TECHNISCHE UNIVERSITÄT DRESDEN

Dresden Center for Digital Linguistics

4472657364656e2043656e74657220666f72204469676974616c204c696e6775697374696373

http://linguistik.zih.tu-dresden.de

KORPUSPRAGMATIK

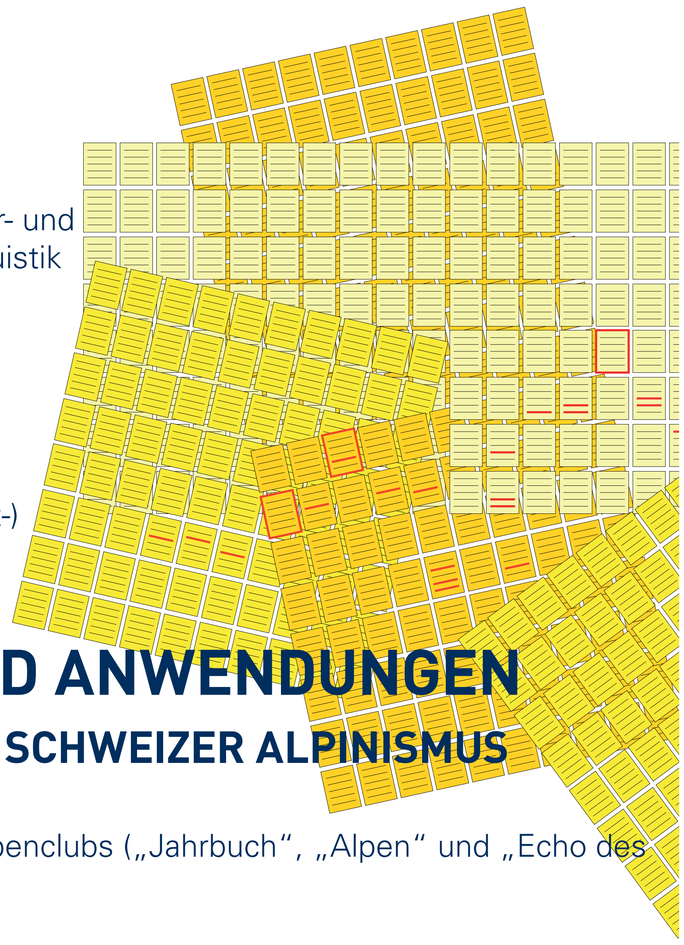
METHODEN UND WERKZEUGE FÜR DIE SOZIAL- UND KULTURWISSENSCHAFTLICH INTERESSIERTE LINGUISTIK

Noah Bubenhofer, Joachim Scharloth, Dresden Center for Digital Linguistics

HYPOTHESEN

Korpuslinguistik identifiziert wiederkehrende Muster des Sprachgebrauchs. In der Kultur- und sozialwissenschaftlichen Linguistik werden Muster mit kulturellen oder sozialen Phänomenen in Zusammenhang gebracht…

- …als Symptom für diese Phänomene;
- ...als diese Phänomene (mit-) konstituierend.



METHODEN UND ANWENDUNGEN

ZEITGEISTANALYSEN SCHWEIZER ALPINISMUS

Korpus

- Periodika des Schweizer Alpenclubs („Jahrbuch“, „Alpen“ und „Echo des Alpes“) von 1864 bis heute.
- Gesamtkorpus:

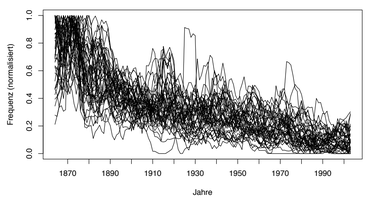
	Tokens	Types	Buchseiten
SAC Jahrbücher, Die Alpen	37,16 Mio.	900.000	87.000
Echo des Alpes	7,45 Mio.	175.000	22.500

- Untersuchungskorpus: 17.116.833 Tokens deutschsprachige Texte mit Mindestlänge 2000 Wörter.

Berechnungsmethode

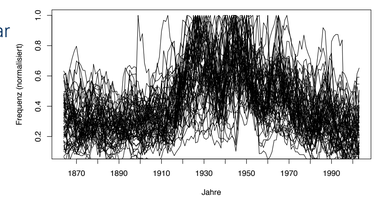
- Berechnung der Frequenzen aller Lexeme (Nomen, Verben, Adjektive, Personalpronomen) pro Jahr im Korpus.
- Auswahl der Lexeme, deren Verteilung über die Jahre ungleichmäßig ist: Verwendung von Gries‘ Deviation of Proportion (DP, Gries 2008a, 2009).
- Clustering der Frequenzverläufe mit dem Ziel, Lexeme mit ähnlichen Entwicklungslinien zu identifizieren und das Korpus datengeleitet zu periodisieren (hierarchisches Clustering, ‚ward‘, 60 Gruppen).
- Visualisierung der Clustergruppen als Wortwolken.

Cluster 23: Typisches Vokabular für die 1870er und 1880er-Jahre (Frequenzverläufe und Wortwolke)



Nomen (21) Adjektive (20) Verben (3) Personalpronomen (0)

Cluster 42: Typisches Vokabular für die 1930er und 1950er-Jahre (Frequenzverläufe und Wortwolke)



Nomen (36) Adjektive (14) Verben (13) Personalpronomen (0)

IDENTIFIZIERUNG UND ANALYSE VON UMBRUCH-ZEITEN IN DEUTSCHLAND SEIT 1946

Ziel

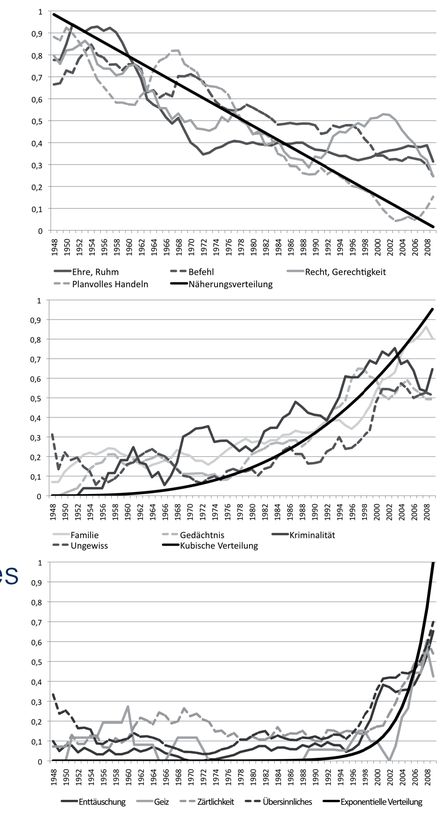
- Veränderung und Vernetzung von Deutungsrahmen („Frames“ in Anschluss an Goffman 1974).
- Datengeleitete Berechnung von sprachlichen Umbrüchen (Kämper 2007) als Indikator für zeitgeschichtliche Umbrüche.

Korpus

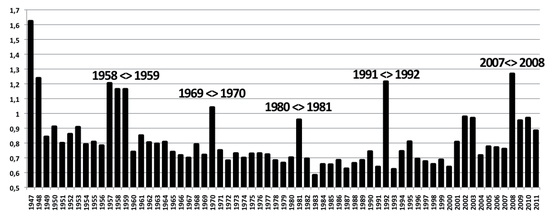
- Wochenzeitung „DIE ZEIT“, 1946 bis 2011, 271.439.149 Tokens.
- POS-Tagging, Stuttgart Tübingen Tagset (STTS).

Berechnungsmethode

- Automatische Frameanalyse auf der Basis des „Deutschen Wortschatz nach Sachgruppen“ (Dornseiff 2004).
- Identifizierung typischer Frequenzverlaufskurven mittels Berechnung der euklidischen Distanz zwischen Näherungsverteilungsvektor und allen Framevektoren (Grafiken rechts).
- Berechnung von Umbrüchen: Für jedes Jahr wird der Betrag der Veränderung der relativen Frequenzen aller Frames im Vergleich zum Vorjahr berechnet und summiert. Stabilität = nur geringfügige Veränderungen im Framehaushalt; Umbruch = ein Teil der Frames nimmt stark zu, ein anderer Teil ab.

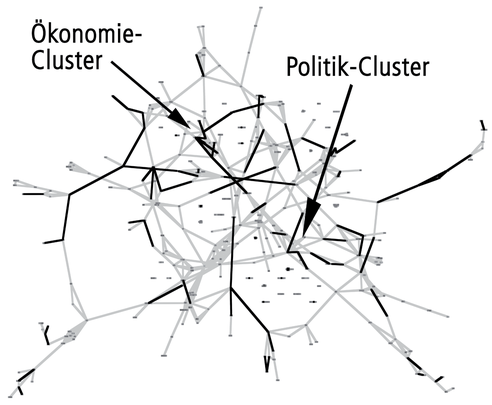


Typische Verlaufskurven von Frames: Annähernd stetig abnehmend, kubisch, exponentiell (von oben nach unten)

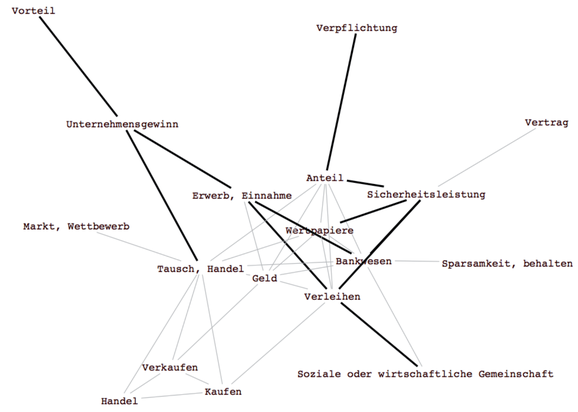


Umbrüche im Framehaushalt

- Berechnung von überzufällig gemeinsam vorkommenden Frames (Frame-Kollokationen) und Visualisierung als Kollokationsgraph.
- Separate Berechnung der typischen Framekollokationen in zwei Zeitabschnitten (vor und nach Umbruch), Visualisierung in einem Kollokationsgraphen, um Cluster visuell zu entdecken, die für Veränderungen im Framehaushalt stehen.



Framekollokationsgraph Umbruch 2004-2007 vs. 2008-2011: Graue Kanten stehen für die Zeit vor dem Umbruch, schwarze Kanten sind typisch für die Zeit des Umbruchs/ nach dem Umbruch. Verdichtungen verweisen auf Framecluster, die weiter analysiert werden können.



Detailansicht des Ökonomie-Clusters: Die mit dem Umbruch neuen Framekollokationen zeigen den Wandel des Wirtschaftssystems: Während die Wirtschaft vorher mit den Frames ‚Handel‘, ‚Kaufen‘, ‚Verkaufen‘, ‚Geld‘, ‚Markt, Wettbewerb‘ beschrieben wird, geht es seit der Bankenkrise um ‚Verleihen‘ (Kreditgeschäfte), ‚Unternehmensgewinn‘, ‚Vorteil‘ und die ‚soziale oder wirtschaftliche Gemeinschaft‘.

Dresden Center for Digital Linguistics
Institut für Germanistik
Noah Bubenhofer

E-Mail: noah.bubenhofer@tu-dresden.de
Web: http://linguistik.zih.tu-dresden.de
Mommsenstr. 13
D-01062 Dresden

Literatur

Bubenhofer, Noah u. a. (Hrsg.) (2011): Text+Berg-Korpus (Release 145), Institut für Computerlinguistik, Universität Zürich. http://www.textberg.ch
Bubenhofer, Noah und Joachim Scharloth (im Druck): „Korpuspragmatische Methoden für kulturanalytische Fragestellungen“, Linguistik als Kulturwissenschaft: Korpus - Kommunikation - Kultur, Giessen Contributions to the Study of Culture, Trier: Wvt Wissenschaftlicher Verlag.
Dornseiff, Franz (2004): Der deutsche Wortschatz nach Sachgruppen, Berlin, New York.
Goffman, Erving (1974): Frame Analysis, Garden City/N.Y.: Doubleday.
Gries, Stefan Thomas (2008): „Dispersions and adjusted frequencies in corpora“, International Journal of Corpus Linguistics 13, S. 403–437 (35).
Gries, Stefan Thomas (2009): „Dispersions and adjusted frequencies in corpora: further explorations“, Language and Computers 71/1, S. 197–212.
Kämper, Heidrun (2007): „Linguistik als Kulturwissenschaft. Am Beispiel einer Geschichte des sprachlichen Umbruchs im 20. Jahrhundert“.
Scharloth, Joachim, David Eugster und Noah Bubenhofer (im Druck): „Das Wuchern der Rhizome. Linguistische Diskursanalyse und Data-driven Turn“, in: Busse, Dietrich und Wolfgang Teubert (Hrsg.): Linguistische Diskursanalyse. Neue Perspektiven, Wiesbaden: VS Verlag.