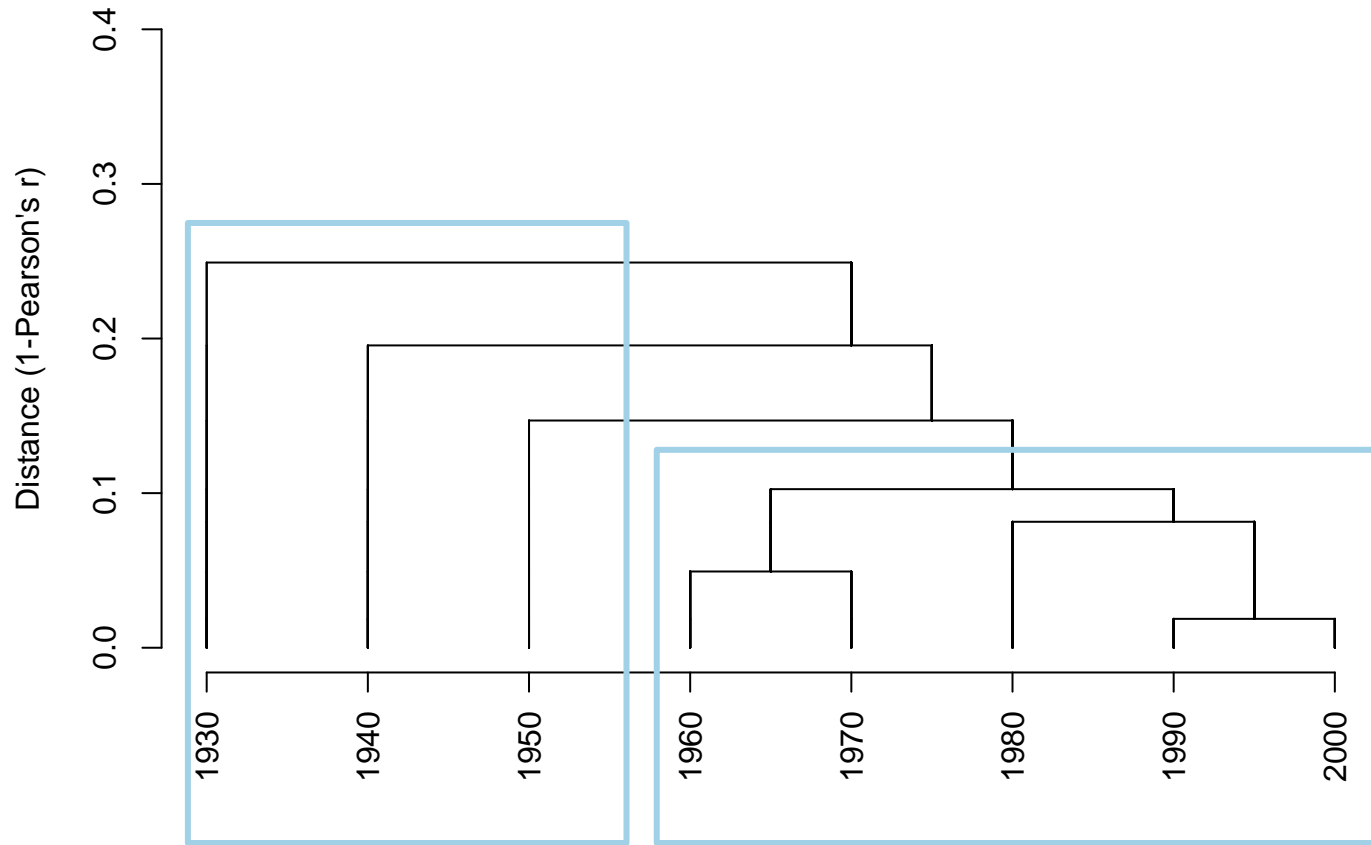
Distributional period clustering

- VNC on the basis of distributional semantic representations of time periods (Perek, in prep.)
- For each period, extract the semantic vector of each lexical item in the distribution from the DSM.
- Multiply each semantic vector by the frequency of occurrence of the lexical item in the construction.
- Add all these vectors: this is the period vector.

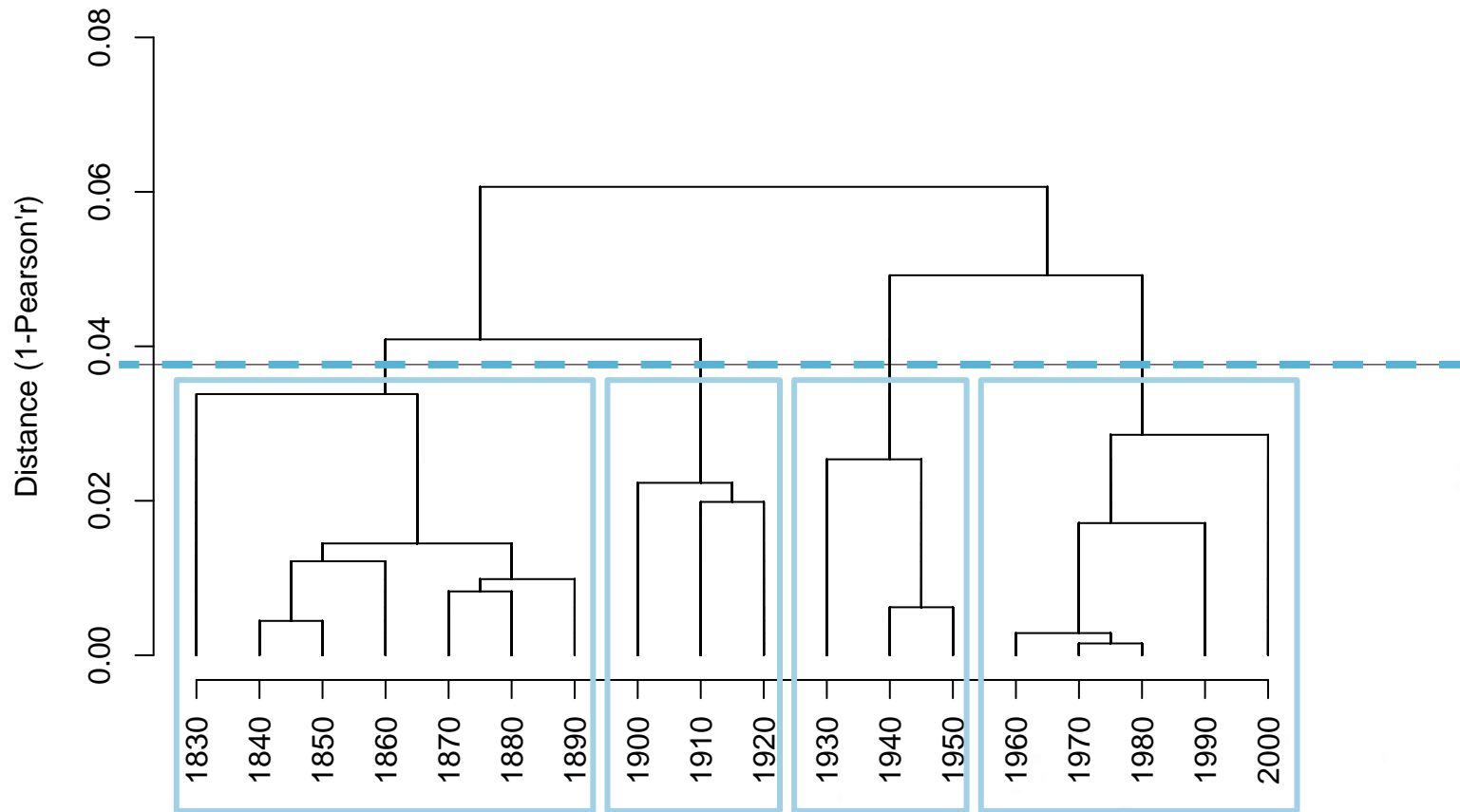
Distributional period clustering

- Similarity between periods is measured by Pearson's r
- The VNC algorithm is run on the period vectors
- The output reveals the semantic history of the construction:
 - Early mergers correspond to periods of semantic stability.
 - Late mergers of large clusters indicate semantic shifts.

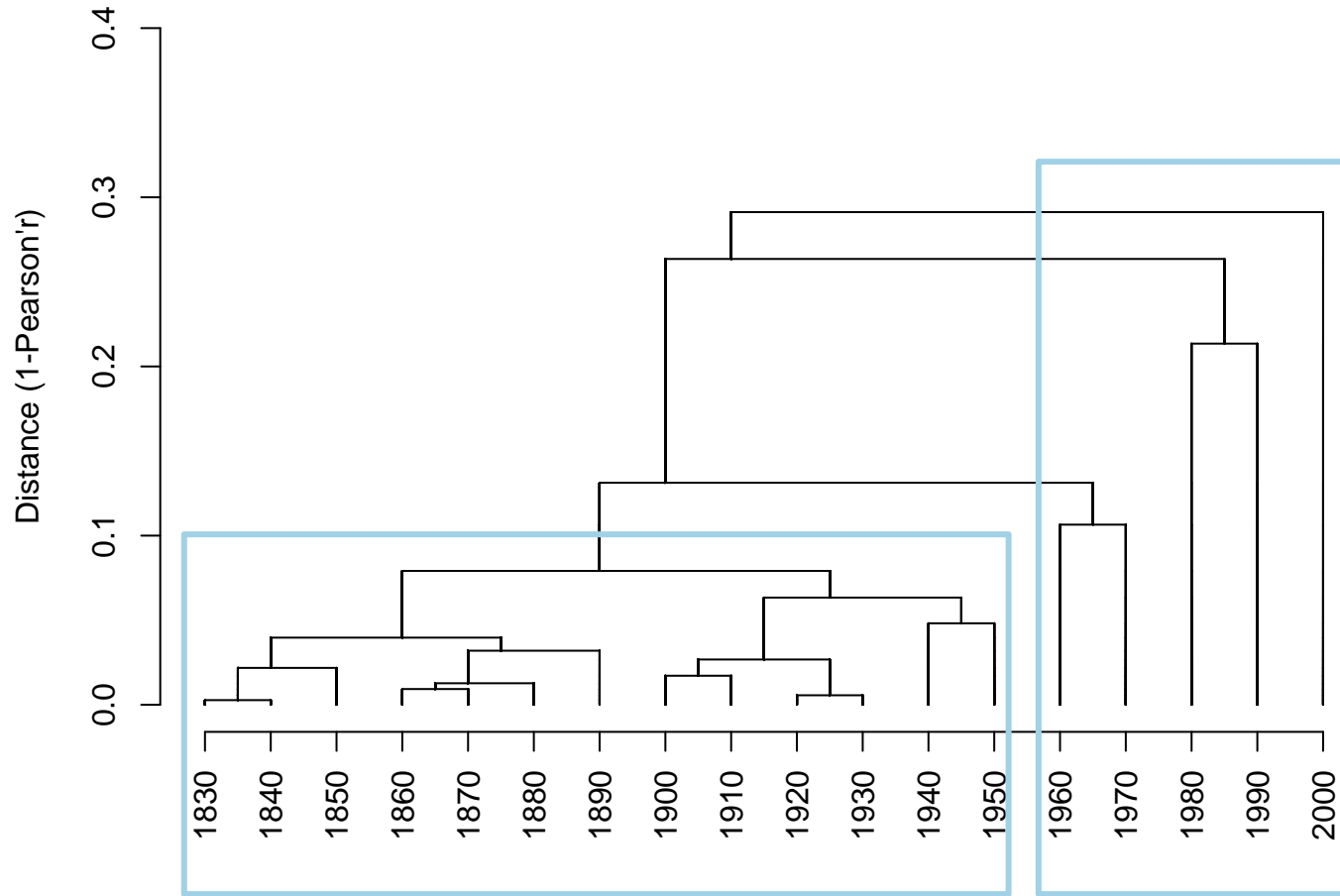
The *hell*-construction



The path-creation *way*-construction



Many a Noun



Summary

- The shapes of the dendrograms indicate different historical scenarios:
 - *Hell*-construction: gradually expanding construction
 - *Way*-construction: variations in distribution in a “fully grown” construction
 - *Many a Noun*: stable then gradually receding construction
- Did we really need distributional semantic information to make these observations?

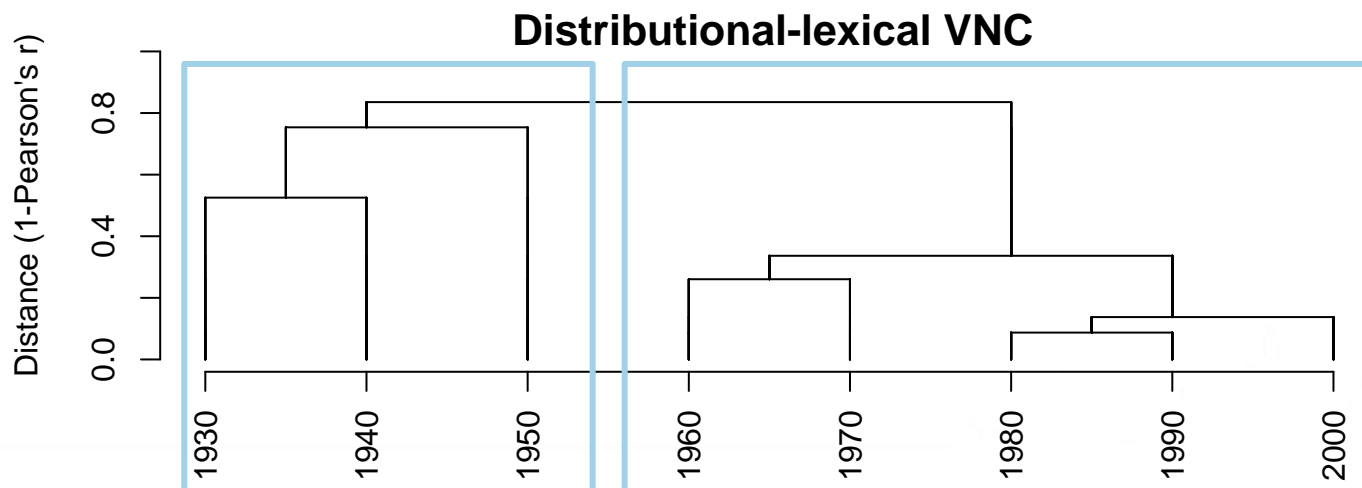
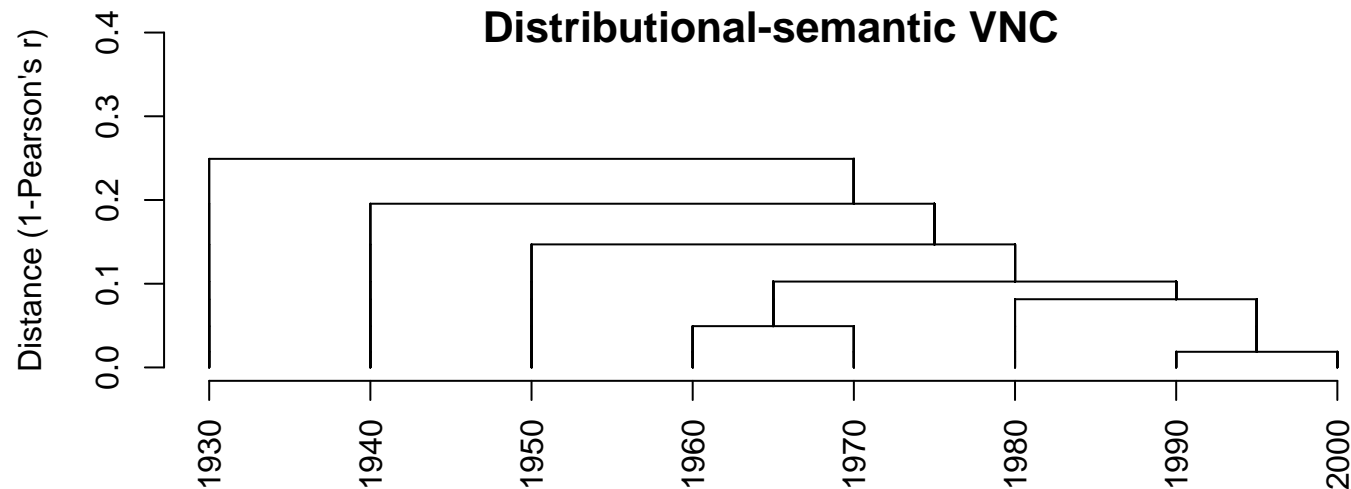
Comparison with “regular” VNC

- Comparison with VNC based on purely distributional-lexical information
 - The representation of each decade is a list of verb-frequency pairings

	1830s	1840s	1850s	...
<i>make</i>	184	167	210	...
<i>fight</i>	9	16	19	...
<i>dig</i>	0	2	2	...
...

- Distance between periods also calculated with Pearson’s r
- The resulting dendrograms have similar shapes, with some crucial differences

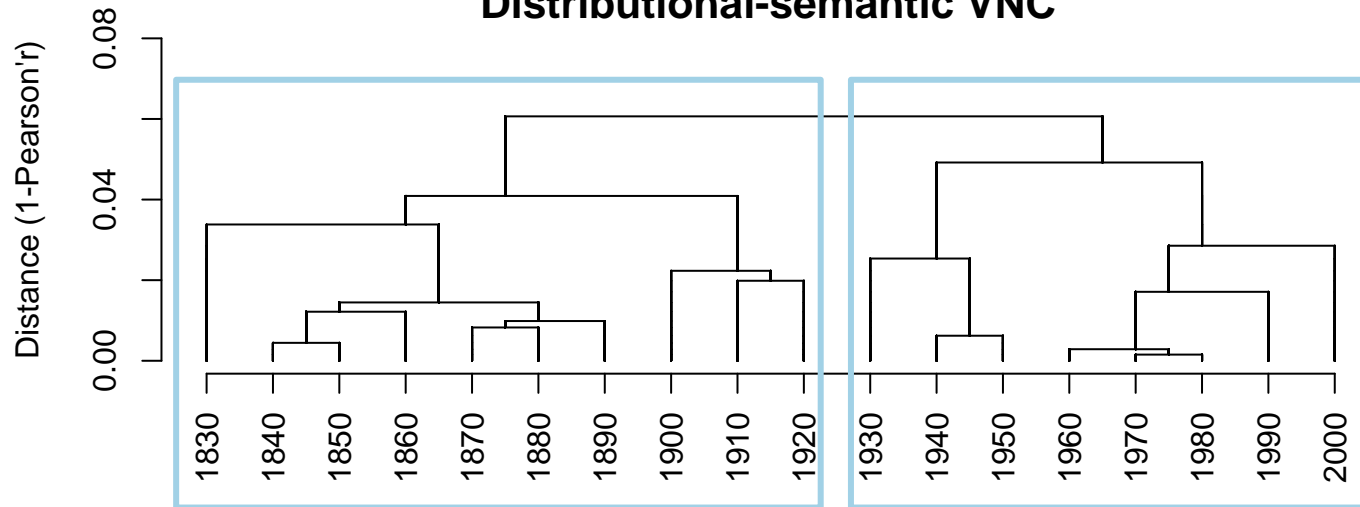
The *hell*-construction



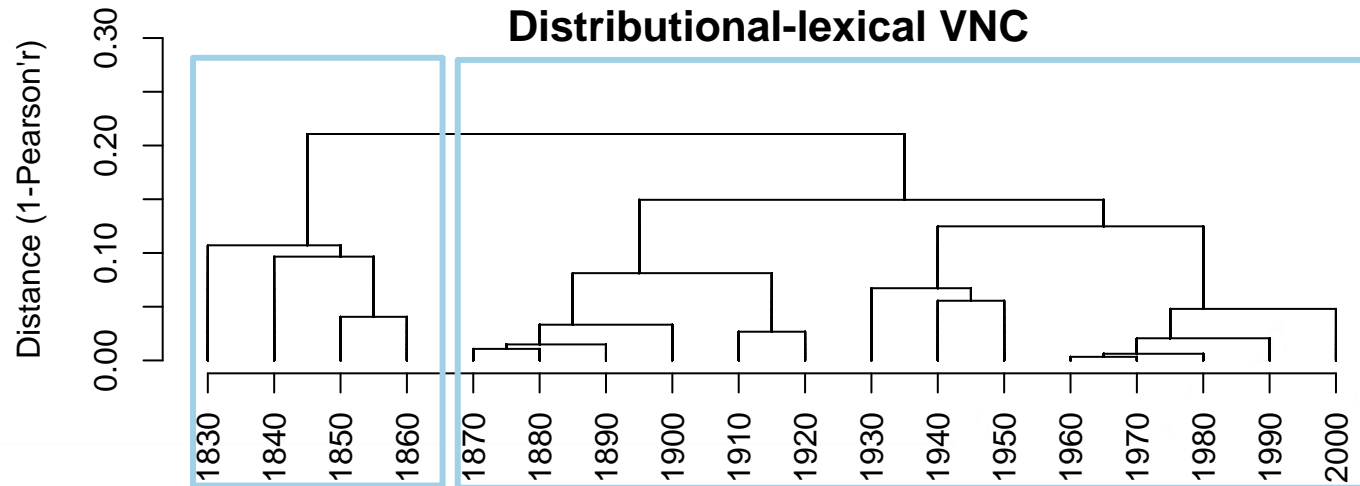
Probably due to an exceptional frequency drop of *beat* and *scare* (50%) in 1950

The *way*-construction

Distributional-semantic VNC

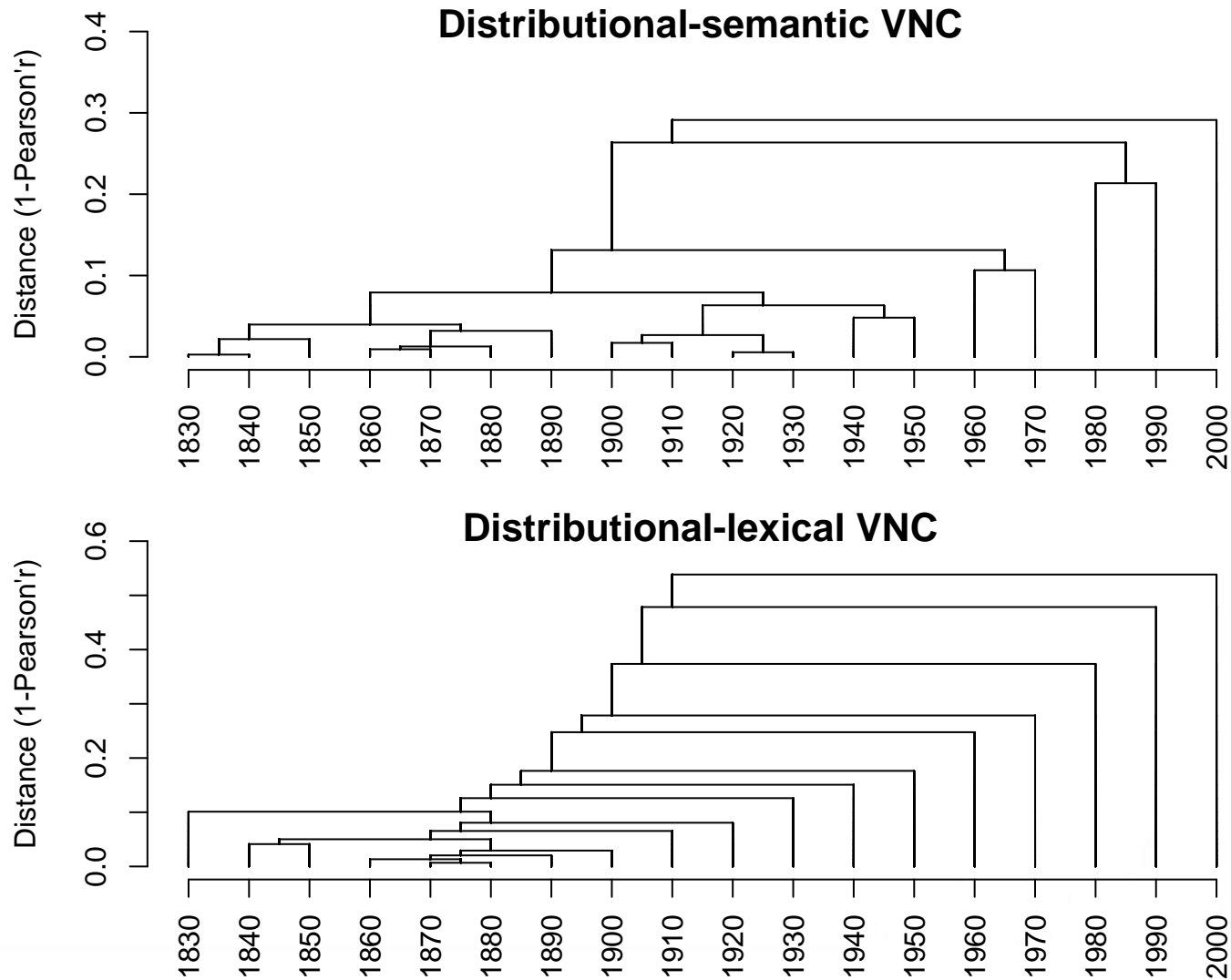


Distributional-lexical VNC



Probably due to the decline of high-frequency *take* after the 1860s

Many a Noun



No clear moment when the distribution starts changing: probably due to the fact that the distribution is centred on several high-frequency members at all times

Summary

- ❑ Distributional period clustering provide precise quantitative measurement to impressionistic observations
- ❑ Helps modelling different kinds of semantic change with dendrograms
- ❑ Less sensitive to distributional quirks that do not have a semantic basis
- ❑ Represents a step forward from regular VNC

Conclusion

- Distributional semantics is a very promising approach (not that this audience needs convincing...)
- Turns the informal notion of meaning into a quantified representation
- Appropriate for the study of constructional change
 - Gives a semantic interpretation to changes in productivity
 - Makes it possible to inform hypotheses about schematicity

Prospects for future research

- Look at the meaning of the construction itself
 - Cf. advances in distributional approaches to compositional semantics
 - Compare distributional semantics of lexemes vs. lexemes in constructions
- Control for semantic change of lexical meaning
 - I.e., by using different distributional representations of the same lexeme in different time periods
 - Especially important for studying earlier time periods



UNIVERSITY OF
BIRMINGHAM

COLLEGE OF
ARTS AND LAW

Grazie per la vostra attenzione!

Thanks for your attention!

f.b.perek@bham.ac.uk

www.fperek.net

