Grammar and Meaning
Analysing the topology of diachronic word embeddings
Yuri Bizzoni, Stefania Degaetano-Ortlieb, Katrin Menzel, Pauline Krielke, Elke Teich
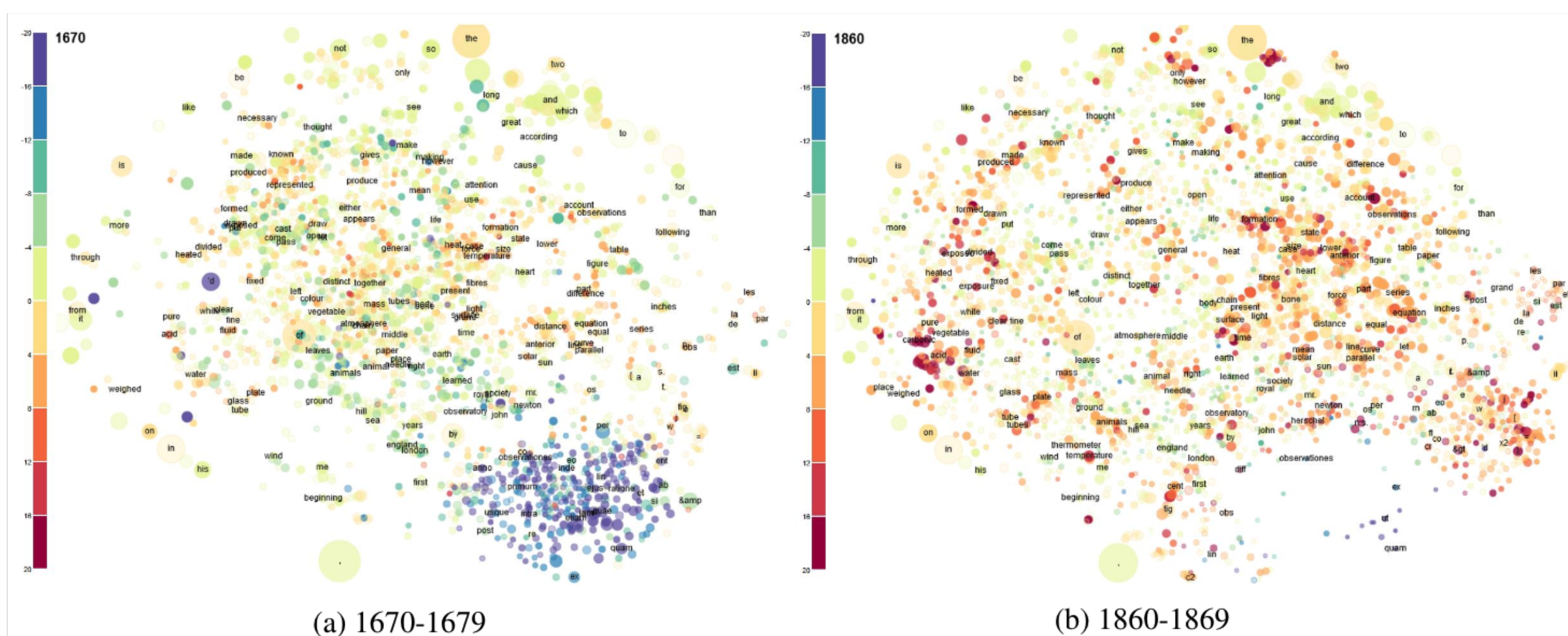
General motivation

- **Language change** in the domain of science, focusing on the Late Modern English period (17-19th century).

- Scientific **style trends**: decrease of entropy, increase of information density, specialization.

- **Diachronic word embeddings**:
  - Clustering and visualization
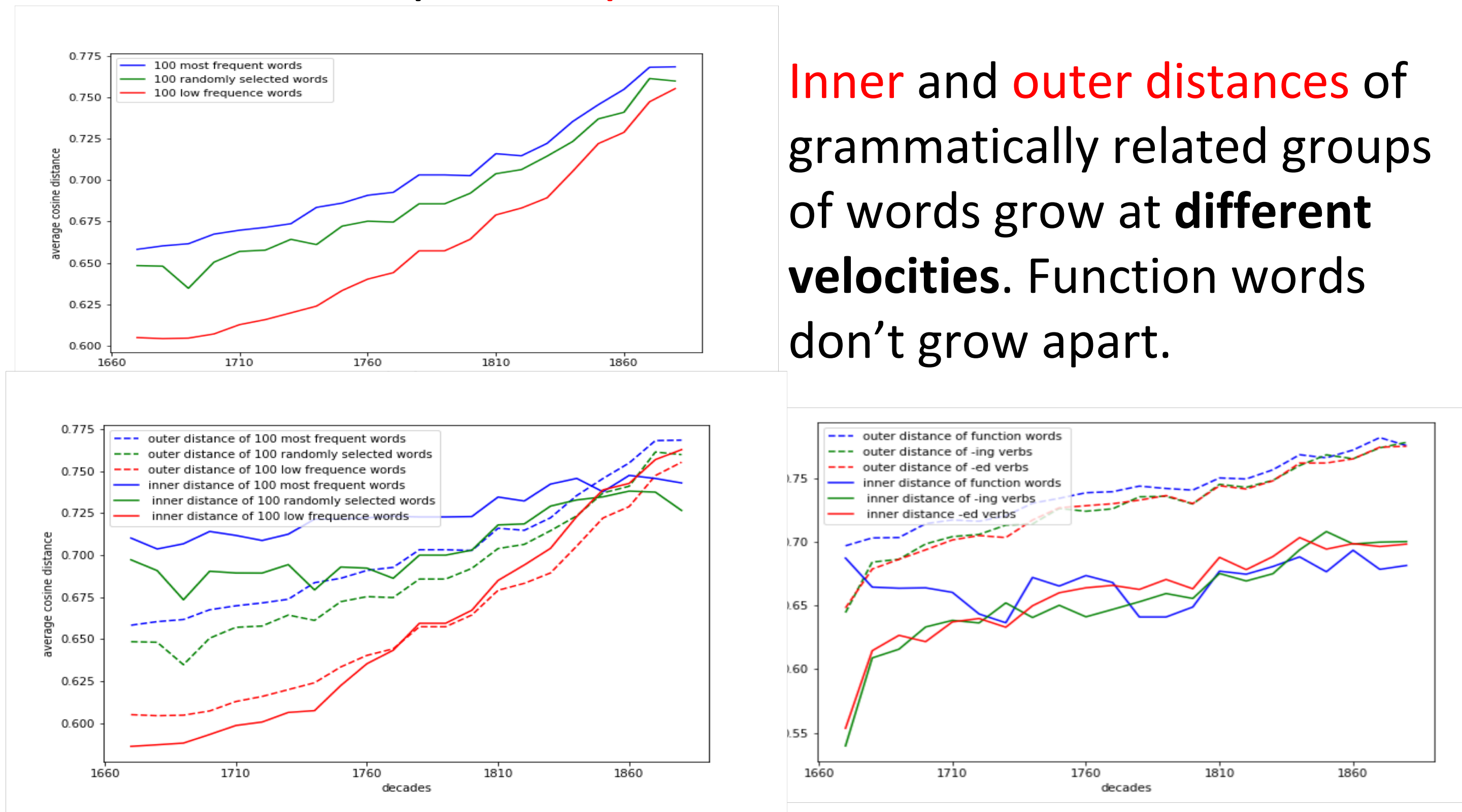 - Comparison with corpus based relative entropy

Data

Royal Society Corpus (RSC): publications Philosophical Transactions of the Royal Society of London 1665-1869 (ca. 32 million tokens).

decade	tokens	lemmata	sentences
1660-69	455,259	369,718	10,860
1670-79	831,190	687,285	17,957
1700-09	780,721	615,770	23,338
1710-19	489,857	383,186	17,510
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
1800-09	1,615,564	1,298,978	45,666
1810-19	1,446,900	1,136,581	42,998
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

Diachronic Word Embeddings

A skip-gram model taking into account word order is fine-tuned for each decade.

(a) 1670-1679   (b) 1860-1869

Tracing the -ings

- **Fix threshold**: lexical pull

- **Dynamic threshold**: functional loose clustering?

- **Three** main **loose clusters**: academic, motion and change of state verbs

- **Popular algorithms**:
1. Affinity Propagation
2. DBSCAN
3. Minibatch K-Means

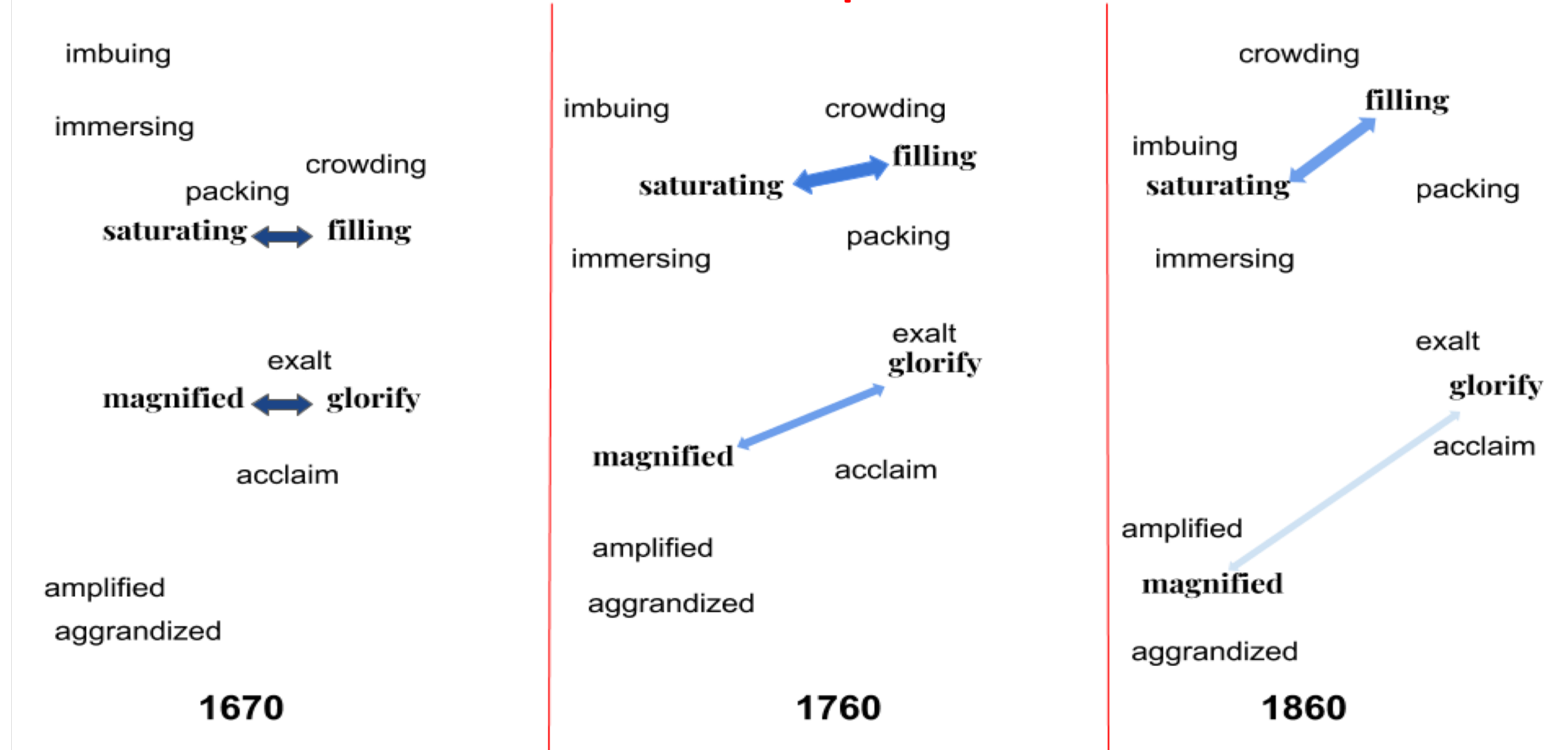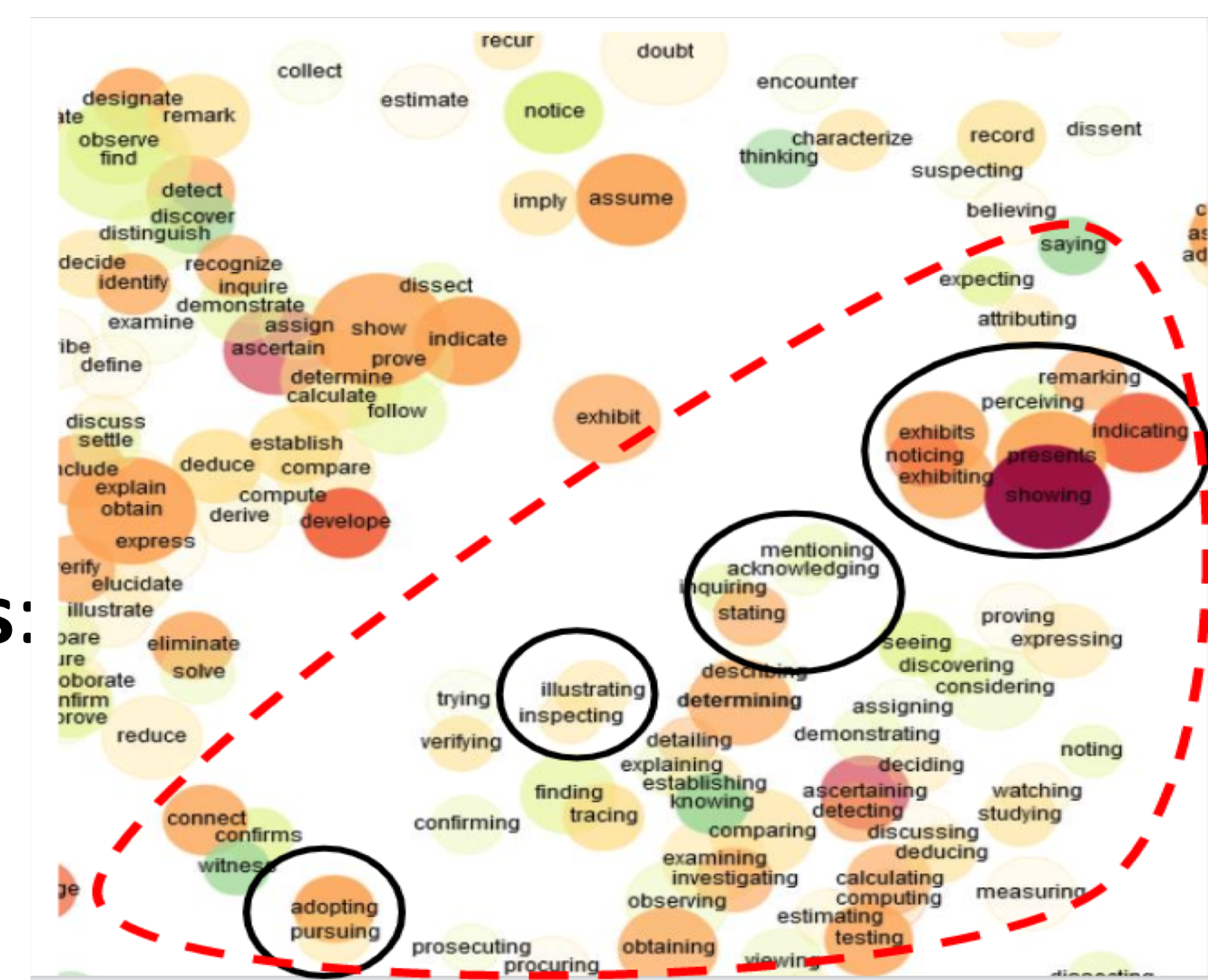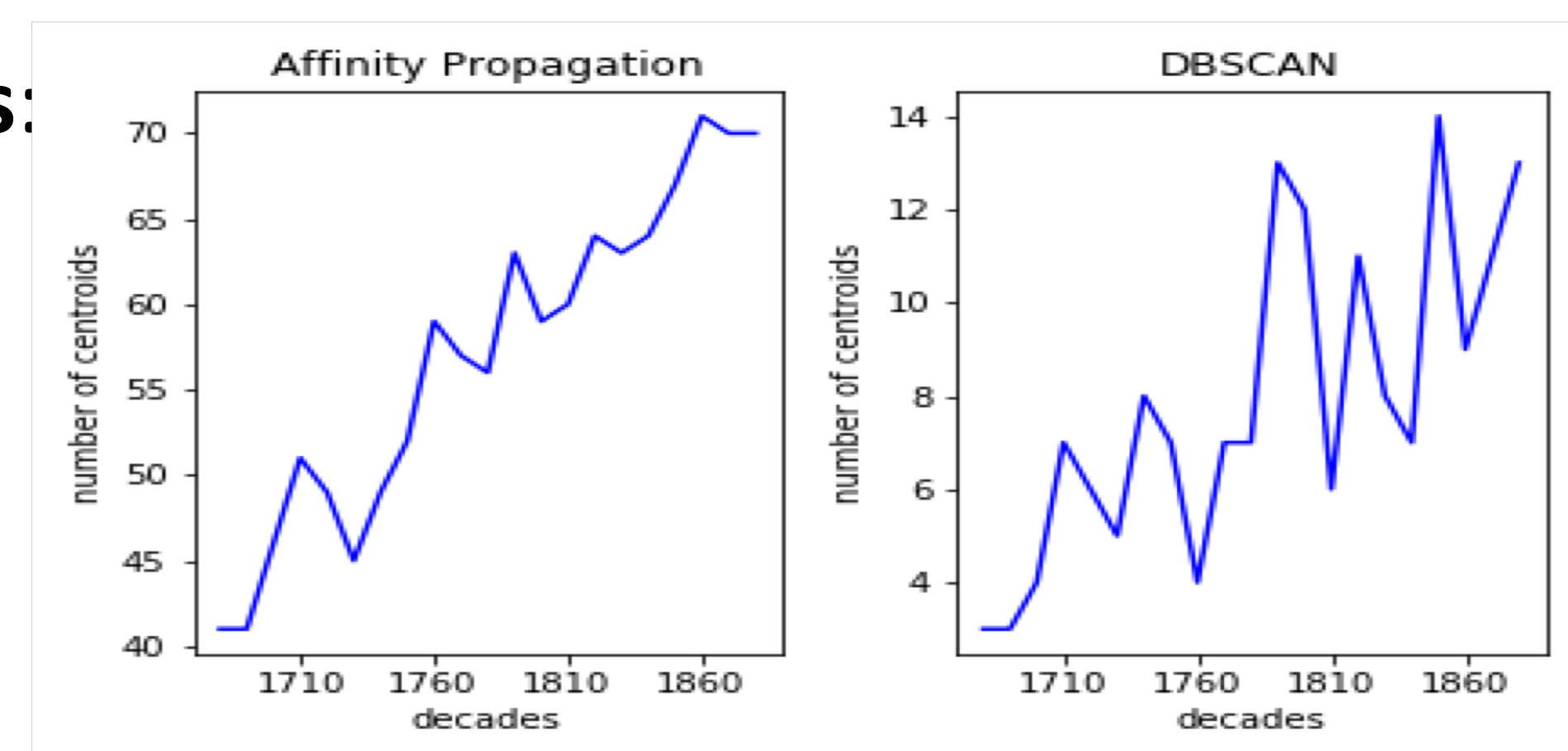## Topology of the space over time

The semantic space expands over time.

Inner and outer distances of grammatically related groups of words grow at **different velocities**. Function words don't grow apart.

Diachronic clusters of -ings

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

Functional and lexical side pull in different directions

Lexico-functional clusters

Kullback-Leibler Divergence to find which **grammatical classes** are **distinctive** of later periods for each group of verbs.

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	Participle	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

Conclusions

- The semantic spaces of the RSC are expanding (due to an increasing lexical specialization?)

- Within this expansion, some words words are forming non-semantic loose clusters

- *-ing* verbs show an increased polyfunctionality, displaying semantic-functional clusters

SOME BIBLIOGRAPHY

Stefania Degaetano-Ortlieb, Hannah Kermes, Ashraf Khamis, and Elke Teich. 2018. An Information-Theoretic Approach to Modeling Diachronic Change in Scientific English. In *From Data to Evidence in English Language Research*, Language and Computers. Brill, Leiden.

Peter Fankhauser, Jörg Knappen, and Elke Teich. 2014. Exploring and Visualizing Variation in Language Resources. In *Proceedings of the 9th LREC*, pages 4125–4128, Reykjavik, Iceland. ELRA.

Peter Fankhauser and Marc Kupietz. 2017. Visualizing Language Change in a Corpus of Contemporary German. In *Proceedings of the Corpus Linguistics International Conference*, Birmingham, UK.