



Peter Fankhauser, Abteilung Zentrale Forschung

NEOLOGISMEN UND BEDEUTUNGSÄNDERUNG

Gemeinsam mit Marc Kupietz

Mitglied der

ÜBERSICHT

- Zwei Phänomene des lexikalischen Wandels
 - Bedeutungsänderung
 - Wortverdrängung
- Zwei Dimensionen zur Modellierung von Wortgebrauch, korpusbasiert
 - (Distributionielle) Semantik
 - Frequenzänderung
- Beispiele
- Abschließende Bemerkungen



Die Zeit ändert alles;
es gibt keinen Grund, warum
die Sprache diesem allgemeinen
Gesetz enthoben sein sollte.

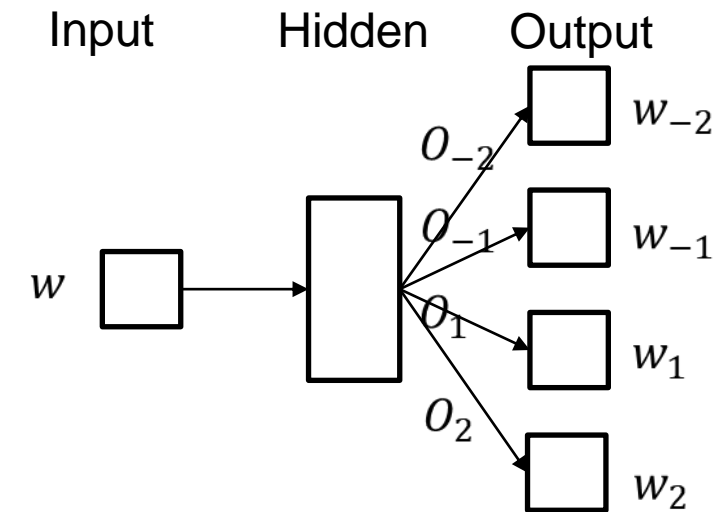
Ferdinand de Saussure,
Cours de linguistique générale
1916/1959

KORPUS: SPIEGEL + DIE ZEIT

- Zeitspanne: 1953 – 2015
- Intervalle: 13*5 Jahre (1953-1957, 1958-1962,...)
- Anzahl Tokens: 570 Mio
- Untersuchte Wörter (Types): Top 50.000 (Frequenz > 527)
- Einbettungsdimensionen: 200
- Visualisierung: <http://corpora.ids-mannheim.de/openlab/diaviz/zeitspiegel.html>

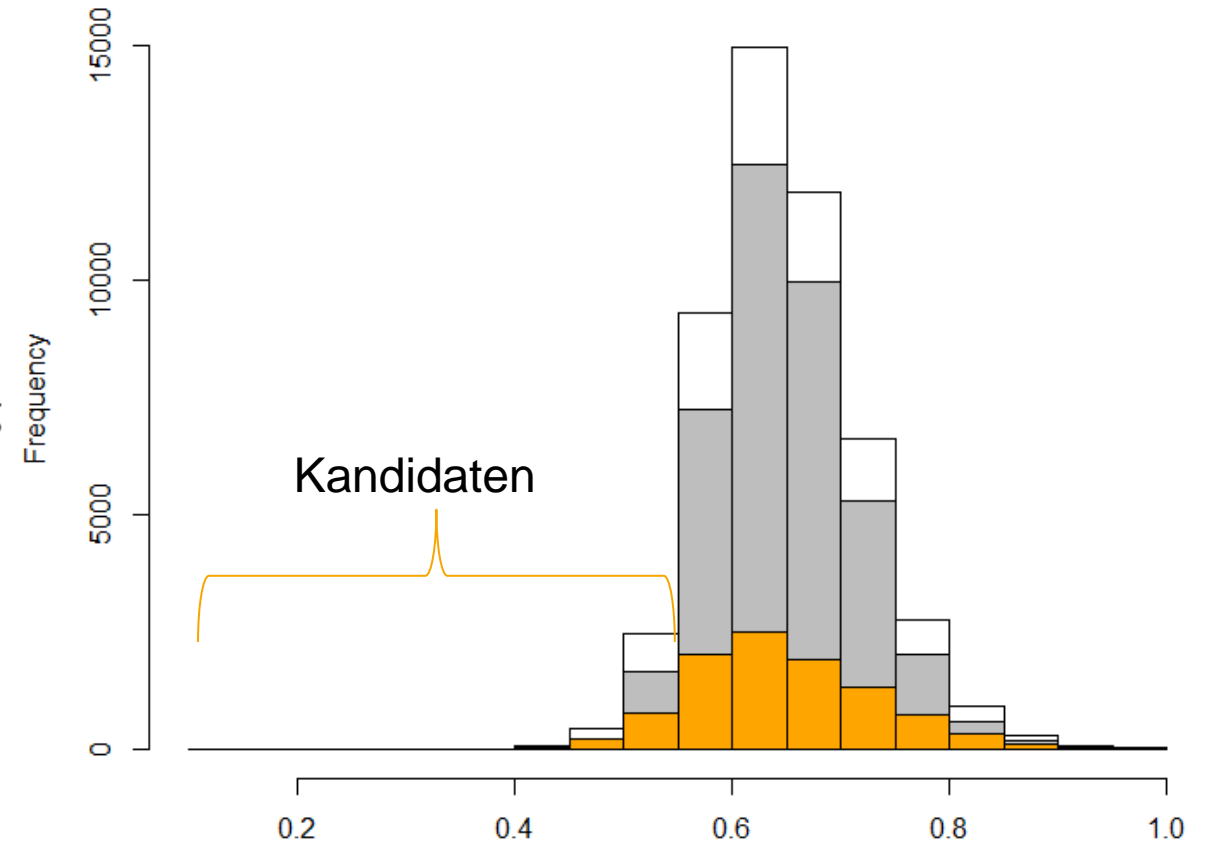
DISTRIBUTIONELLE SEMANTIK: REPRÄSENTATION VON WORTGEBRAUCH IN WENIGEN DIMENSIONEN

- Wort-Kookurrenz-Vektoren [Belica et al. 2010]
- Shifted $PMI(w_{-1}, w)$ [Levy, Goldberg 2014]
- Anzahl der Dimensionen =
Vokabulargröße (* Kontextgröße)
- Worteinbettungen
 - Strukturiertes Skipgram (Wang2Vec [Ling et al. 2015])
 - Lerne Abbildung von Wort w auf Links/Rechts-Kontext $(w_{-2}, w_{-1}, w_1, w_2)$
über Zwischenebene mit wenigen Dimensionen (100-200)
- Diachrone Worteinbettungen [Kim et al. 2014], [Hamilton et al. 2016], [Fankhauser, Kupietz 2017]
 - Beginne mit randomisiertem Neuronales Netz
 - Initialisiere Neuronales Netz im Intervall $t + 1$ mit NN im Intervall t



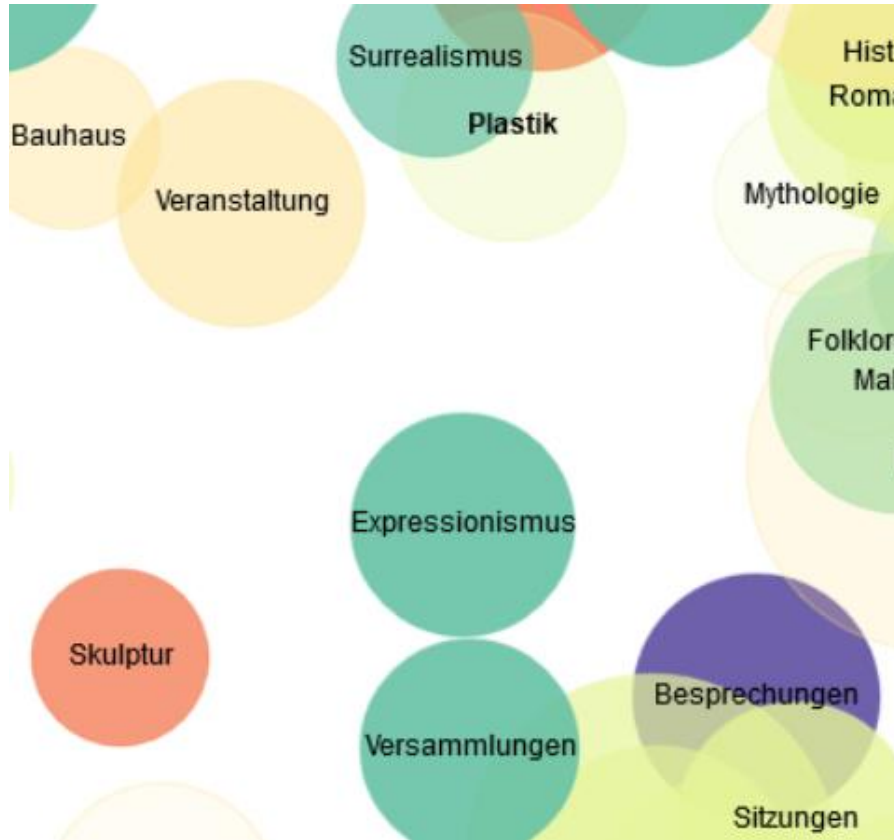
ERKENNUNG VON BEDEUTUNGSÄNDERUNG

- Vergleiche Worteinbettungen x_1, x_n
 - Intervall t_1 des ersten Auftretens (>25)
 - Intervall t_n des letzten Auftretens
- Cosinusähnlichkeit
 - $\frac{x_1 x_n}{\|x_1\| \|x_n\|}$
 - Geringe Cosinusähnlichkeit -> Bedeutungsänderung
- Caveats
 - Pro Intervall: Dominante (häufigste) Bedeutung
 - Für t_1 : Manchmal zu wenige Belegstellen
 - Eigennamen (~ 20%) ändern häufig ihre Bedeutung

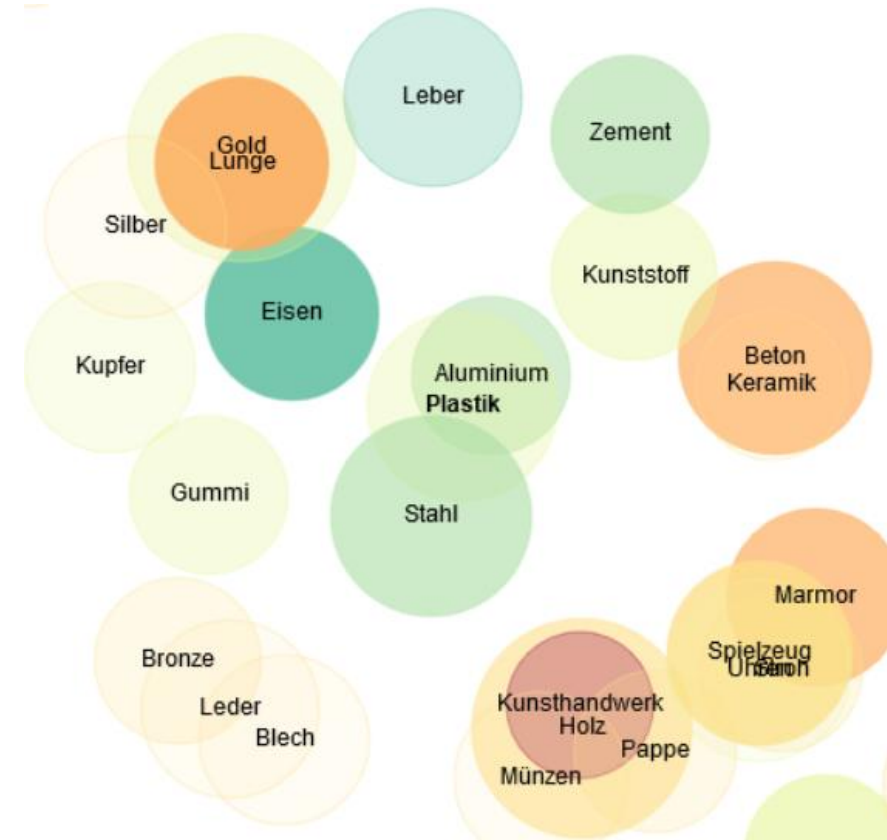


Verteilung Cosinusähnlichkeit für **Eigennamen** vs. Rest

PARADIGMATISCHE NACHBARSCHAFT VON PLASTIK ÜBER DIE ZEIT



1960



2015

BEDEUTUNGSÄNDERUNG: SEMANTISCHE VERSCHIEBUNG

Wort	Sim	Von	Nachbarn1	Anz1	Bis	Nachbarn2	Anz2
Portal	0.482	1958	Bahnsteig,Vorraum,Mausoleum	100	2015	Webseite,Netzwerk,Programm	263
Soli	0.496	1963	Bläser,Komödien,Bilderbogen	29	2015	Steuererhöhungen,Maut,Solidarpakt	85
Installationen	0.502	1968	Generatoren,Klimaanlagen,Anschlüsse	26	2015	Plastiken,Zeichnungen,Collagen	111
Installation	0.504	1958	Bebauung,Verkleinerung,Bewässerung	37	2015	Ausstellung,Installationen,Retrospektive	202
abgeschlagen	0.513	1958	abgewehrt,entwaffnet,überrollt	63	2015	zurückgefallen,Spitzenreiter,abgerutscht	97
Plastik	0.523	1958	Photographie,Baukunst,Gotik	228	2015	Edelstahl,Gummi,Kunststoff	449
Einstieg	0.525	1963	Absprung,Parcours,Auslauf	29	2015	Generationswechsel,Börsengang,Sprung	393
Ex	0.529	1958	Multimillionär,Colonel,Geheimdienstchef	49	2015	Scheidung,Partnerin,Freundin	137
vereinnahmt	0.531	1958	einbehalten,abgegolten,aufgestockt	46	2015	abgestempelt,aufgesogen,verhöhnt	57
Organ	0.542	1958	Zentralorgan,Wochenblatt,Parteiorgan	480	2015	Gehirn,Muskel,Organe	247
Aufschrei	0.560	1958	Stöhnen,Krampf,Seufzer	64	2015	Entrüstung,Aufruhr,Protest	287

BEDEUTUNGSÄNDERUNG: KONKRET VS. ABSTRAKT

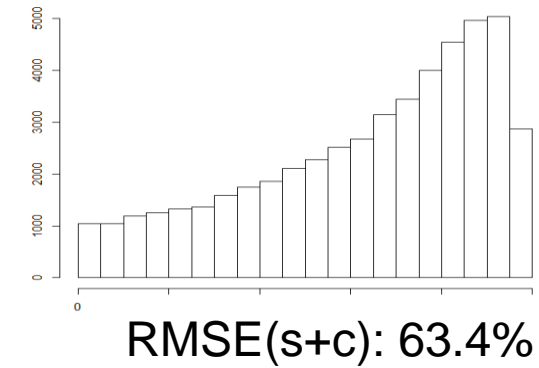
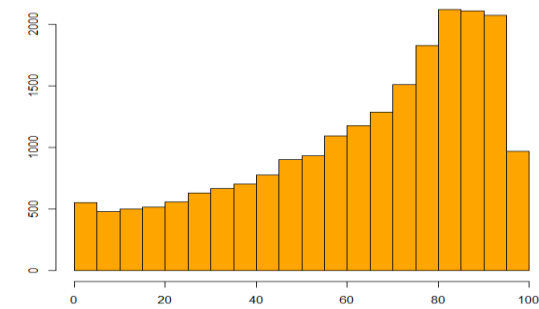
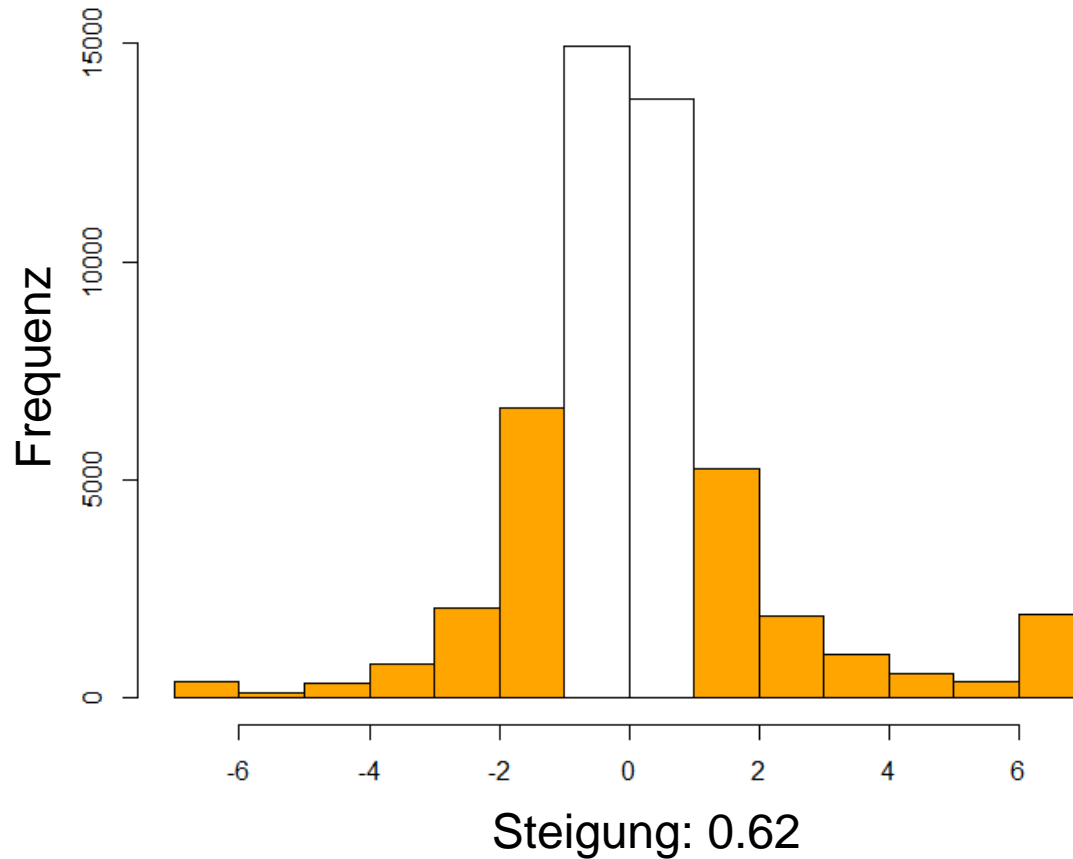
Wort	Sim	Von	Nachbarn1	Anz1	Bis	Nachbarn2	Anz2
Exzellenz	0.508	1958	Hoheit,Schwägerin,Gastgeberin	99	2015	Internationalität,Spitzenleistungen,Effizienz	114
Ikonen	0.522	1963	Vitrinen,Exponate,Pelze	69	2015	Ikone,Idole,Symbole	128
Filz	0.529	1963	Tüchern,Lehm,Tücher	31	2015	Klüngel,Korruption,Beton	107
Selbstbedie- nungsladen	0.557	1958	Plattenspieler,Fachhandel,Speicher	35	2015	Spielwiese,Armutszeugnis,Anachronismus	29
Prüfstand	0.485	1963	Werksgelände,Rennstrecke,Bohrloch	31	2015	Tagesordnung,Pranger,Lot	192
Rüstungen	0.495	1958	Autarkie,Machtmittel,Organisationsformen	37	2013	Mützen,Gewänder,Hüte	43
Illustration	0.521	1958	Neuausgabe,Widerlegung,Charakterisierung	108	2015	Karikatur,Foto,Illustrationen	207
Baumarkt	0.521	1958	Wohnungsmarkt,Geldmarkt,Arbeitsmarkt	77	2015	Wasserhahn,Sperrmüll,Kiosk	172
Übergewicht	0.536	1958	Reservat,Vordringen,Staatsgebiet	159	2015	Diabetes,Rheuma,Asthma	228
E-Mail	0.540	1993	E-Mails,Internet,Computer	297	2015	SMS,Mail,Brief	2036

BEDEUTUNGSÄNDERUNG: WORTKLASSEN, GRAMMATIKALISIERUNG

Wort	Sim	Von	Nachbarn1	Anz1	Bis	Nachbarn2	Anz2
verkommen	0.460	1958	überlisten,herausnehmen,impfen	28	2015	mutiert,degradiert,mutieren	237
angesagt	0.487	1958	geritten,gefeuert,angetragen	60	2015	hip,ausgebrochen,wichtig	250
gezielt	0.506	1958	zurückzuhalten,zugehört,umgegangen	81	2015	verstärkt,routinemäßig,wahllos	989
ferner	0.542	1958	überdies,demnach,demzufolge	1262	2015	näher,absehbarer,ähnlicher	243
vermessen	0.551	1958	auszudenken,überspitzt,deprimierend	69	2015	durchleuchten,analysieren,erfassen	322
Hauptsache	0.506	1958	Endeffekt,Folgezeit,Quantität	291	2015	denn,geil,Aber	518
Verglichen	0.539	1958	Gesucht,profitierte,Unterstützt	45	2015	verglichen,Gegensatz,Vergleich	171
Vorfeld	0.540	1958	Staatsgebiet,Machtbereich,Kriegsschauplatz	53	2015	Beginn,Bekanntwerden,Zuge	411
Stande	0.553	1958	Zeitgeist,Aussterben,Umbruch	66	2008	Stocken,zuwege,Lot	326
Kippe	0.559	1968	Kloake,Müllkippe,Wanne	40	2015	Hand,Bredouille,Klemme	147



SPIEGEL/ZEIT: VERTEILUNG VON STEIGUNG, PSEUDO-RMSE



ERKENNUNG VON WORTVERDRÄNGUNG

- Im letzten Intervall (2013-2015)
 - Für jedes Wort w
 - Bestimme nächste k Nachbarn w_k (nach Cosinus-Ähnlichkeit)
 - Vergleiche Steigungskoeffizienten s, s_k
 - Sortiere Wortpaare w, w_k nach $s * s_k$
 - Entgegengesetzte Steigung: Wortverdrängung
 - Gleiches Vorzeichen: Parallele Entwicklung
- Normalfall: Steigung von nächsten Nachbarn sind positiv korreliert
N1: 0.41, N2: 0.36, N3: 0.35

PARADIGMATISCHE NACHBARSCHAFT VON: RATIONALISIERUNG/DIGITALISIERUNG



1960



2015

WORTVERDRÄNGUNG (1)

Orthographie

Sinkt	Steigt
Mikrophon	Mikro
Konsensus	Konsens
Paragraph	Paragrafen
Anläßlich	anlässlich
placiert	platziert

Eigennamen

Sinkt	Steigt
Strauß	Stoiber
Chruschtschow	Putin
KPD	PDS
Bundespost	Telekom
Harrisburg	Fukushima

WORTVERDRÄNGUNG (2)

Neue Dinge

Sinkt	Steigt
Unterredung	Telefonat
Telephongespräche	E-Mails
Mitteilungen	Mails
Telephon	Handy
brieflich	telefonisch
telephonieren	surfen
Konstrukteur	Entwickler
Couturiers	Labels

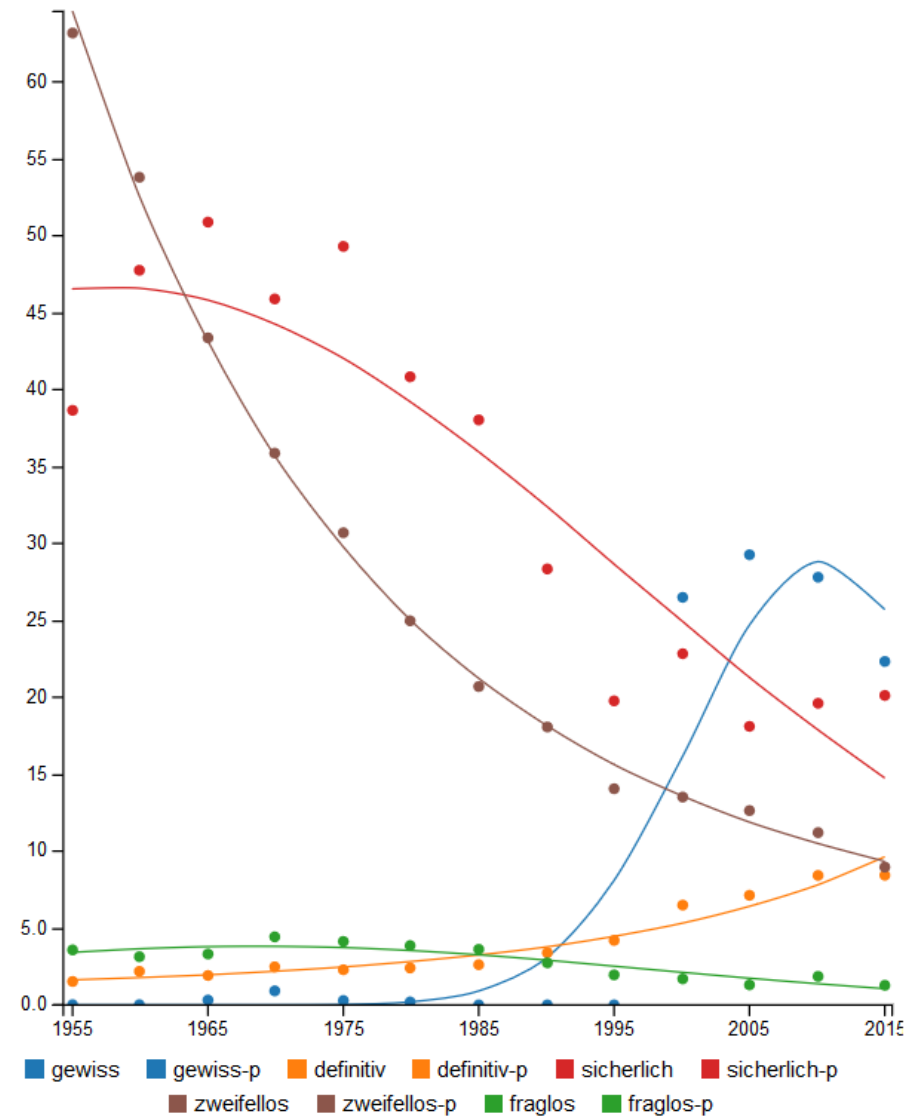
Spinning/Framing

Sinkt	Steigt
Rationalisierung	Digitalisierung
rationalisieren	optimieren
Kompromiß	Deal
Zusätze	Aromen
Selbstmord	Suizid

WORTVERDRÄNGUNG (3)

Stilistisch

Sinkt	Steigt
vermittels	mithilfe
unzweideutig	unmissverständlich
zuungunsten	zulasten
Gegenwärtig	Zurzeit
fraglos	gewiss
sicherlich	gewiss
eindeutig	definitiv



ABSCHLIEßENDE BEMERKUNGEN

- Zwei Varianten von lexikalischen Wandel
 - Bedeutungsänderung: Unterschiedlicher Wortgebrauch über die Zeit
 - Wortverdrängung: Alternatives Wort für ähnliche Bedeutung über die Zeit
- Anwendungen
 - Mikroanalyse
 - Exploration: Auffinden von Kandidaten mit Evidenz in Korpus
 - Verifikation: Überprüfen von Kandidaten
 - Makroanalyse
 - Visuelle Korrelation von Bedeutung mit Frequenzentwicklung
 - Verteilung von Maßen im Registervergleich
 - Bedingungen für Bedeutungsänderung und Wortverdrängung: paradigmatische/syntagmatische Produktivität(?)
- Weitere Arbeiten
 - Vergleich mit/Evaluierung auf Neologismen-Liste
 - Leistungsfähigere Word-Embeddings (Fasttext, Elmo, BERT)

REFERENZEN

- (1) Belica, C., Keibel, H., Kupietz, M., Perkuhn, R. (2010): An empiricist's view of the ontology of lexical-semantic relations. In Storjohann, P. (ed.): Lexical-Semantic Relations. Theoretical and practical perspectives. Amsterdam: Benjamins, 115-144.
- (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 (2016). 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) Levy, O. & Goldberg Y. (2014). Neural Word Embedding as Implicit Matrix Factorization. In Advances in Neural Information Processing Systems 27 (NIPS 2014).
- (6) Ling, W., C. Dyer, A. Black & I. Trancoso (2015). Two/Too Simple Adaptations of Word2Vec for Syntax Problems. Human Language Technologies (NAACL HLT 2015)