INSTITUT FÜR DEUTSCHE SPRACHE

Peter Fankhauser, Abteilung Zentrale Forschung

VISUAL CORRELATION FOR EXPLORING PARADIGMATIC LANGUAGE CHANGE

Joint work with Marc Kupietz and Elke Teich

Mitglied der Leibniz-Gemeinschaft

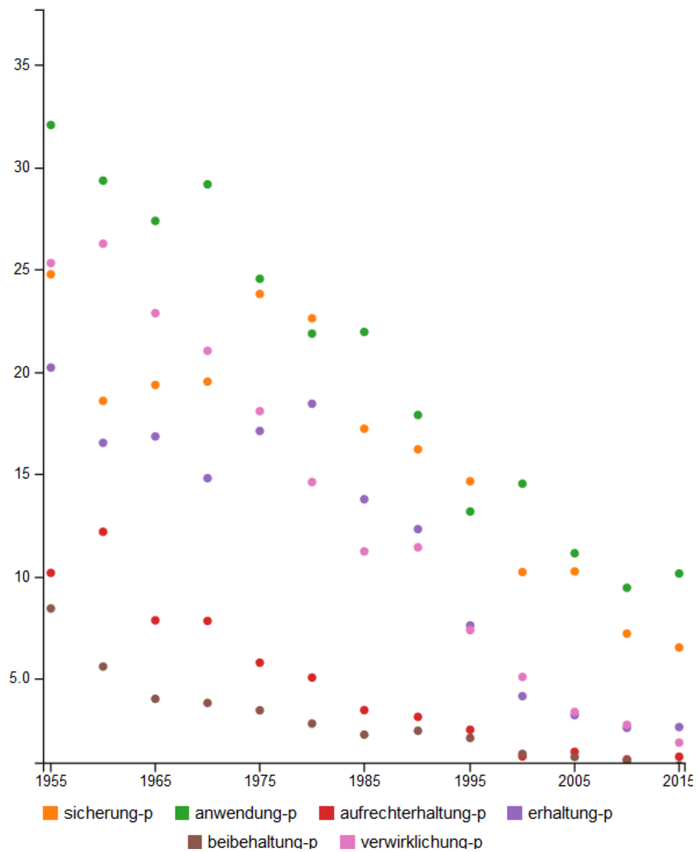
OVERVIEW

- Goal: Explore paradigmatic language change
  - Words with similar usage context
  - Rise and fall together
- Approach: Visually correlate
  - Frequency Change
  - (Distributional) Semantics of Words
- Some Examples
- Concluding Remarks

Time changes all things. There is no reason why language should escape this universal law.

Ferdinand de Saussure,
Course in General Linguistics
1916/1959

EXAMPLE:
DECREASE OF NOMINALIZATION WITH „-UNG“



Sicherung
Anwendung
Aufrechterhaltung
Erhaltung
Beibehaltung
Verwirklichung
(...)

VISUALIZING FREQUENCY CHANGE BY COLOR

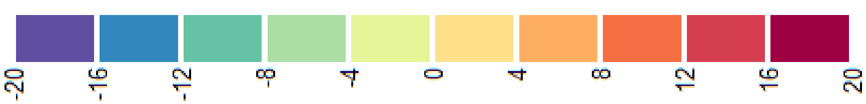
- Fit Logistic Growth Curve to Timed Frequencies *p(t)*

$$p(t) = \frac{1}{1 + e^{-k - s * t}}$$

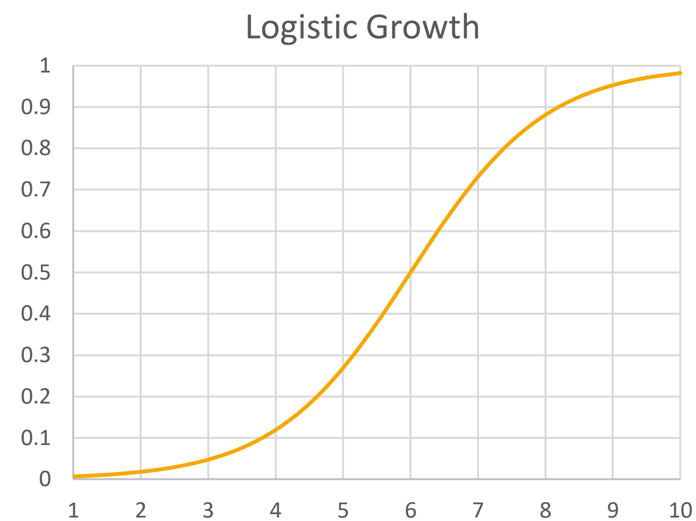
- *k* … Intercept
  s … Slope
  t … Time
- Equivalently: Logit

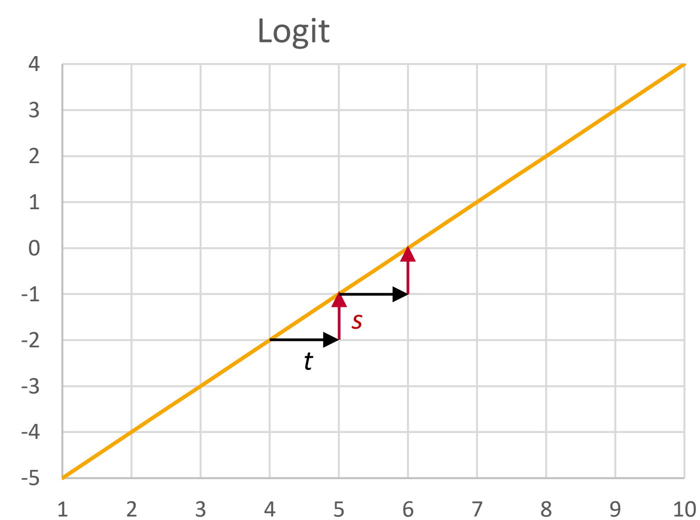
$$\ln\left(\frac{p(t)}{1 - p(t)}\right) = k + s * t$$

- Map Slope *s* to Color

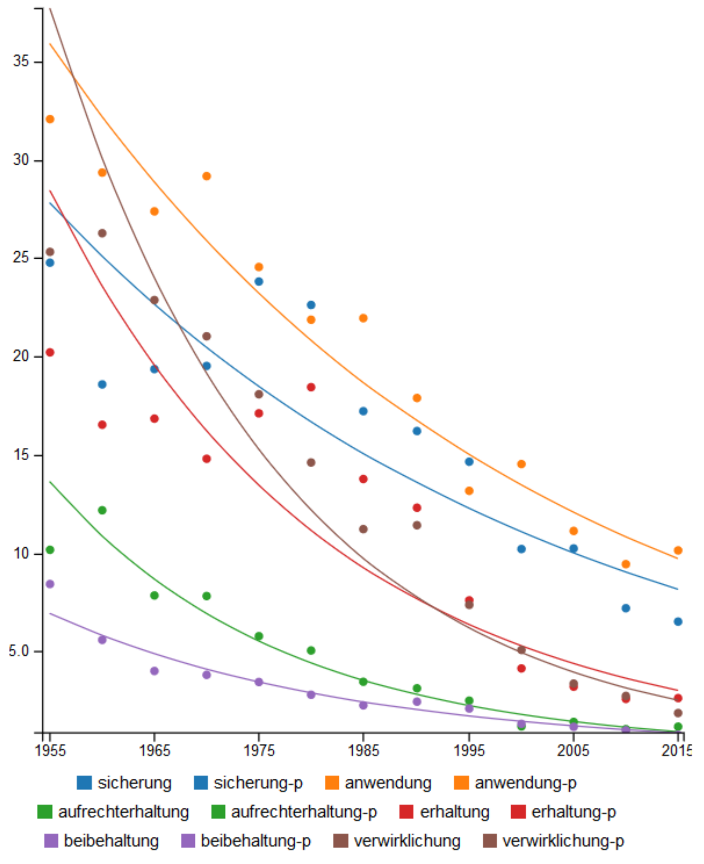


Similar Slope : Same Color

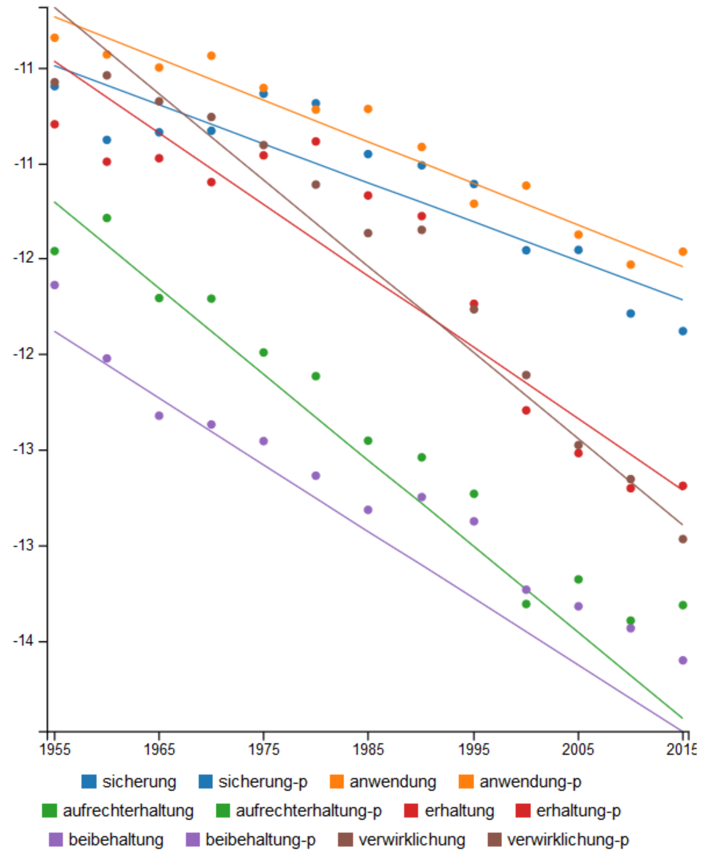




EXAMPLE:
FREQUENCIES AND FITTED CURVES

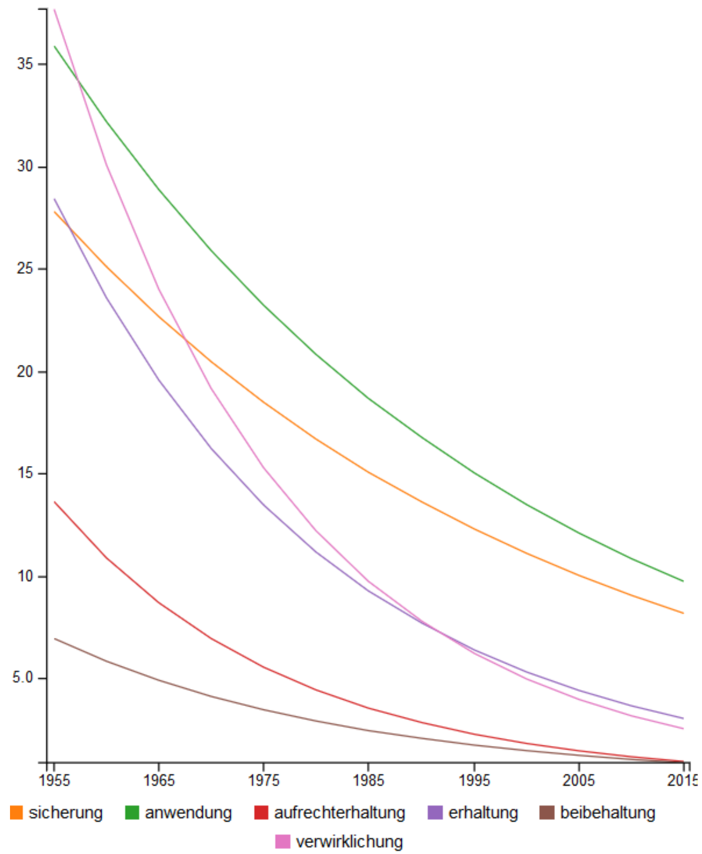


Freq per Mio

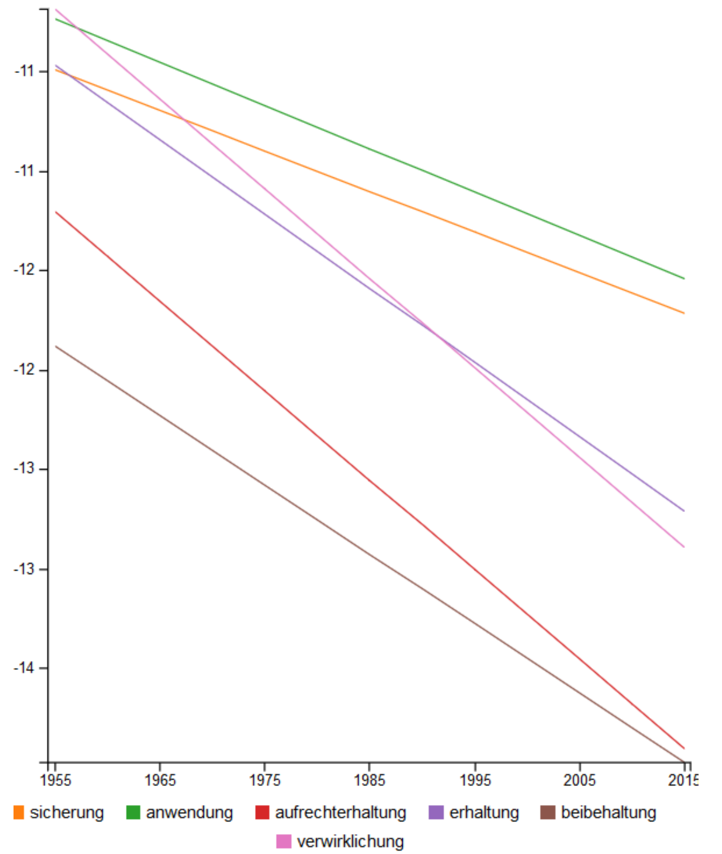


Logit(FpM)

EXAMPLE:
FITTED CURVES

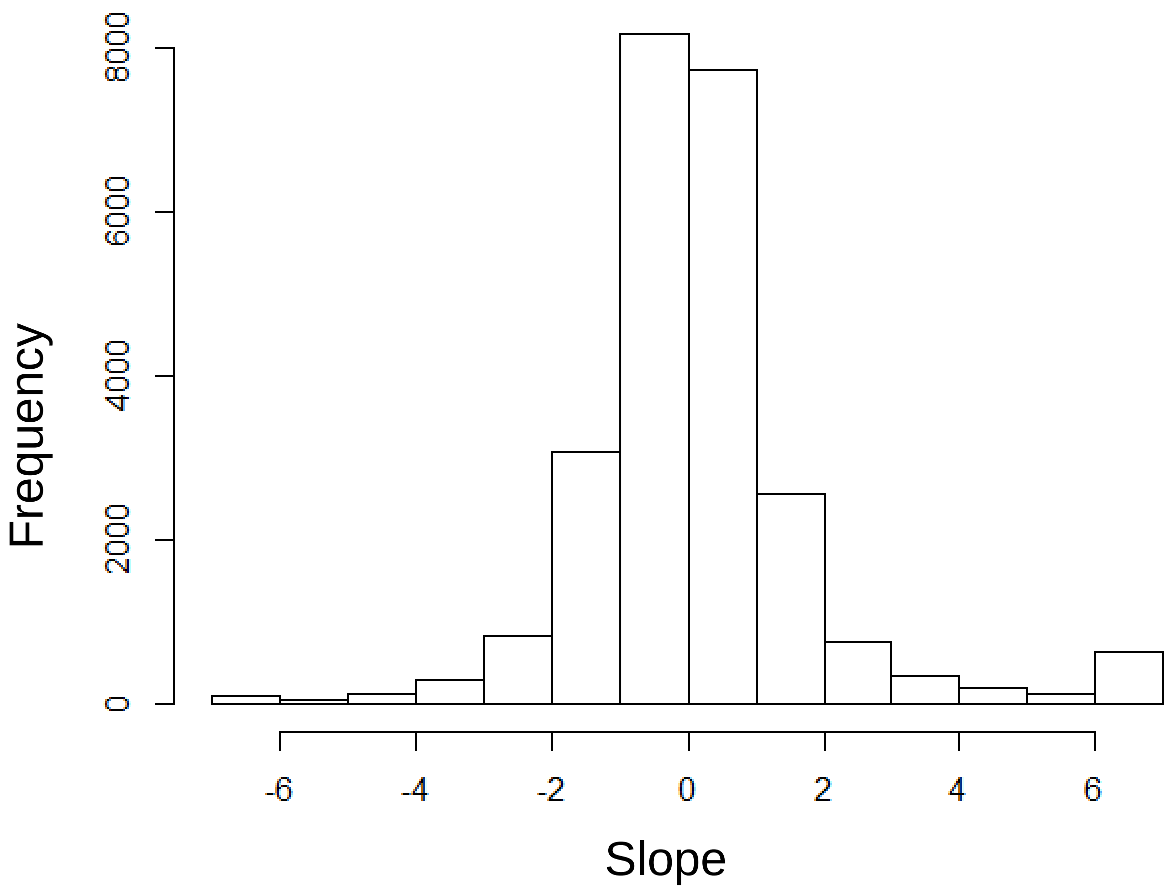


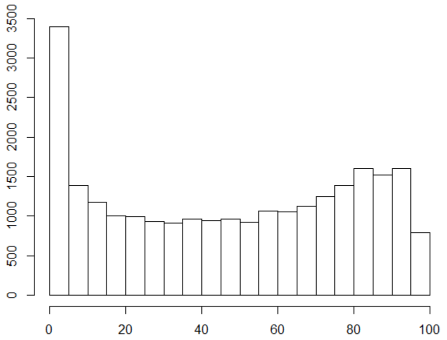
Freq per Mio



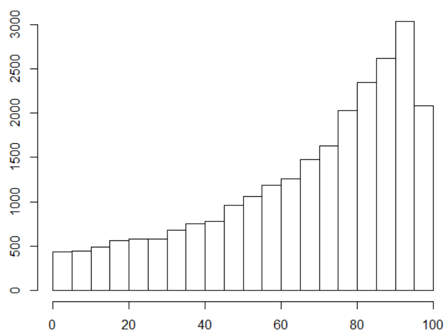
Logit(FpM)

SPIEGEL/ZEIT: DISTRIBUTION OF SLOPES, ROOT MEAN SQUARE ERRORS (RMSE)





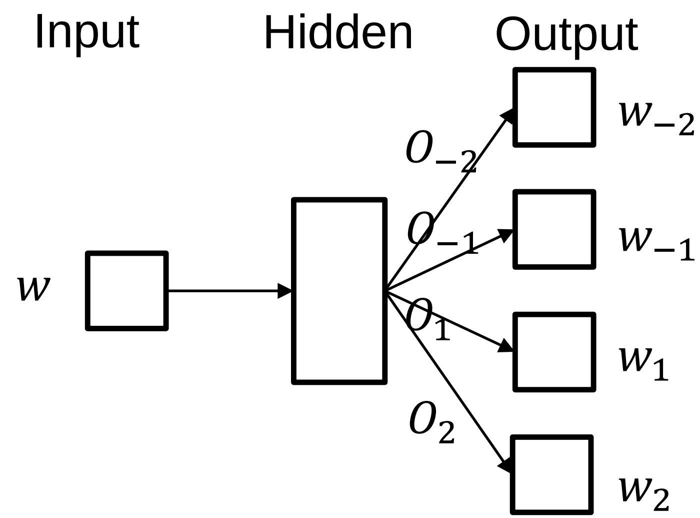
RMSE1: 47.6%



RMSE2: 66.6%

REPRESENTING WORD USAGE IN FEW DIMENSIONS

- Word Co-Occurrence Vectors
  - $p(w_{-1}|w)$
  - Number of dimensions = vocabulary size (* context size)
- Word Embeddings
  - Structured Skipgram (Wang2Vec [4])
  - Learn mapping from word w to left/right context $(w_{-2}, w_{-1,} w_{1,} w_{2,})$ via hidden layer with few dimensions (100-200)
- Diachronic Word Embeddings [7]
  - Start with random Neural Net
  - Initialize Neural Net for Time $t+1$ by Neural Net for Time t



VISUALIZING WORD EMBEDDINGS IN TWO DIMENSIONS

- T-Distributed Stochastic Neighbor Embedding (T-SNE) [8]
  - Map from n Dimensions to 2 Dimensions
  - Given: Probability of Word Vectors x_i and x_j: $p_{ij} = \frac{e^{-\left\|x_i - x_j\right\|^2/2\sigma^2}}{\sum_{k \neq l} e^{-\left\|x_k - x_l\right\|^2/2\sigma^2}}$
  - Seek: Word Coordinates y_i and y_j, with: $q_{ij} = \frac{\left(1+\left\|y_i - y_j\right\|^2\right)^{-1}}{\sum_{k \neq l}(1+\|y_k - y_l\|^2)^{-1}}$
  - Such that the KL-Divergence between *P* and *Q* is minimal: $KL(P||Q) = \sum_{i,j} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right)$
- Preserves *local* structure, compromising global structure
  - Close x_i, x_j (should) remain close y_i, y_j
  - Larger distances do not need to be preserved accurately
- Caveat: Global position has no interpretation

EXAMPLE: NOMINALIZATION WITH „-UNG“ VISUAL CORRELATION

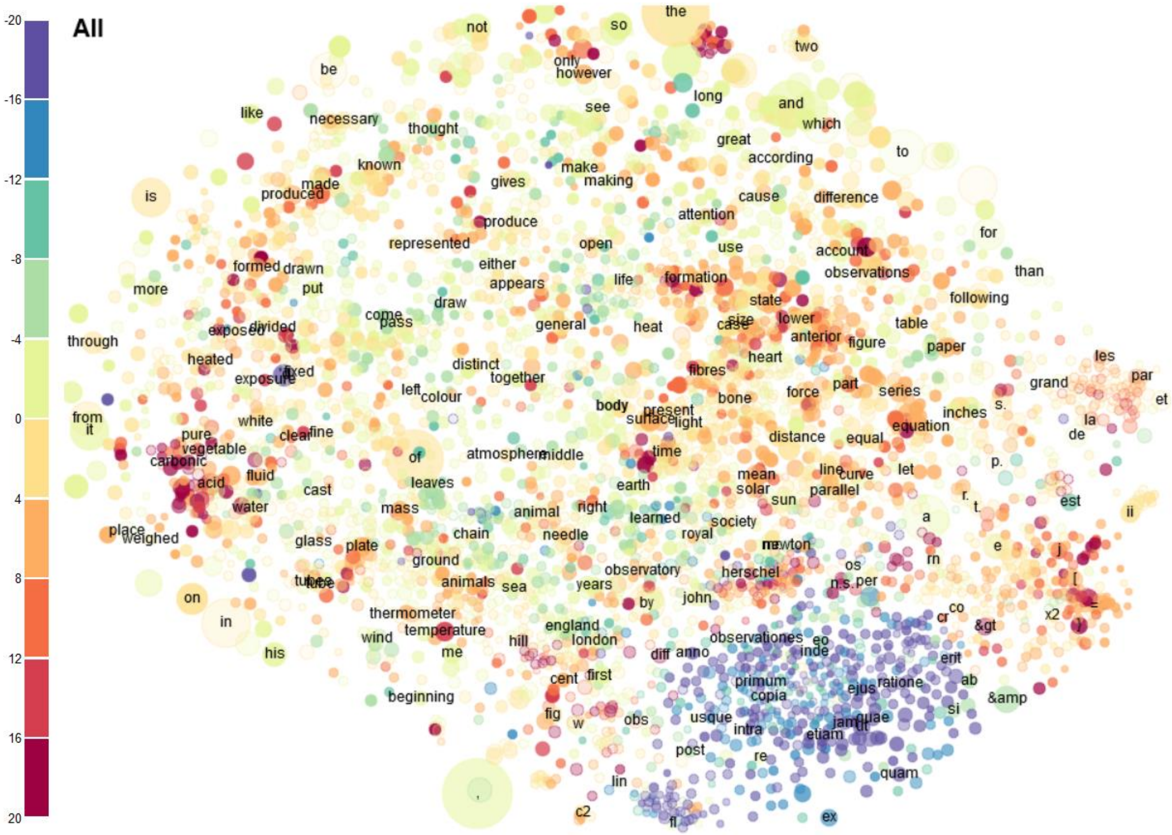
1955

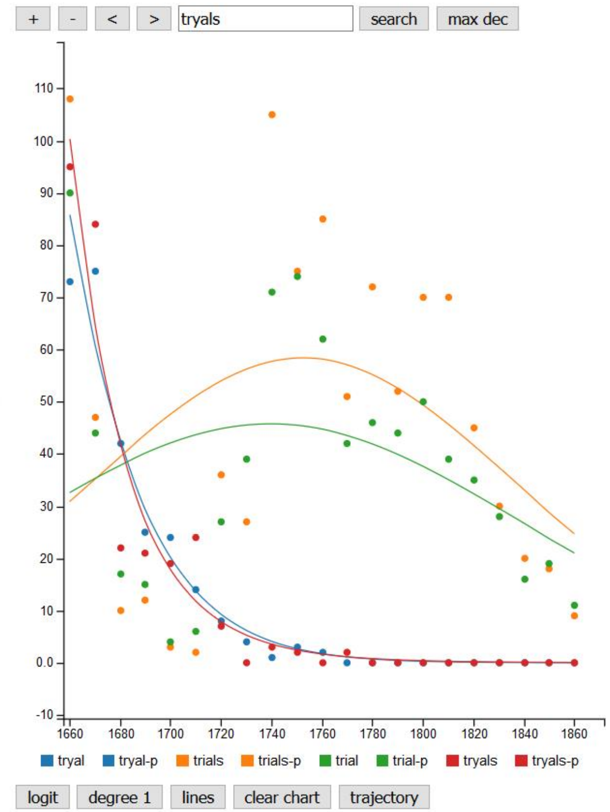
2015

CORPORA: SUMMARY INFO

	Royal Society	Spiegel/Zeit	DeReKo News
Span	1665-1869	1953-2015	2000-2015
Timeslices	10 years	5 years	1 year
Tokens	35 Mio	570 Mio	18825 Mio
Visualized Types	18700	25000	25000
RMSE1	31.1%	47.6%	44.4%
RMSE2	46.2%	66.6%	59.2%
Median Slope	0.093	-0.014	-0.023
Embedding Dim	100	200	200
Nearest Neighbor Slope Correlation	0.77	0.43	0.42
http://corpora.ids-mannheim.de/diaviz/	royalsociety.html	zeitspiegel.html	dereko.html

VISUALIZATION OVERVIEW
ROYAL SOCIETY CORPUS





CORRELATION BETWEEN FREQUENCY CHANGE AND USAGE SIMILARITY (RSC)

- Coefficients of Frequency Change

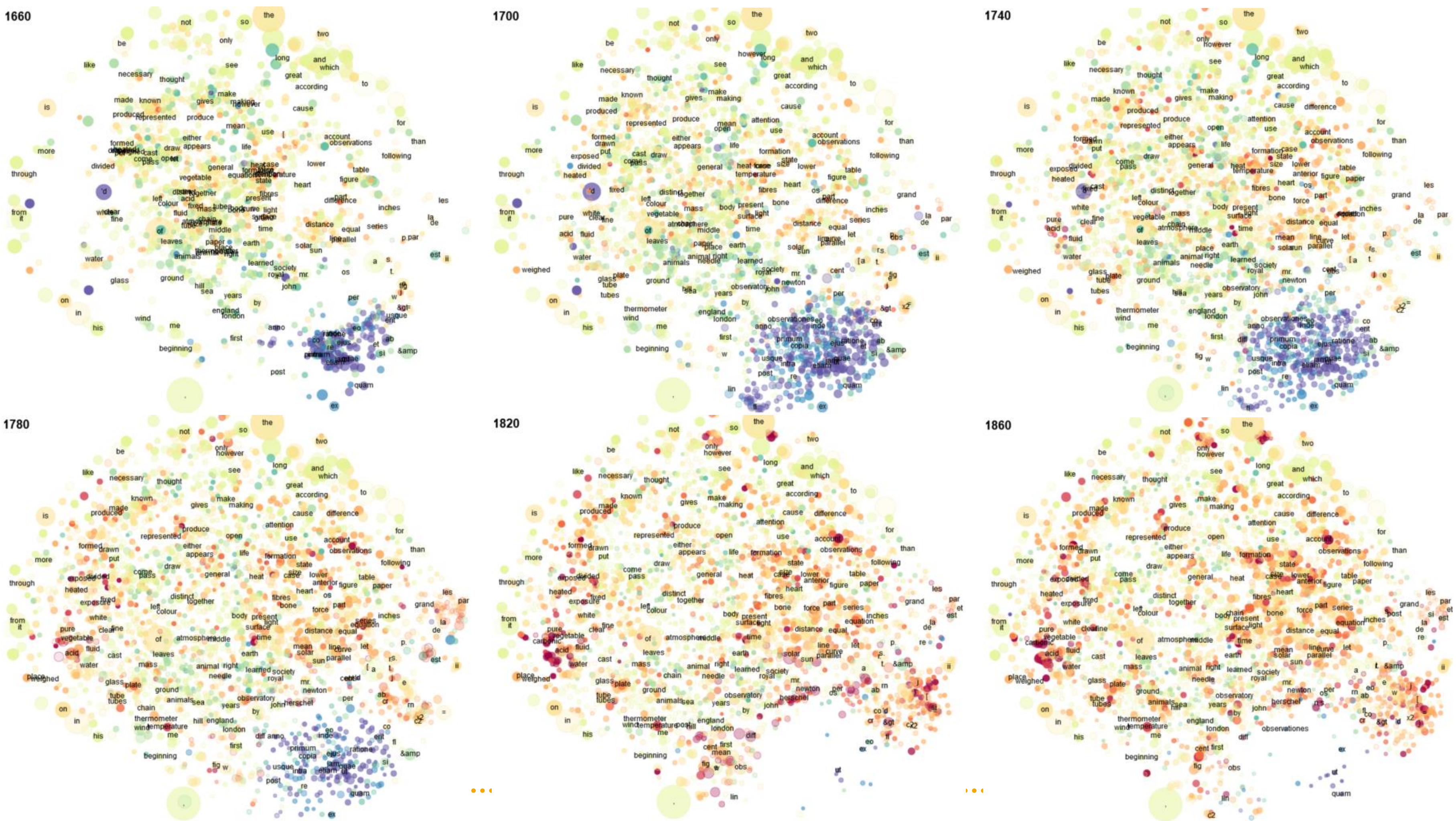
$$\ln\left(\frac{p(t)}{1-p(t)}\right) = k + s * t + c * t^2$$

k ... Intercept: Start
s ... Slope: Rate of Change
c ... Curvature: Change of Slope

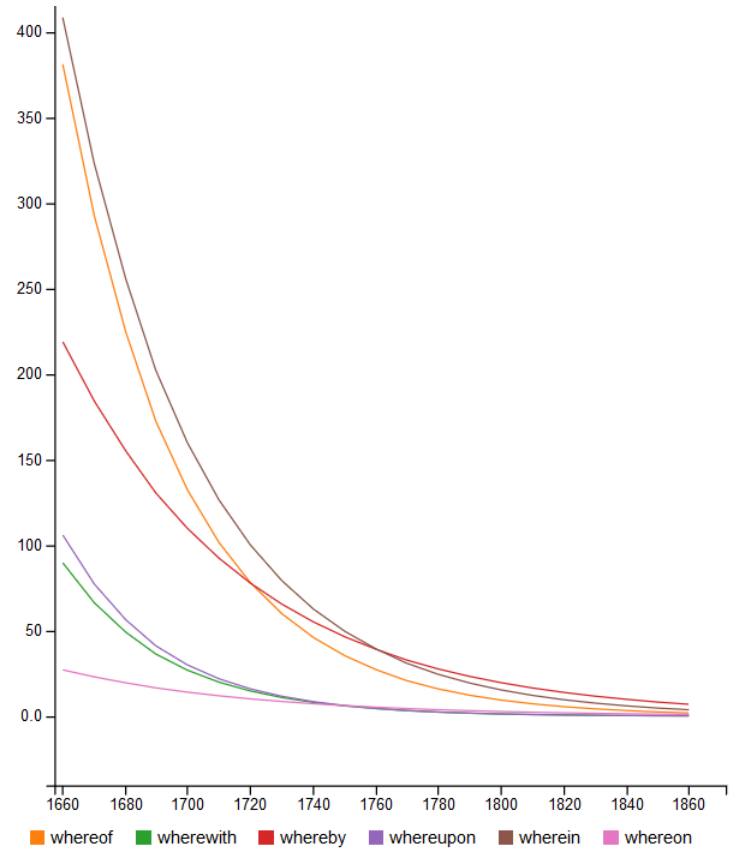
- Spearman Rank Correlation ρ
  - Strongest between Slopes of Nearest Neighbors (*NN*)
  - Curvature stronger than Intercept
  - Decreasing with increasing distance between neighbors

NN	*k*	*s*	*c*
1	0.53	0.77	0.63
2	0.49	0.74	0.59
3	0.46	0.73	0.56

VOCABULARY DIVERSIFICATION IN THE ROYAL SOCIETY CORPUS

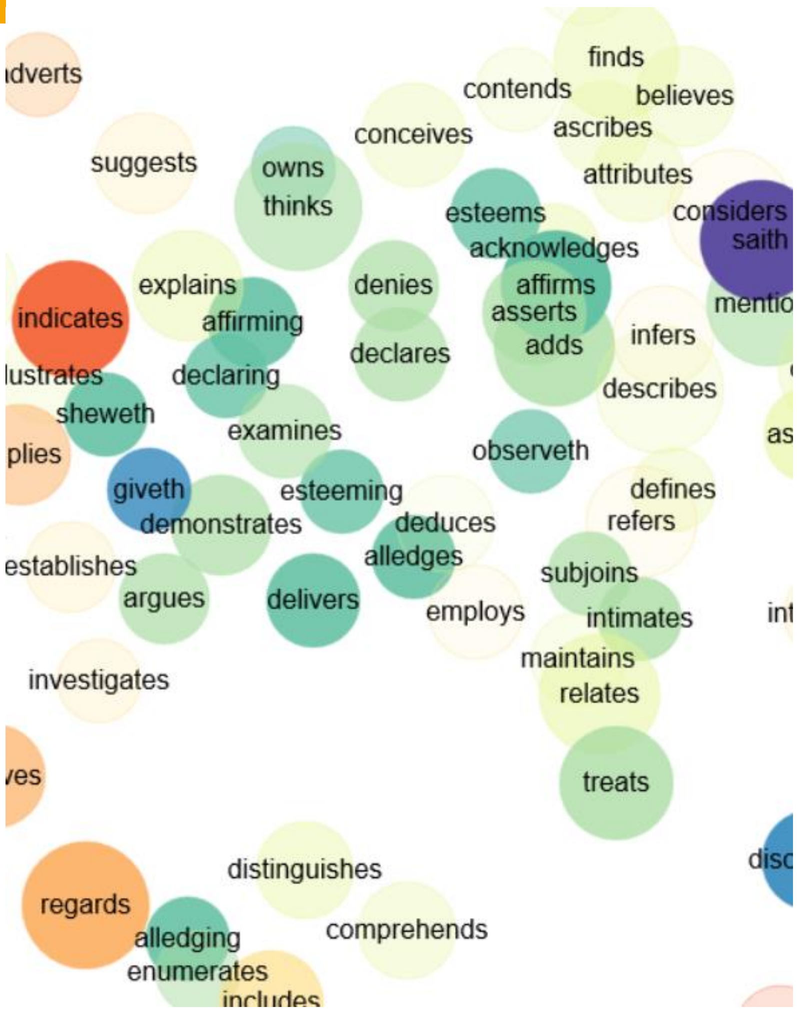


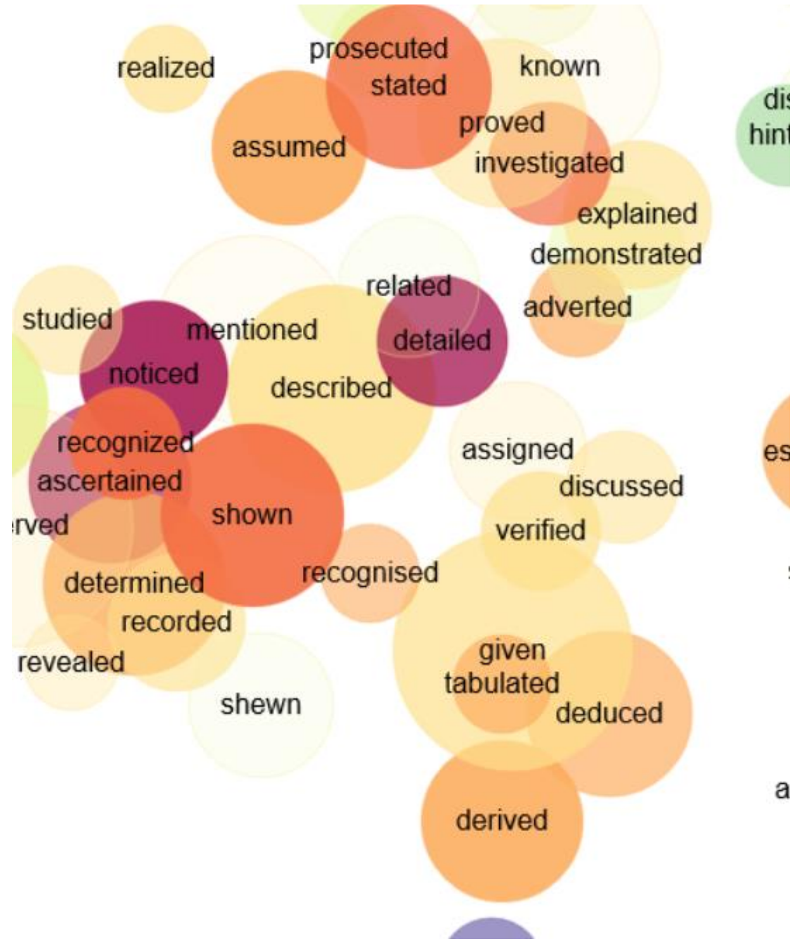
GRAMMAR
WH-ADVERBS DOWN



GRAMMAR: COMMUNICATIVE/MENTAL VERBS

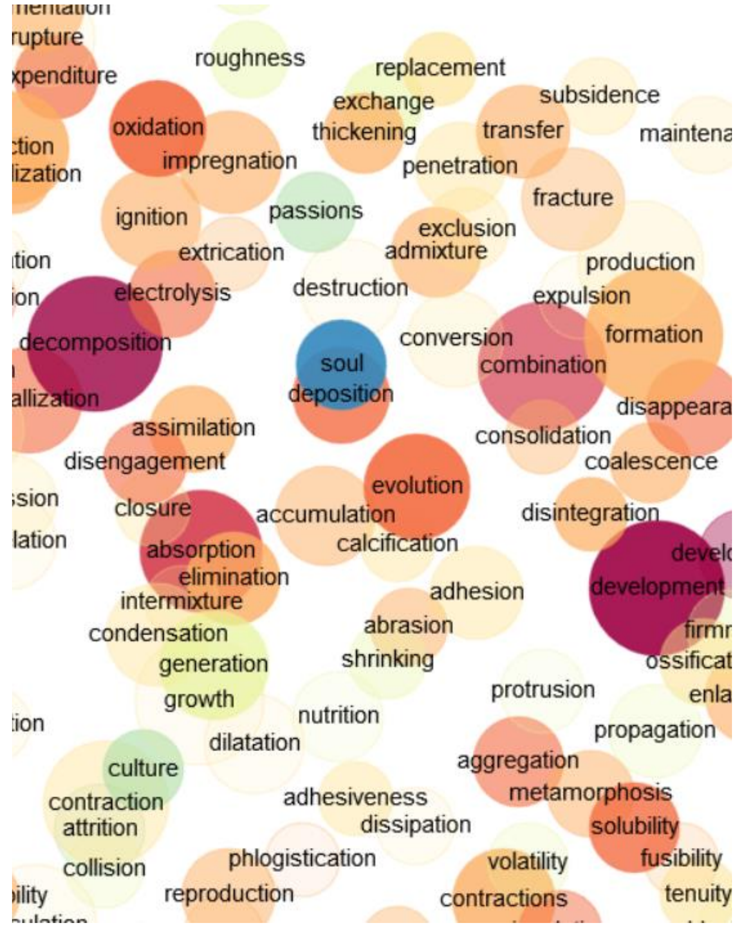
PRESENT TENSE DOWN PASSIVE/PAST UP





GRAMMAR/STYLE

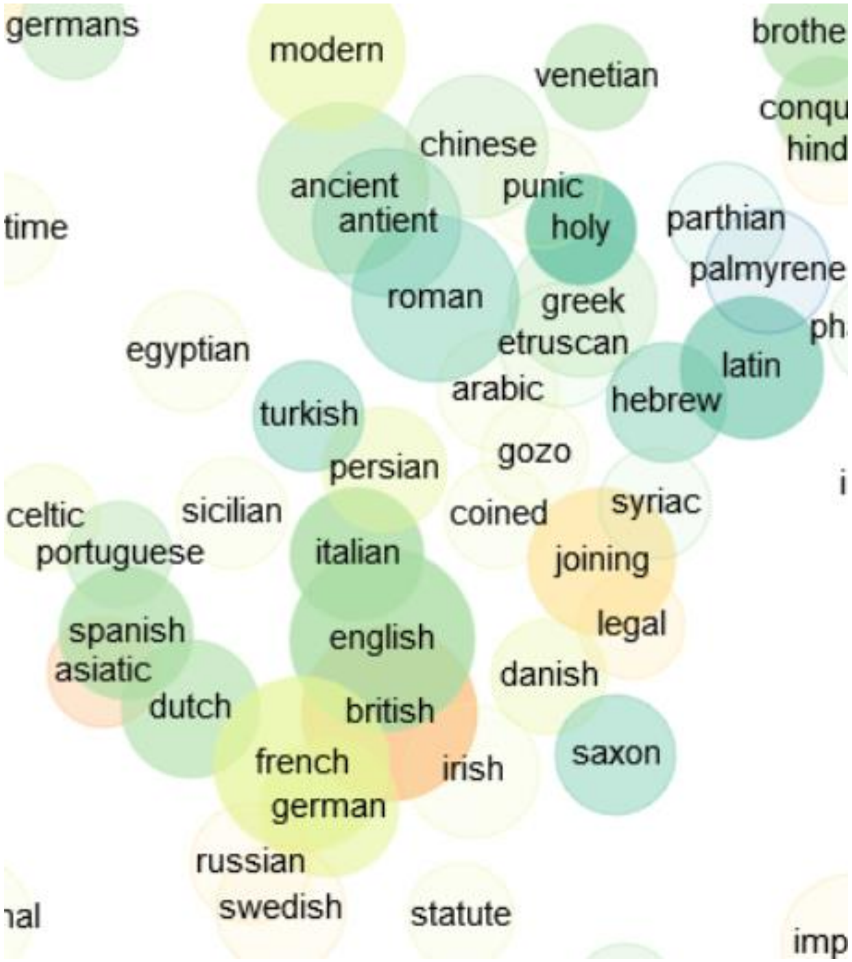
PERSON ADJECTIVES DOWN PROCESS NOUNS UP

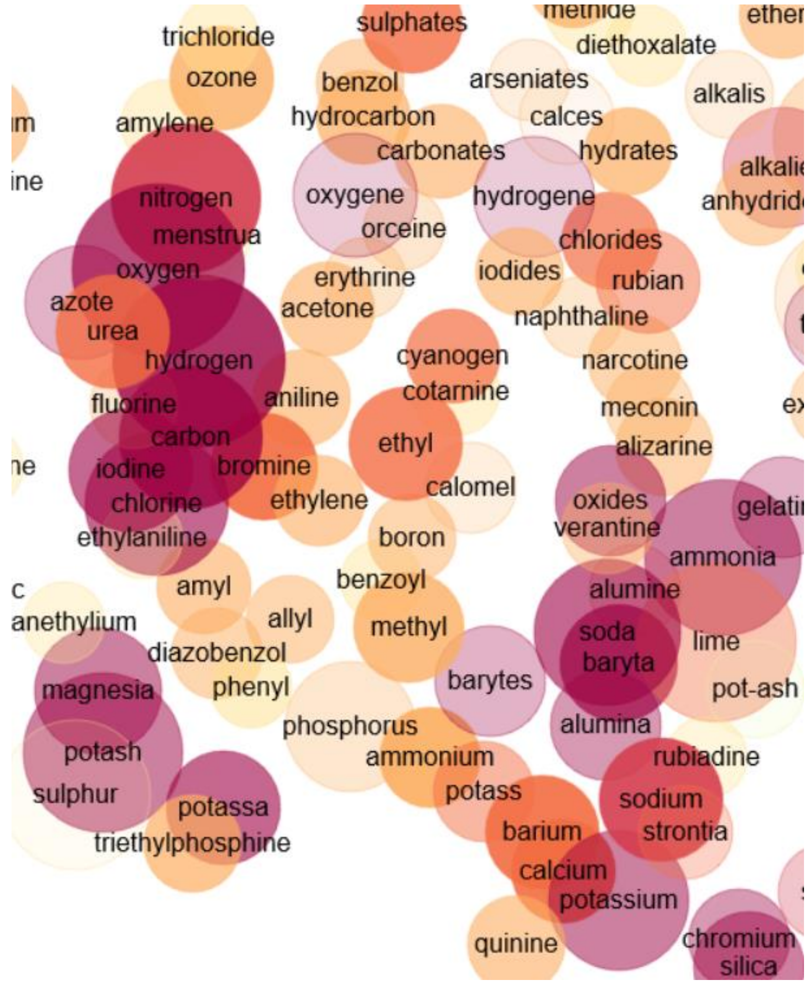


THEME

PROVENANCE DOWN

CHEMISTRY UP

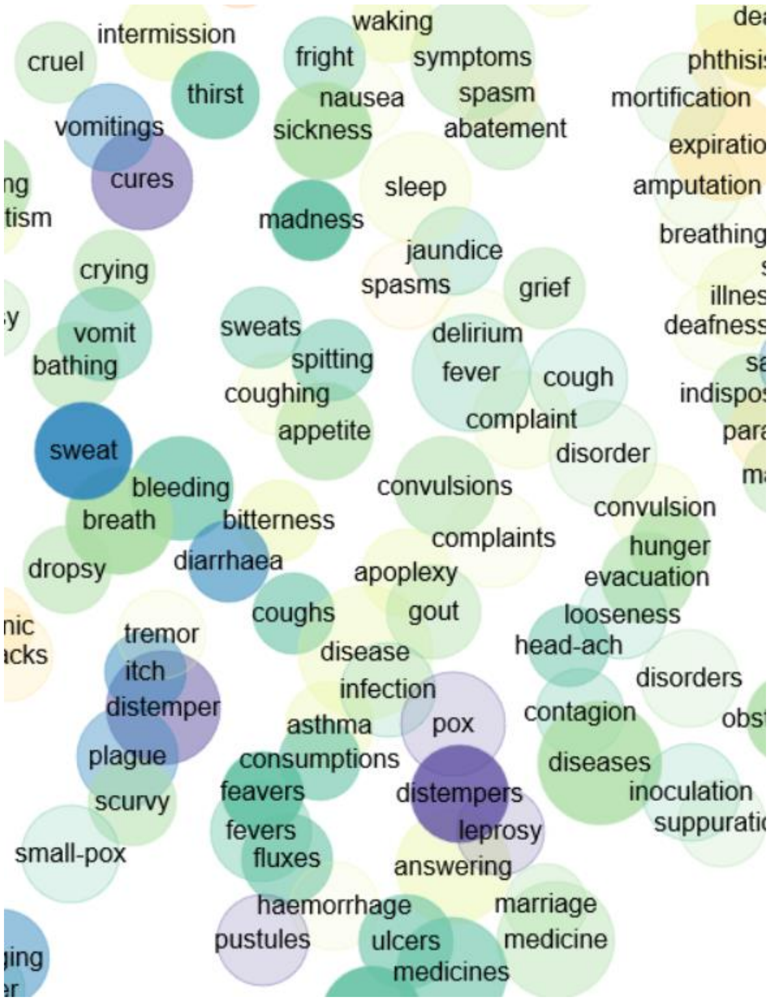




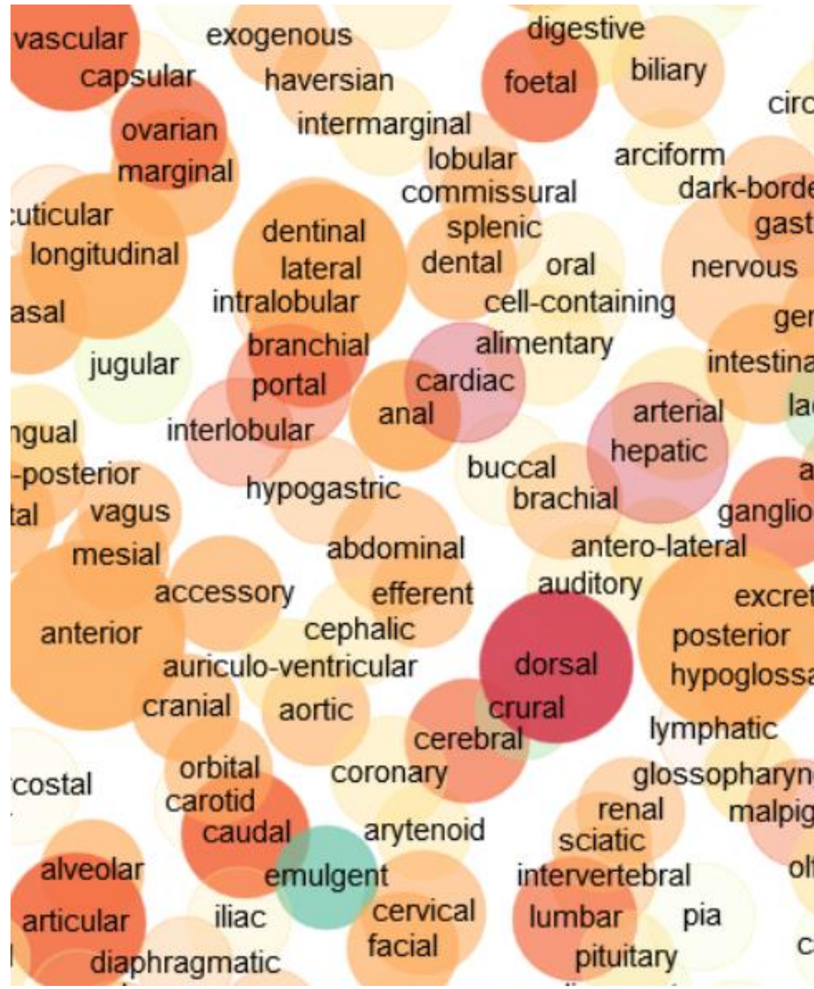
INSTITUT FÜR DEUTSCHE SPRACHE

THEME: MEDICINE

SYMPTOMS DOWN



ANATOMY UP



NEAREST NEIGHBOR WINNERS AND LOSERS IN THE ROYAL SOCIETY CORPUS

Down	Up
bigness	size
splendor	brilliancy
curious	interesting
loadstone	magnet
impertinent	unnecessary
plentiful	abundant
remembrance	recollection
hindred/hindered	prevented
contrived	constructed
truths	facts

SUMMARY AND RELATED WORK

- Paradigmatic Frequency Change
  - Paradigmatically related words rise and fall together
  - Not only due to theme
- Lehrer 1985 [6]: Parallel Change
  - Paradigmatically related words change/extend their meaning together
  - E.g.: derogatory Ape/Baboon/Gorilla
  - But: Cf. Xu/Kemp 2015 [9]
- Kroch 1989 [5]: Constant (Equal) Rate Hypothesis
  - Language change has the same rate independent of context
  - E.g.: Periphrastic Do
- Dubossarsky et al. 2015 [1]: „Marginal" words more likely to change meaning
  - Correlation between distance from cluster center and meaning change
- Hamilton et al. 2015 [3]: Two „Laws" of semantic change
  - Meaning change vs. Frequency vs. Polysemy

REFERENCES

(1) Dubossarsky, H., Y. Tsvetkov, C. Dyer & E. Grossman (2015). A bottom up approach to category mapping and meaning change. In: Word Structure and Word Usage. Proceedings of the NetWordS Final Conference, 66-70. Pisa.

(2) Fankhauser, P., M. Kupietz (2017). Visualizing Language Change in a Corpus of Contemporary German. Corpus Linguistics Conference 2017.

(3) Hamilton, W., J. Leskovec & D. Jurawsky (2015). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. ACL 2016

(4) Kim, Y., Y. Chiu, K. Hanaki, D. Hedge & S. Petrov (2014). Temporal Analysis of Language through Neural Language Models. ACL 2014 Workshop on Language Technologies and Computational Social Science

(5) Kroch, A. (1989). Reflexes of grammar in patterns of language change. In Language Variation and Change 1 (3), 199-244.

(6) Lehrer, A. (1985). The influence of semantic fields on semantic change. In J. Fisiak (Ed.), Historical semantics: Historical word formation (pp. 283-296). Berlin: Mouton de Gruyter.

(7) Ling, W., C. Dyer, A. Black & I. Trancoso (2015). Two/Too Simple Adaptations of Word2Vec for Syntax Problems. Human Language Technologies (NAACL HLT 2015)

(8) Van der Maaten, L., G.E. Hinton (2008). Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 9(Nov):2579-2605, 2008.

(9) Xu, Y. & C. Kemp (2015). A Computational Evaluation of Two Laws of Semantic Change. CogSci 2015.

NEAREST NEIGHBOR WINNERS AND LOSERS IN THE SPIEGEL/ZEIT CORPUS

Down	Up
Carters	Obamas
DM	Euro
Brandt	Scharping
Herberger	Klinsmann
Rhodesien	Simbabwe
Industrialisierung	Globalisierung
Argwohn	Misstrauen
Erkenntnis	Gewissheit
fraglich	ungewiss
Grundbesitz	Immobilien

NEAREST NEIGHBOR WINNERS AND LOSERS IN DEREKO NEWS

Down	Up
Stoiber	Guttenberg
Tschernobyl	Fukushima
Windows	Android
Blair	Cameron
Scharping	Bahr
Kindergeld	Betreuungsgeld
PCs	Smartphones
Handys	Smartphones
Neuverschuldung	Schuldenbremse
Prospekte	Flyer
Selbstmord	Suizid

CORRELATION BETWEEN FREQUENCY CHANGE AND USAGE SIMILARITY: SPIEGEL/ZEIT & DEREKO

- Coefficients of Frequency Change

$$\ln\left(\frac{p(t)}{1-p(t)}\right) = k + s * t + c * t^2$$

k … Intercept: Start
s … Slope: Rate of Change
c … Curvature: Change of Slope

- Spearman Rank Correlation ρ
  - Strongest between Slopes of Nearest Neighbors (*NN*)
  - Curvature stronger than Intercept
  - Decreasing with increasing distance between neighbors

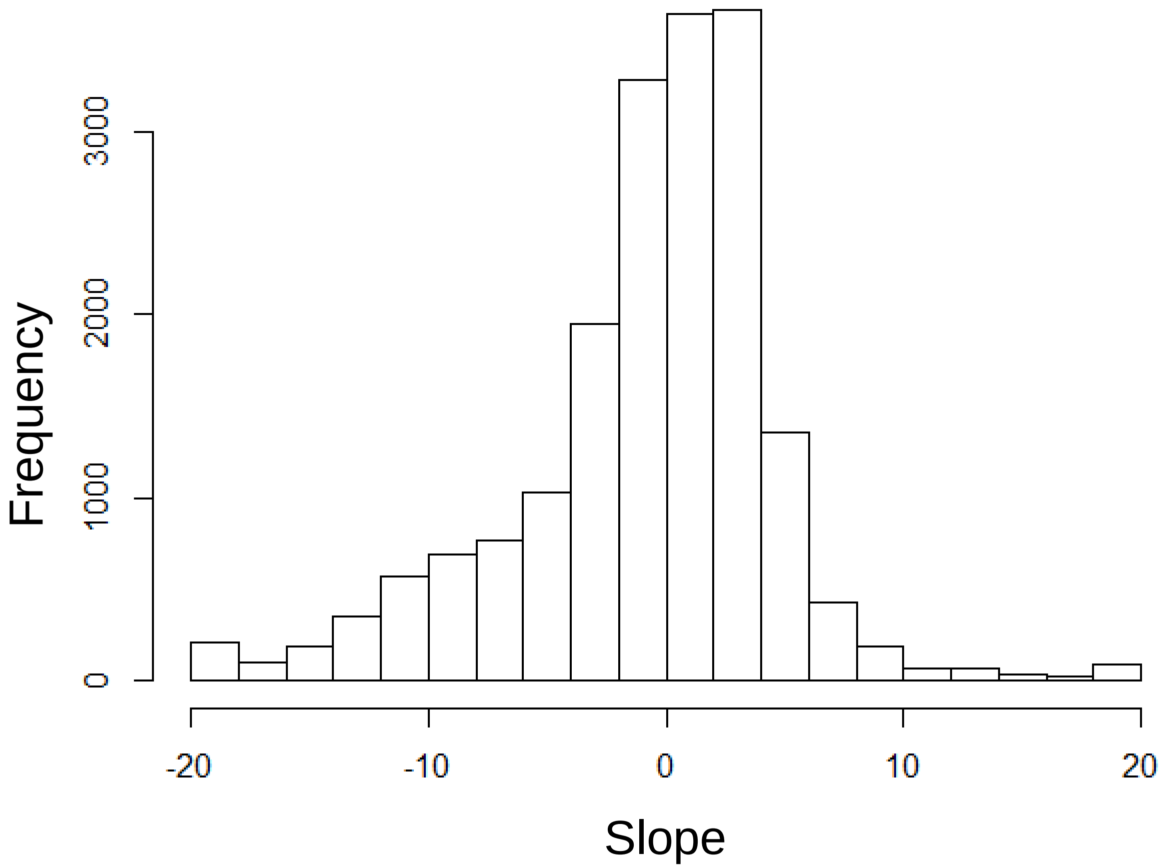
NN	*k*	*s*	*c*
1	0.33	0.43	0.40
2	0.27	0.39	0.33
3	0.25	0.34	0.27

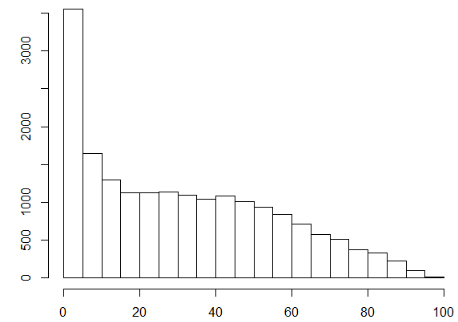
Spiegel/Zeit

NN	*k*	*s*	*c*
1	0.33	0.42	0.48
2	0.26	0.36	0.41
3	0.23	0.32	0.38

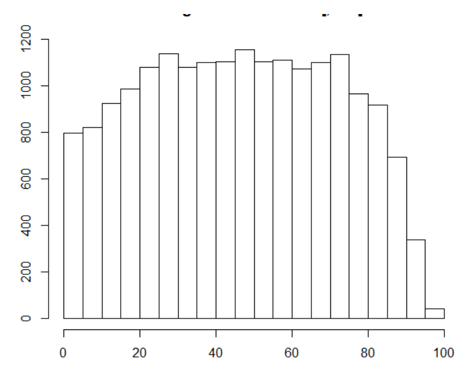
DeReKo

RSC: DISTRIBUTION OF SLOPES, ROOT MEAN SQUARE ERRORS (RMSE)





RMSE1: 31.1%



RMSE2: 46.2%