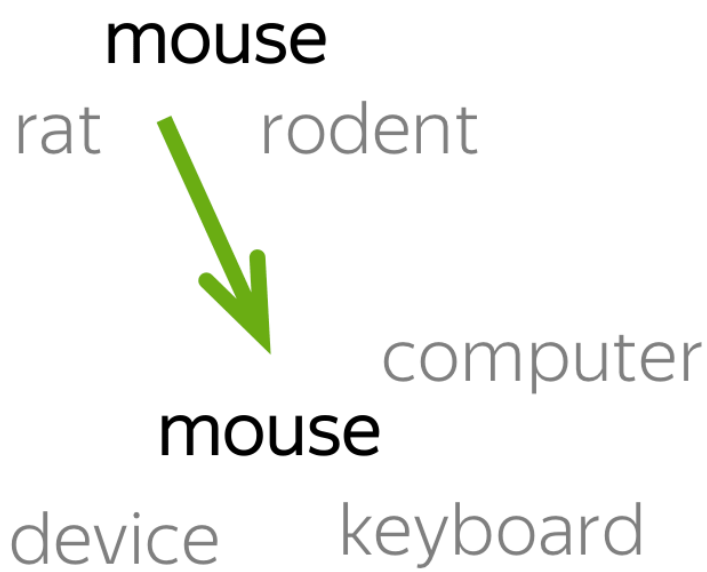
Grammar and Meaning: Analysing the topology of diachronic word embeddings

Katharina Altrichter (6 CP)
12.12.2023

Motivation and background

Motivation


majestic
full of awe
awful
solemn
terrible
awful
horrible
awful
wonderful
awfully
weird


mouse
rat
rodent
computer
mouse
device
keyboard

Motivation

Can we also use word embeddings to analyze changes in grammatical use?

- **Function words**
  - Articles (a, the)
  - Determiners (this, that)
  - Conjunctions (and, but)
  - Prepositions (in, on) …
- **Polyfunctional words**
  - Gerunds (“swimming is my favorite hobby)
  - Participles (“the painted walls”)
  - Interrogative words (what, why) …

Background

- Authors: Bizzoni et al., Universität des Saarlandes
- Published 2019 in *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, Association for Computational Linguistics

Aim of the paper:

- Topological analysis of diachronic word embeddings
- Focus: “-ing”-verbs
  - Grammatical patterns
  - Syntagmatic properties

Word embeddings

Dataset

- Dataset: Royal Society Corpus (RSC) [1]
  - publications of the Philosophical Transactions and Proceedings of the Royal Society of London from 1665 to 1869
  - 32 million tokens, 10000 documents
  - metadata and linguistic annotations (lemma, sentence boundaries, etc)
- Interesting for analyzing scientific writing

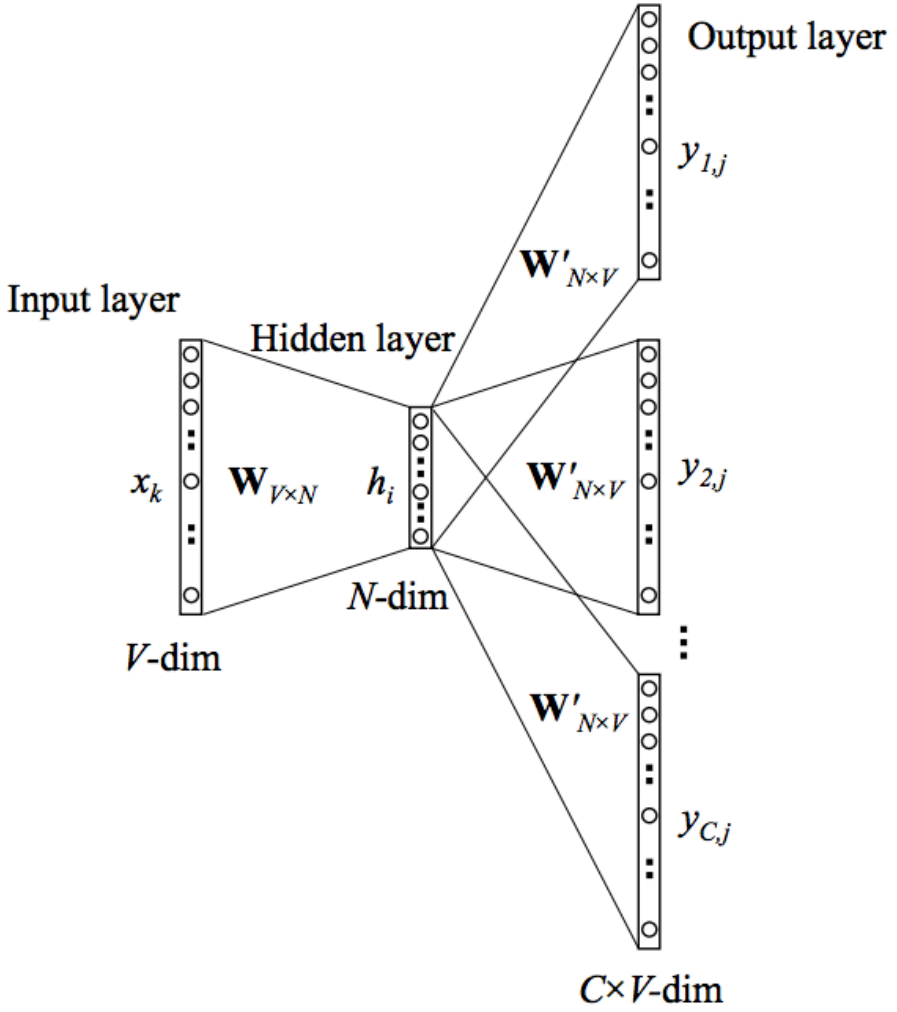
decade	tokens	lemma	sentences
1660-69	455,259	369,718	10,860
1670-79	831,190	687,285	17,957
1680-89	573,018	466,795	13,230
1690-99	723,389	581,821	17,886
1700-09	780,721	615,770	23,338
1710-19	489,857	383,186	17,510
1720-29	538,145	427,016	12,499
1730-39	599,977	473,164	16,444
1740-49	1,006,093	804,523	26,673
1750-59	1,179,112	919,169	34,162
1760-69	972,672	734,938	27,506
1770-79	1,501,388	1,146,489	41,412
1780-89	1,354,124	1,052,006	37,082
1790-99	1,335,484	1,043,913	36,727
1800-09	1,615,564	1,298,978	45,666
1810-19	1,446,900	1,136,581	42,998
1820-29	1,408,473	1,064,613	43,701
1830-39	2,613,486	2,035,107	81,500
1840-49	2,028,140	1,565,654	70,745
1850-59	4,610,380	3,585,299	146,085
1860-69	5,889,353	4,474,432	202,488
total	31,952,725	24,866,457	966,469

[1] Open source, available at http://fedora.clarin-d.uni-saarland.de/rsc_v4/

Computation of word embeddings
Methodology

Structured Skip-gram method [2]

- Words represented as vectors
- Prediction of context given a target word
  - compute probability for each neighbor word in a context window
  - Dot product between target word vector and context word vector + application of softmax function
- Iterative training with positive (context words) and negative (other words) examples



[2] Wang Ling et al, 2015.

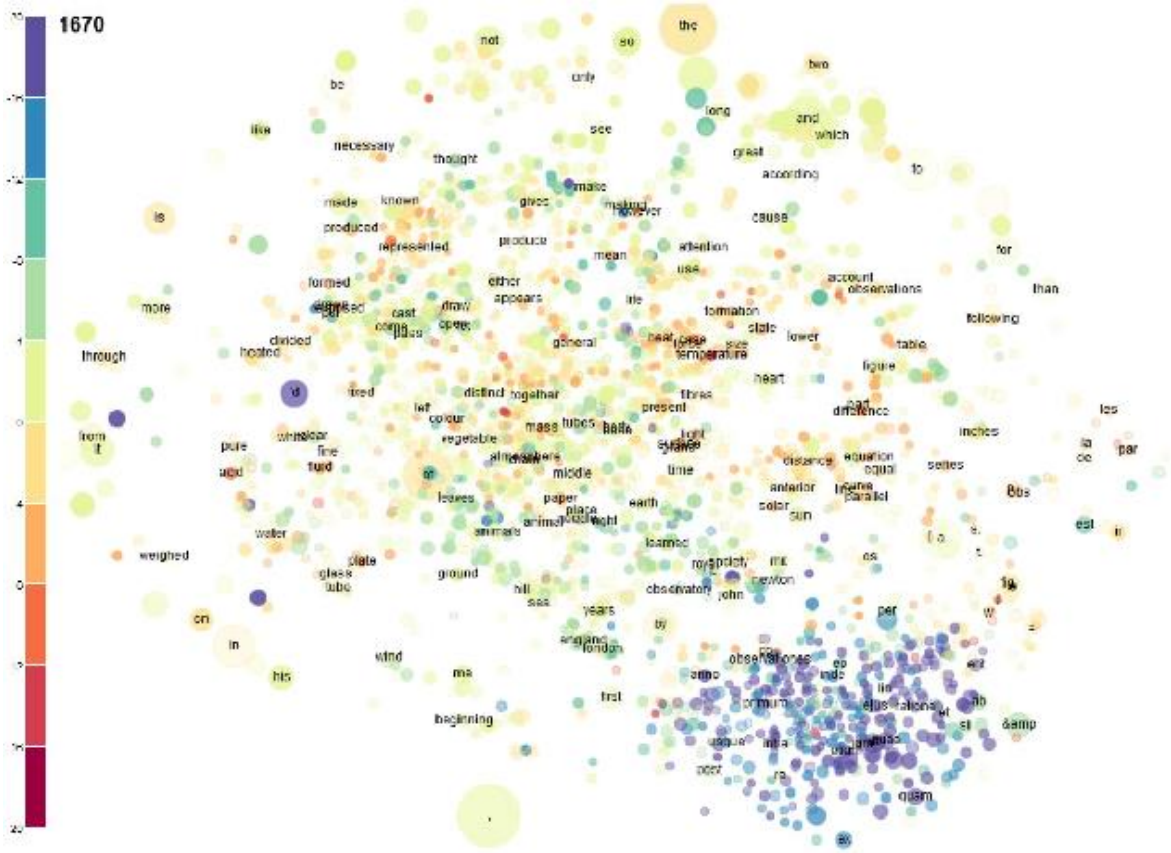
Computation of word embeddings
Methodology

Structured Skip-gram method [2]

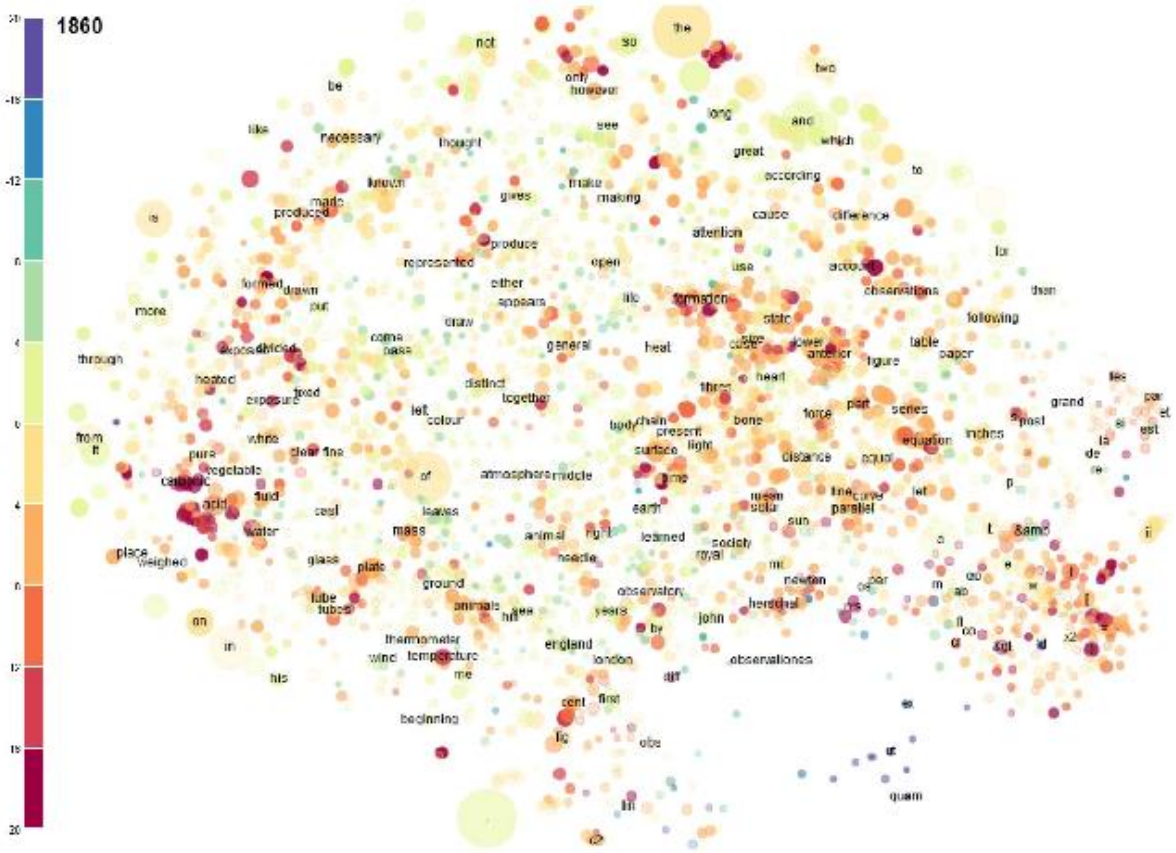
- Takes word order into account → **capture grammatical structures**
- No multiword expressions → **model becomes as agnostic to corpus as possible**
- Embeddings calculated for every decade of the dataset
- Vocabulary: 117.165 100-dimensional points
- Dimensionality reduction using t-distributed stochastic neighbor embedding [3]

[2] Wang Ling et al, 2015.
[3] Laurens van der Maaten and Geoffrey Hinton. 2008.

Computation of word embeddings
Visualization in space



(a) 1670-1679



(b) 1860-1869

color: increasing (red) and decreasing (green) frequency, size of the bubbles: relative frequency

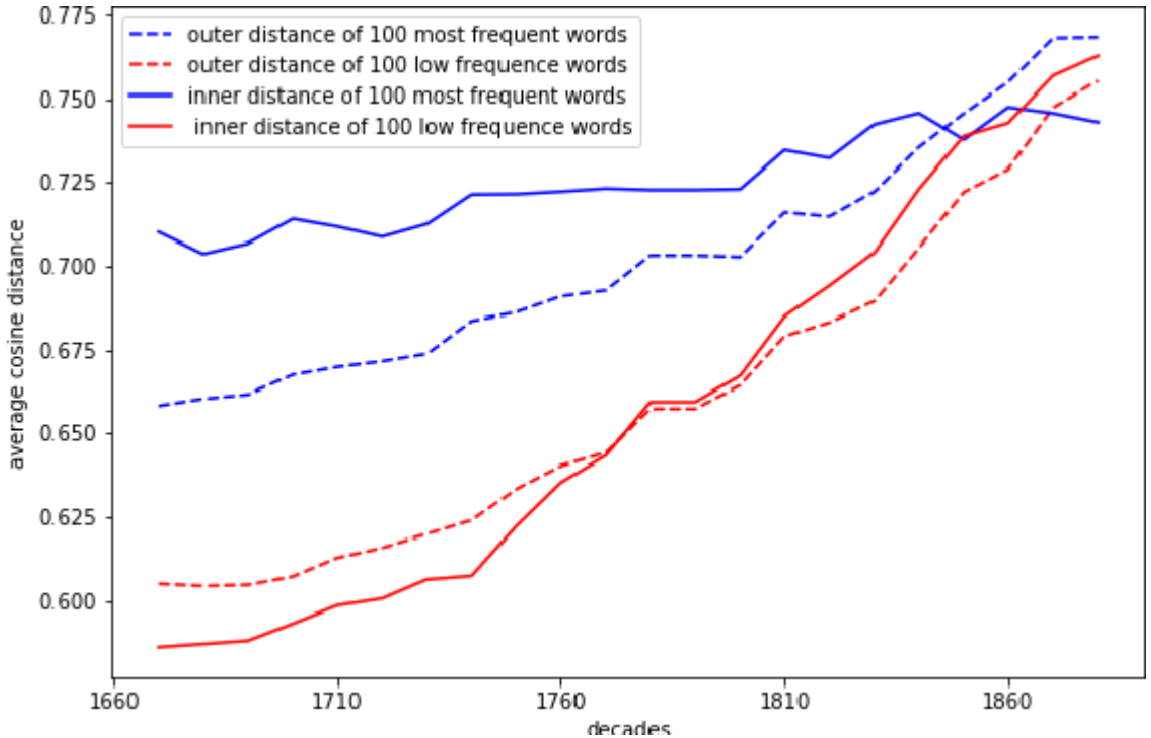
Topological analysis

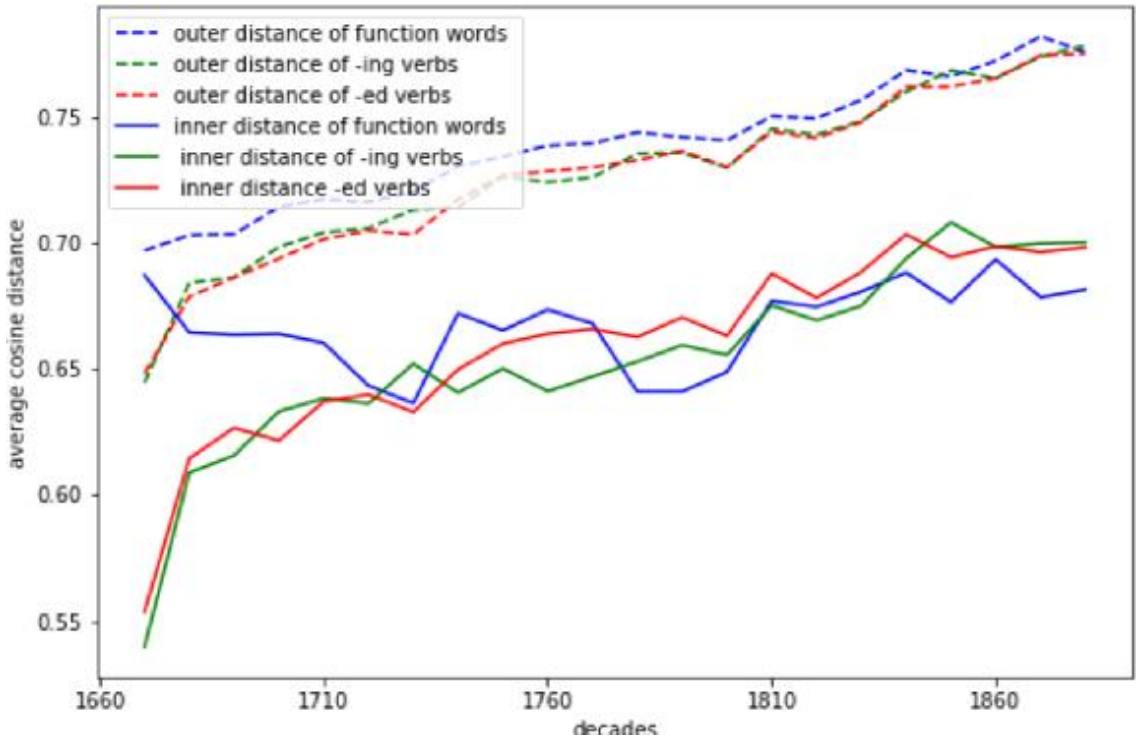
Topological analysis
Methodology

Analysis of the expansion of space

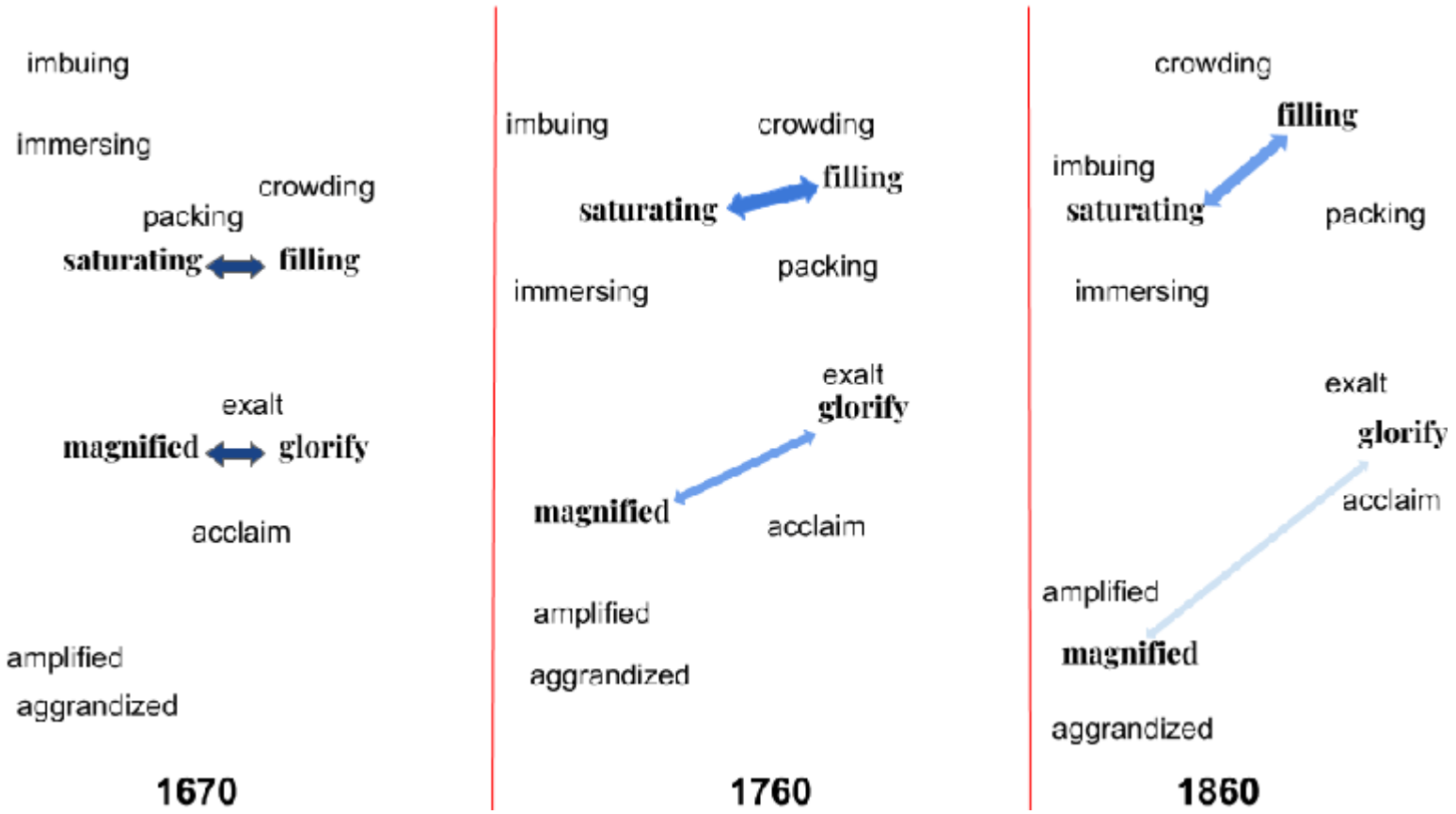
- Measure distances of lexical, function and polyfunctional words (cosine distance)
  - **Average distance** to the rest of the vocabulary
  - **Inner distance**: degree of similarity within a group of words
  - **Outer distance**: degree of isolation of one group to the rest of the vocabulary

Topological analysis
Inner and outer distances





Topological analysis
Semantic diversification



Topological analysis
Key Findings

- **Whole dataset:** all distances increase over time

  → Low-frequency words have a **highly specialized meaning**

- **Function words:** outer distance increases, inner distance remains stable

  → **isolation** but **no contextual specialization**

- **Polyfunctional words**: average inner distance lower than average outer distance

  → Changes most likely due to their **lexical rather than their grammatical side**

“-ing”-forms

Grammatical change

“-ing”-Forms: Grammatical change
Methodology

Diachronic clustering of “-ing”-forms

- Extraction of all “-ing”-words that are either **gerunds or participles**
- Formation of word clusters by analyzing **nearest neighbors**
- **Dynamic threshold**: average distance of nearest neighbors for given decade + 0.05

  → observations are independent from the general expansion of the space
- Later: application of **clustering algorithms**
  - Affinity Propagation [4]
  - DBSCAN [5]
  - Minibatch K-Means [6]

[4] Brendan J Frey and Delbert Dueck, 2007.
[5] Thanh N. Tran et al, 2013.
[6] David Sculley, 2010.

“-ing”-Forms: Grammatical change
Key Findings

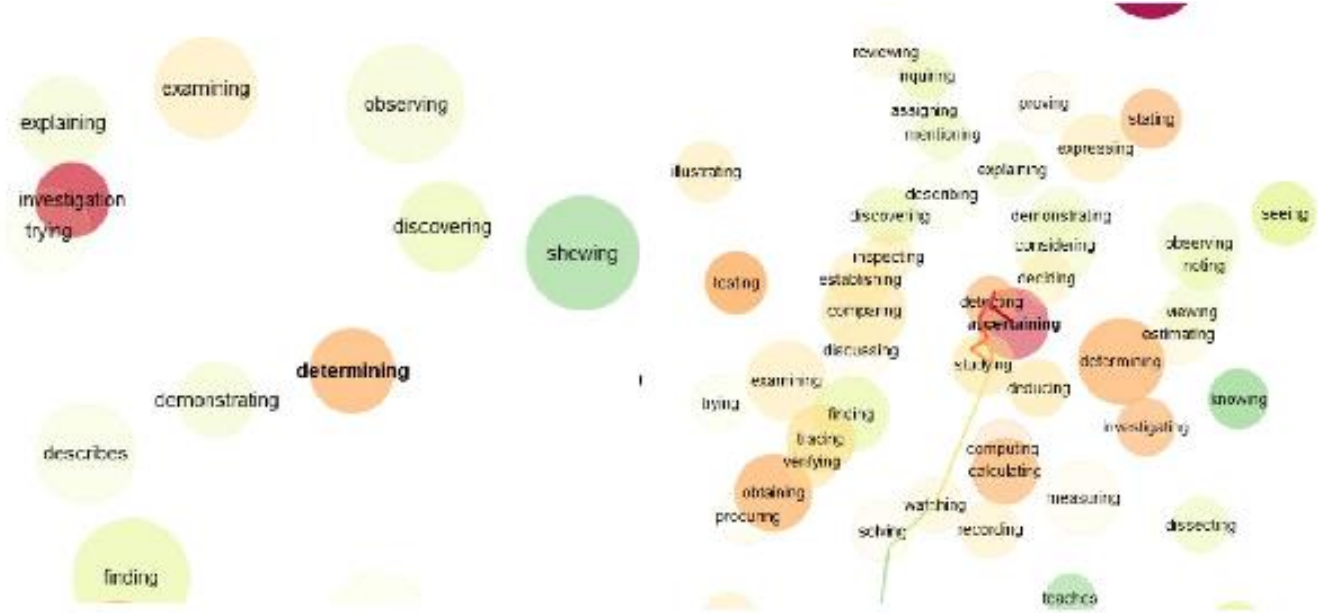
After first clustering:

- Density of “-ing”-clusters diminishes over time

  → possible specialization of words

- Words at the center of clusters can be grouped into three categories:
  - Academic verbs (i.e. *examining*)
  - Motion verbs (i.e. *falling)*
  - Change-of-state verbs (i.e. *warming)*

“-ing”-Forms: Grammatical change
Cluster formation of academic verbs



(a) 1670s (b) 1860s

“-ing”-Forms: Grammatical change
Key Findings

Application of clustering algorithms:

- **Affinity Propagation**
  - No need for predetermined number of centroids
  - Tendency towards many smaller clusters
  - Increasing number of clusters over time

- **DBSCAN**
  - No predetermined number of centroids either
  - Fixed threshold, fixed minimum number of neighbors
  - Fewer, but still increasing number of clusters over time

“-ing”-Forms: Grammatical change
Key Findings

Application of clustering algorithms:

- **Minibatch K-Means**
  - Requires predetermined number of centroids → 3 categories
  - Distance between three centroids increases over time
  - Results very similar to prior observations
  - Beginning of 19th century: growing distributional differences

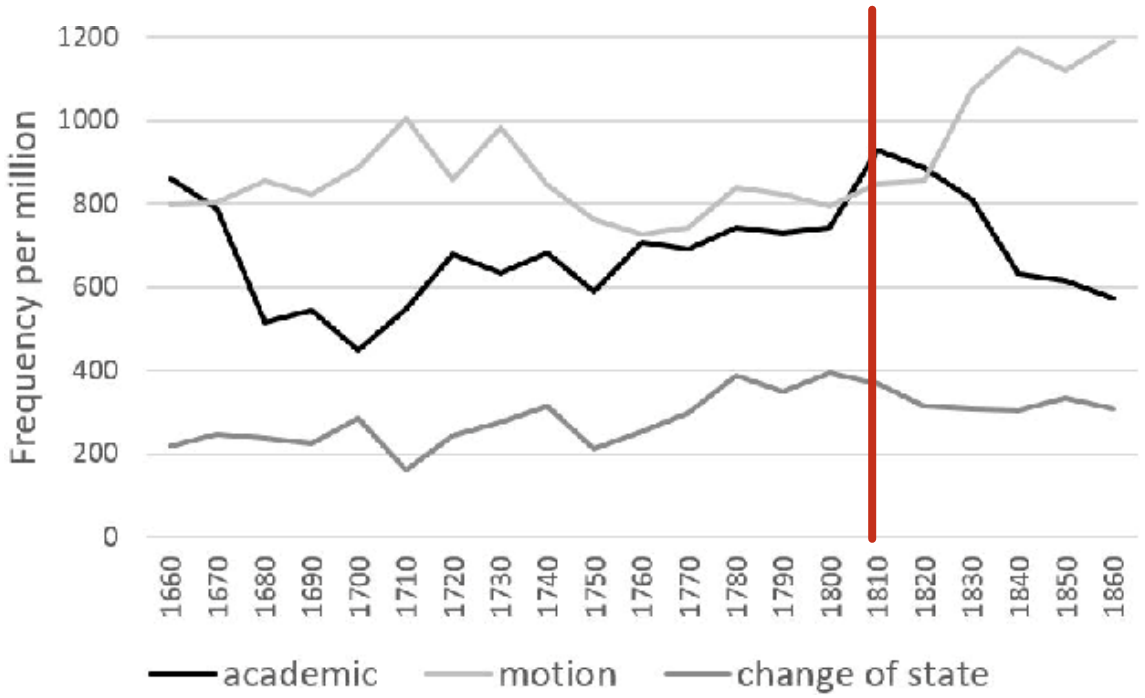
“-ing”-Forms: Grammatical change

Clustering

Decade	Affinity Propagation (AP)	DBSCAN	Minibatch KMeans
1660	*Extending, reaching, proceeding.*	*Abounding, according, adding.*	*Detaching, wetting, squeezing.*
	Crying, coughing, sweating.	*Whiting, widening, willing.*	*Verifying, deciding, transferring.*
	Shading, scattering, tracing.		*Playing, retiring, accumulating.*
1760	*Pricking, stimulating, snapping.*	*Abating, abounding, abstracting.*	*Arranging, attaching, immersing.*
	Following, lowing, preceding.	*Lessening.*	*Arranging, studying, illustrating.*
	Informing, troubling, acquainting.	*Deducting, subtracting, weighing.*	*Interlacing, arranging, transforming.*
1860	*Nourishing, binding, imbibing.*	*Abounding, absorbing, abstracting.*	*Determining, establishing, studying.*
	Snapping, widening, pricking.	*Integrating, introducing, putting.*	*Passing, extending, running.*
	Stimulating, promoting, biting.	*Arching, running, sweeping.*	*Purifying, agitating, warming.*

“-ing”-Forms: Grammatical change
Frequency distribution

- Particularly interesting: time period before 1810 compared to time period after 1840

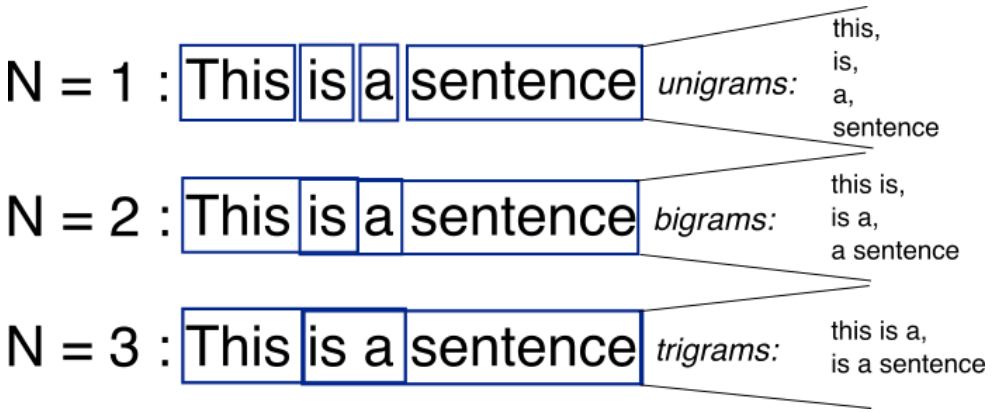


“-ing”-forms

Syntagmatic change

“-ing”-Forms: Syntagmatic change
Methodology

- Check if “-ing”-forms differ in terms of grammatical classes (gerund vs. participle)
- Consider top 30 verbs derived from previous clustering
- Extract part-of-speech ngrams of “-ing”-expressions



“-ing”-Forms: Syntagmatic change
Methodology

- Measure distinctiveness of syntagmatic contexts per time period and feature
- Relative entropy (Kullback-Leibler Divergence [7]):

$$D_{feature}\,(T_1\,||T_2\,) = p(feature|T_1\,)\log_2\frac{p(feature|T_1)}{p(feature|T_2)}$$

[7] Solomon Kullback and Richard A. Leibler, 1951.

“-ing”-Forms: Syntagmatic change
Top 5 ngrams for the 1850s period

POS ngram	class	relative entropy (KLD)	example
Academic verbs			
SENT.IN.VVG	Gerund	0.0620	. ***In examining*** *the laws*
VVN.IN.VVG	Gerund	0.0587	*the formulae* ***employed in finding*** *these logarithms*
NN.IN.VVG	Gerund	0.0492	*Potasse for the* ***purpose of ascertaining*** *whether*
IN.RB.VVG	Gerund	0.0183	*opportunity* ***of sufficiently investigating*** *the errors*
SENT.RB.VVG	Gerund	0.0110	. ***Hence considering*** *an equation*
Motion verbs			
JJ.NN.VVG	Participle	0.0412	*the* ***smaller extremity lying*** *in contact with*
(.,.VVG	Participle	0.0370	*the tangential force (F),* ***forming*** *two equal*
JJ.NNS.VVG	Participle	0.0362	*refracting the* ***visual rays passing*** *thorough them*
IN.NNS.VVG	Participle	0.0327	*dark cloud* ***of ashes falling*** *from the volcano*
SENT.IN.VVG	Gerund	0.0270	. ***After passing*** *the central layer*
Change-of-state verbs			
VVN.IN.VVG	Gerund	0.1116	*more strongly* ***magnetized by placing*** *them*
SENT.IN.VVG	Gerund	0.0630	. ***By heating*** *it to above the boiling*
VVZ.IN.VVG	Gerund	0.0590	***crystallizes on cooling***
NN.,.VVG	Participle	0.0254	*a deep oblique* ***fold , penetrating*** *from the inner side*
JJ.NN.VVG	Participle	0.0235	*the* ***chylo-aqueous fluid filling*** *the ciliated*

IN: preposition, JJ: adjective, NN(S): common noun (pl.), RB: adverb, SENT: full stop, VVG: *ing*-form, VVN: participle, VVZ: present tense

“-ing”-Forms: Syntagmatic change
Key Findings

- **Comparison of ngrams** in this time period with respect to grammatical class:
  - Academic verbs: gerunds
  - Change-of-state verbs: gerunds (most distinctively)
  - Motion verbs: participles

Conclusion and discussion

Conclusion and discussion
Summary

- Word embeddings can be used to analyze changes in **lexical and grammatical use**
  - Different groups of words exhibit different topological behaviors
  - Polyfunctional words are influenced by both lexis and grammar
- "-ing"-verbs form clusters according to **three categories**
- Connection between verb category and **grammatical class** (gerunds, participles)
- 19th century seems to have a major influence on linguistic changes

Conclusion and discussion
Questions

1. **Is this a robust method to quantify changes in grammatical patterns?**
2. **How important is the interplay between lexical and grammatical meaning with regards to large language models?**

References

- **Yuri Bizzoni et al. 2019. Grammar and meaning: Analysing the topology of diachronic word embeddings. In: Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, pages 175-18, Florence, Italy. ACL.**
- DeepAI: Ngrams [online]. URL: https://deepai.org/machine-learning-glossary-and-terms/n-gram (accessed 16.11.2023)
- Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data Using t-SNE. Journal of Machine Learning Research, 9. November 2008, pages 2579–2605.
- Wang Ling et al. 2015. Two/too Simple Adaptations of word2vec for Syntax Problems. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1299–1304, Denver, Colorado. ACL.
- Thanh N Tran et al. 2013. Revised DBSCAN Algorithm to Cluster Data with Dense Adjacent Clusters. Chemometrics and Intelligent Laboratory Systems, 120:92–96.
- Solomon Kullback and Richard A. Leibler. 1951. On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1):79–86.
- David Sculley. 2010. Web-scale K-Means Clustering. In Proceedings of the 19th International Conference on World Wide Web, pages 1177–1178, Raleigh, North Carolina, USA. ACM.
- Brendan J Frey and Delbert Dueck. 2007. Clustering by Passing Messages between Data Points. Science, 315(5814):972–976.
- Derivation of word2vec formula and simple implementation in python [online]. URL: https://www.fatalerrors.org/a/derivation-of-word2vec-formula-and-simple-implementation-in-python.html (accessed 21.11.2023)
- Semantic change [online]. URL: https://lena-voita.github.io/resources/lectures/word_emb/research/semantic_change-min.png (accessed 21.11.2023)