

# Das Deutsche in mehrsprachigen europäischen Öffentlichkeiten

Programmatik der Vergleichenden Diskurslinguistik  
und Anforderungen an offene digitale Forschungsinfrastrukturen

# Gliederung

- (1) Europäische Öffentlichkeiten
- (2) Vergleichende Diskursanalysen
- (3) Vergleichen: Erkenntnisprozesse & Praktiken
- (4) Transformationen: Inkommensurabilität und Kommensurabilität der Diskurse
- (5) Auswahl und Verarbeitung von Daten
- (6) Methoden
- (7) Analysetools
- (8) Anforderungen an offene digitale Forschungsinfrastrukturen
- (9) Fazit

# Warum wir zu diesem Thema sprechen



MORCDA - Making Open Research Data Suitable  
for Comparative Discourse Analysis



# Europäische Öffentlichkeiten

# SRG-Initiative

Eidg. Vorlage: Volksinitiative «200 Franken sind genug!»

Endresultat von 08.03.2026

JA

38.1%

1'188'459 Stimmen

Standesstimmen

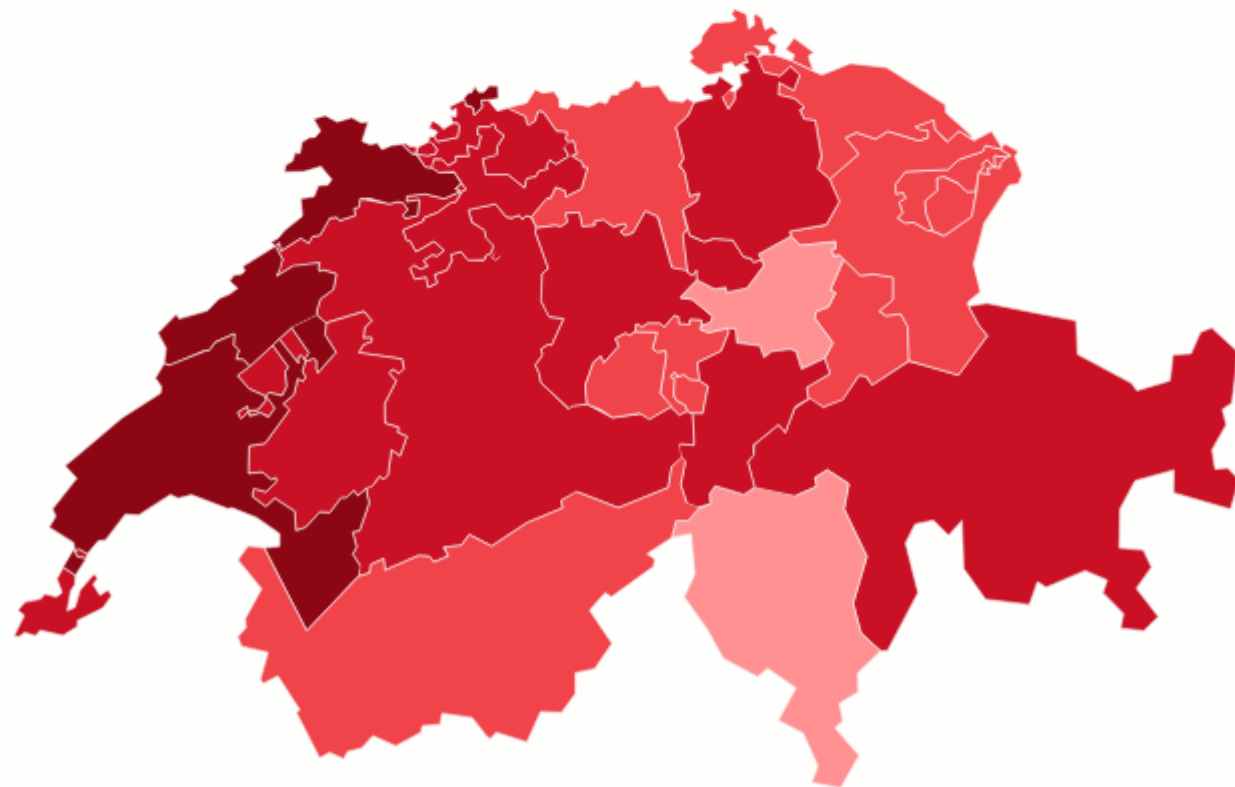
0

NEIN

61.9%

1'934'819 Stimmen

23



<https://www.srf.ch/news/schweiz/abstimmungen-8-3-2026/abstimmungen-vom-8-maerz-srg-initiative-scheitert-am-volks-und-staendemehr>

# Gibt es europäische Öffentlichkeiten?

- öffentlich-rechtlicher Rundfunk wird in vielen europäischen Staaten infrage gestellt
- Diskurse über öffentlich-rechtlichen Rundfunk in Europa als Beispiel für einen Gegenstand der Vergleichenden Diskursanalyse
- Journalistische Öffentlichkeiten werden insbesondere länderspezifischen hergestellt
- transnationale, translinguale, transkulturelle, transregionale etc. Öffentlichkeiten fehlen in Europa
- Diskurse enden in der Regel nicht an Landes- oder Zeitungsgrenzen

# Vergleichende Diskursanalysen

# Was ist Vergleichende Diskurslinguistik?

Vergleichende Diskursanalyse wird eingesetzt, wenn mindestens zwei Diskurse oder Diskursausprägungen in Beziehung gesetzt werden zum Zweck, Unterschiede und Gemeinsamkeiten in Wissens- und Machtordnungen bestimmen zu können.

Was ermöglicht die Vergleichende Diskurslinguistik?

- Für die Linguistik z.B.: Hinterfragung von z.B. medienspezifischen oder thematischen Diskursannahmen durch Kontrasterfahrungen; Auseinandersetzung mit Praktik des Vergleichens
- Für die Berufspraxis z.B.: Organisationen agieren mit interkulturellen Teams; sie müssen in einer globalen Öffentlichkeit ihre Stakeholder adressieren
- Lehre: internationale Germanistik kann auf Vorwissen der Studierenden aufbauen und verstärkt auf digitale Verfahren des Vergleichens setzen [?]
- Für Gesellschaften in Europa: Meinungsbildung und europäische Integration durch grenzüberschreitende Öffentlichkeit

# Vergleichen wird komplizierter und leichter zugleich

These 1: Vergleichen von Diskursen ist kompliziert, weil Diskurse kaum zu fassende Theoriegrößen sind, die sich in beispielsweise Sprache, Medien, Zeiträumen, beteiligten Akteuren und Ereignissen unterscheiden können.

These 2: Vergleichen von Diskursen wird einfacher, weil Diskurse digital in Korpora erfasst und mithilfe von statistischen und hermeneutischen Verfahren analysiert werden können.

# Vergleichen: Erkenntnisprozesse & Praktiken

# Vergleichen als Untersuchungsgegenstand

Semantiken und Praktiken von Vergleich in der Gesellschaft:

In-Verbindung-bringen; In-Beziehung-Setzen; Gleichsetzen; Gleichstellen; als unvergleichbar deklarieren; Einzigartigkeit konstruieren, semantische Nuancen zwischen *unvergleichbar* und *unvergleichlich* verhandeln, ...

# Vergleichen als Normalfall

Vergleichen ist ein alltägliches kognitives Verfahren, das wir alle seit unserer frühesten Kindheit nutzen: Ball ≠ Puppe, Puppe ≠ Teddy etc.

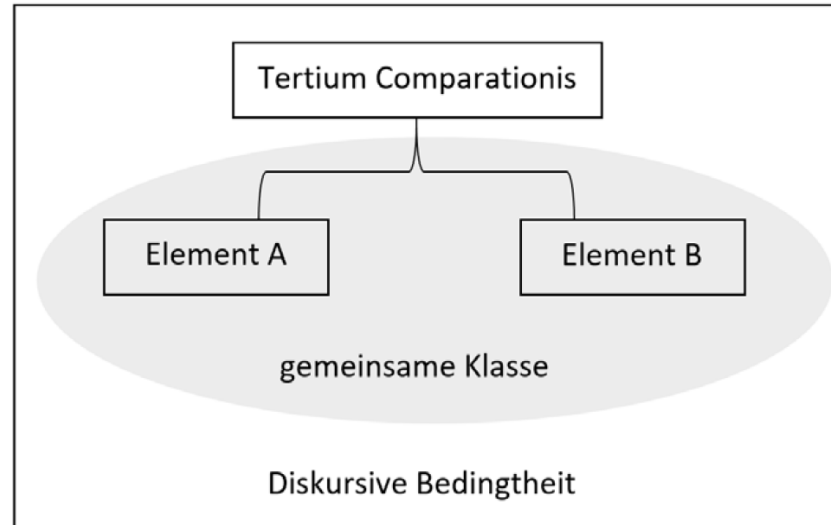
Die meisten Repräsentativa können sinnhaft nur aufgrund von Vergleichen geäußert werden: *Er ist liberal* hat folgende diskursive Bedingung:

Fast immer ist da impliziert: anderswo ist es anders. Sehr viele Aussagen also, sowohl in der Alltagssprache als auch im wissenschaftlichen Diskurs, enthalten faktisch einen Vergleich

Gauger, H.-M. (2000): „Der Vergleich in der Sprachwissenschaft“. In: Zima, P. V., Kacianka, R. & Strutz, J. (Hrsg.): *Vergleichende Wissenschaften. Interdisziplinarität und Interkulturalität in den Komparatistiken*. Tübingen: Narr, 149–172, hier: S. 150.

# Vergleichen als Methodologie

Aus diskursanalytischer Sicht beruht jeder Vergleichsprozess auf fünf Elementen:



Ein Apfel und eine Birne sind nur dann vergleichbar, wenn wir erstens von einer gemeinsamen Klasse (z.B. Frucht, Obst, Lebensmittel) ausgehen und sie unter mindestens einem *tertium comparationis* (z.B. Form, Preis, Vitamingehalt) betrachten. Wir können also festhalten: Entgegen der landläufigen Meinung lassen sich sehr wohl Äpfel mit Birnen vergleichen, sofern wir sie derselben Klasse zuordnen und sofern mindestens ein sinnvolles *tertium comparationis* verwendet wird.

# Vergleichen als machtvolle Praktik

„Nicht der Vergleich, sondern das Vergleichen, das *doing comparisons*, [...] im Zentrum stehen [muss], wenn verstanden werden soll, wie wir unsere Welt nicht allein beobachten und erschließen, sondern auch verändern.“

Grave, J. (2015): „Vergleichen als Praxis. Vorüberlegungen zu einer praxistheoretisch orientierten Untersuchung von Vergleichen“. In: Epple, A. & Erhart, W. (Hg.): *Die Welt beobachten. Praktiken des Vergleichens*. Campus, 135–159; hier 147; Herv. im Orig.).

- Vergleichen verändert Weltwahrnehmungen.
- Man kann sich gegen Vergleiche nicht wehren.

# Transformationen: Inkommensurabilität und Kommensurabilität der Diskurse

# Digitalität

- Systemtheoretisch und damit funktional aufgefasst, reduziert Digitalität zunächst gesellschaftliche Komplexität: analoger Aspektreichtum von Zeichen wird digital verarmt.
- Die einfache binäre Logik standardisiert z.B. Sprachgebrauch und Korpora, wodurch Vergleiche einfacher werden.

Hacking, I. (1982). Biopower and the Avalanche of Printed Numbers. *Humanities in Society* 5, 279–195.

Nassehi, A. (2019). *Muster: Theorie der digitalen Gesellschaft*. C.H. Beck.

# Singularität

## Vergleichslogik des Allgemeinen

- Beispiele: Einkommen, Schulnoten, Arbeitsleistung
  - geringe Komplexität der Einheiten; Differenz nach außen; eindeutige Vergleichsparameter
- 

## Vergleichslogik des Besonderen (seit der Moderne)

- Beispiele: Glücklichkeit, Lebensläufe; TV-Serien
- hohe Binnenkomplexität der Einheiten, uneindeutige Vergleichsparameter

Diskurse sind Singularitäten, sind also inkommensurabel.

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a Social Process. *Annual Review of Sociology*, 24, 313–343.

Fourcade, M. (2016). Ordinalization: Lewis A. Coser Memorial Award for Theoretical Agenda Setting 2014. *Sociological Theory*, 34(3), 175–195. <https://doi.org/10.1177/0735275116665876>

Reckwitz, A. (2021). *Die Gesellschaft der Singularitäten: Zum Strukturwandel der Moderne* (5. Auflage, wissenschaftliche Sonderausgabe). Suhrkamp.

# Transformation der Singularität durch Digitalität [?]

Doing Singularity: Die Logik des Besonderen braucht standardisierte, digitale Vergleichstechnologien, um Singularitäten (z.B. Diskurse) sichtbar, bewertbar und verwertbar zu machen.

Digitalisierung schafft die Voraussetzungen dafür, dass Singularitäten einen Wert erhalten können.

Reckwitz, A. (2021). *Die Gesellschaft der Singularitäten: Zum Strukturwandel der Moderne* (5. Auflage, wissenschaftliche Sonderausgabe). Suhrkamp.

# Vergleichende Diskursanalyse: Methodologische Herausforderungen und Forschungsinfrastrukturen

# Open Science und Europäische Daten

- Transformation durch **Open Science**: transnationale europäische Infrastrukturen & Daten



- Virtual Language Observatory
- CLARIN Resource Families
- Federated Content Search
- Language Resource Switchboard



# Welche Rolle spielen Forschungsinfrastrukturen für die Vergleichende Diskursanalyse (und andersherum)?

- Praktisch-methodische Relevanz: Infrastrukturen senken Eintritts- und Opportunitätskosten vergleichender Analysen (z.B. unter Aspekten der Ressourcenknappheit und fehlenden technischen Kompetenzen)
- Strukturelle Passung: In ihrer Grundlogik sind Forschungsinfrastrukturen europäisch ausgerichtet, netzwerkorientiert und auf Interoperabilität sowie grenzüberschreitende Nachnutzung angelegt (etwa im Umfeld von CLARIN, SSH-/EOSC-Ökosystemen)
- Unter einer Vergleichsperspektive werden die komplexen Herausforderungen von Infrastrukturen besonders evident

Welche Anforderungen ergeben sich für linguistische Infrastrukturen, wenn sie für vergleichende Diskursanalysen genutzt werden sollen?



Welche methodologischen Herausforderungen bestehen, wenn vergleichende Diskursanalysen korpusbasiert durchgeführt werden?

# Vergleichende Diskursanalysen mit Korpora: Methodologische Herausforderungen

# Herausforderungen der vergleichenden Diskursanalyse mit Korpora

Herausforderungen bestehen in Bezug auf

... die Auswahl von Sprachdaten/Korpora

...die Aufbereitung von Sprachdaten zu Korpora

... die Anwendung von Methoden der Korpusanalyse

# Auswahl von Daten/Korpora

# I. Auswahl von Daten: Datenäquivalenz

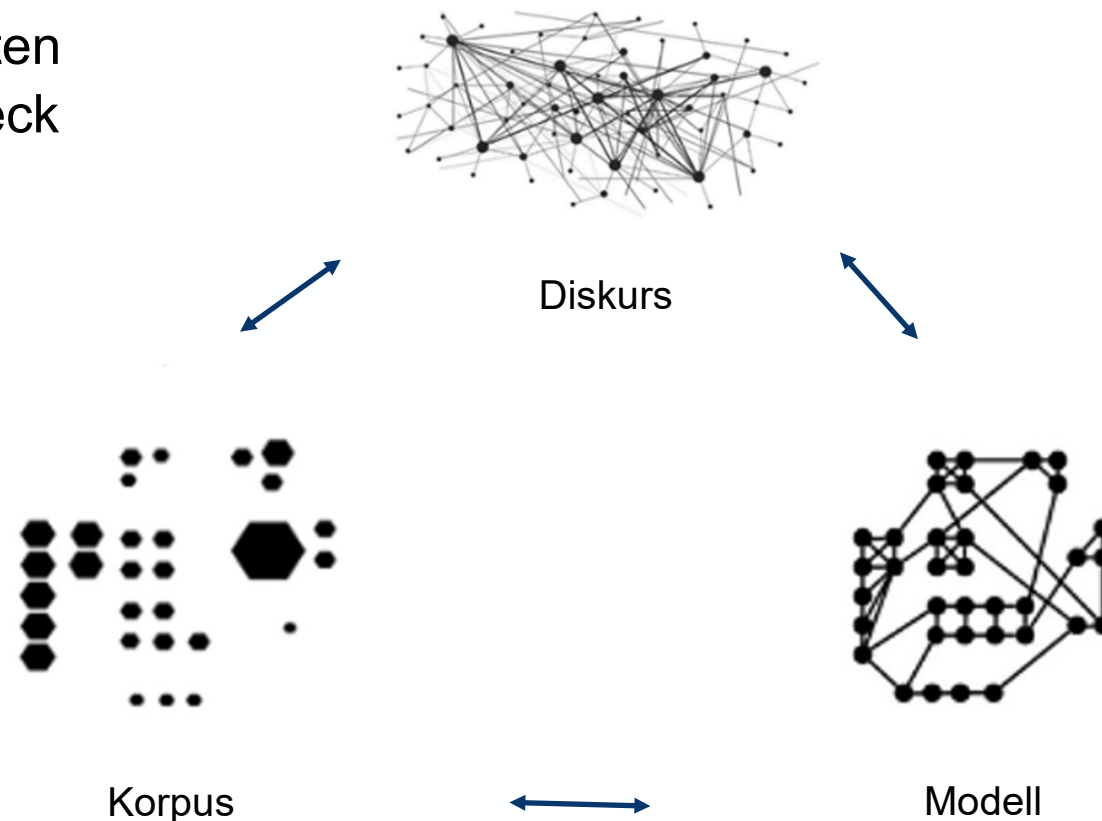
Notwendige Voraussetzung: Untersuchung äquivalenter Datensätze

Datenäquivalenz: der Grad, zu dem zwei (oder auch mehrere) Datensätze (Korpora) für eine konkrete Fragestellung so aufeinander bezogen werden können, dass Unterschiede in den Ergebnissen **plausibel als diskursiv oder inhaltlich interpretiert** werden können und **nicht primär als Artefakt von Modellierung, Datenerhebung, Aufbereitung, Methodenwahl oder infrastrukturellen Bedingungen**.

# Vom Diskurs zum Korpus: Modellierung

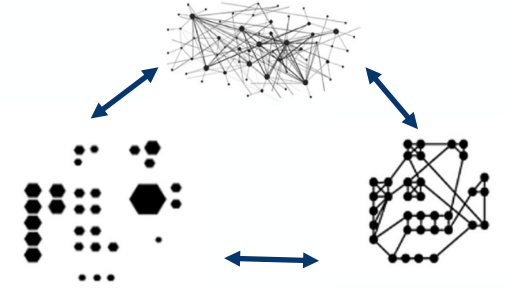
- Modell (Stachowiak, 1973; vgl. Epstein 2008)
  - Abbild eines Originals
  - Abstraktion/Verkürzung von Eigenschaften
  - für einen spezifischen Verwendungszweck
- Modellparameter
  - Akteure
  - Zeiträume
  - Schlüsselwörter
  - Sprache(n)
  - ...

Stücheli-Herlach et al. (2022)  
Dreesen et al. (in Vorb.)



# Vom Diskurs zum Korpus: Modellierung

- Herausforderung: Modellierung und Datenäquivalenz
- ein Modell vs. zwei Modelle
- Relation der Modelle zueinander (vgl. unilaterale und bilaterale Vergleichsmethoden (Czachur 2020, S. 207))
- bewusste oder unbewusste Dominanz eines Modells gegenüber anderen Modellen
- Beispiele aus Forschungsarbeiten:
  - Suremann (in Vorb.), Vergleich Facebook-Kommentare Irland mit Leser\*innenkommentaren Schweiz
  - Eysette (2025), Taylor (2014): Vergleich von Diskursen, die in unterschiedlich funktionierenden Mediensysteme stattfinden
  - Truan (2019) zu institutionellen Besonderheiten bei der Erstellung von vergleichenden Parlamentskorpora



# Vom Diskurs zum Korpus: Modellierung

- Herausforderung: Suchwörter im semantischen Kernbereich bestimmen
- Dokumente gelten als „diskursrelevant“, wenn definierte Suchwörter vorkommen
  
- Herausforderung aus Perspektive Vergleich:
  - diskurs-/kulturspezifische Suchwörter identifizieren
  - formale statt funktionale Äquivalenz: Fokus auf formal äquivalente Lexikalisierungen (z.B. Kognaten, Internationalismen, Entlehnungen) statt funktional äquivalenter Begriffe (z.B. Schröter/Storjohann (2013): *financial crisis* vs. *Wirtschaftskrise*)
  
- Lösungsansatz: Distributionelle Semantik und Word Embeddings



# Auswahl von Daten: Zugangsbeschränkungen

Herausforderung: Zugangsbedingungen schränken die Äquivalenz von Daten ein

1. rechtliche und institutionelle Zugangsbedingungen, Lizenzmodelle
  2. Verfügbarkeit in bestimmten Datenbanken
  3. unterschiedliche Archivierungspraktiken
  4. API-Beschränkungen und Plattformrichtlinien («post-API age», Freelon 2018)
  5. nationale Rechtslagen
- Storjohann/Schröter (2011): «Dieser Unterschied zwischen dem deutschen und englischen Korpus ist nicht unproblematisch und ist vor allem durch die Zugänglichkeit des Materials bedingt.» (S.186)

# Aufbereitung von Sprachdaten zu Korpora

# Datenaufbereitung

- Herausforderung: korpus- und computerlinguistische Aufbereitung schränkt Datenäquivalenz ein
- Die Aufbereitung von Daten implementiert neue Äquivalenzannahmen: Sobald Dokumente formatiert, bereinigt und annotiert werden, treffen wir Entscheidungen darüber, was als „gleich“ gilt.
- Zwei Beispiel
  1. **Annotation von Wortarten:** Zuordnung zu Kategorien setzt unterschiedliche Formen/Funktionen als gleichartig)
    - Universal Dependencies (de Marneffe et al., 2021): Framework für konsistente (sprachübergreifende) morphosyntaktische Annotation; ggw. 200 Baumbanken in über 150 Sprachen; 17 Wortarten, 37 syntaktische Relationen, morphosyntaktische Annotationen
  2. **Datenstruktur:** Übersetzung in ein gemeinsames Schema implementiert Äquivalenz

# Methoden der Korpusanalyse

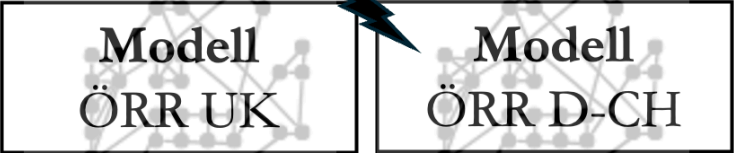
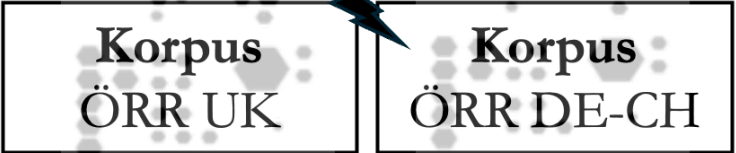
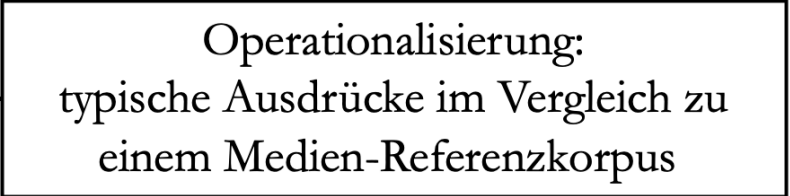
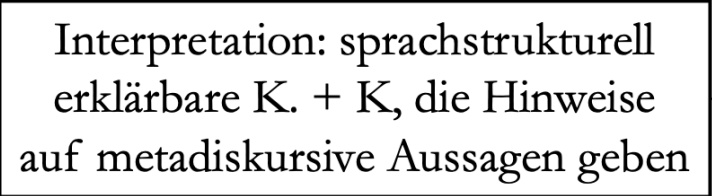
# Korpuslinguistische Methoden in der Vergleichenden Diskursanalyse

- Korpuslinguistische Berechnungen produzieren zunächst nichts weiter als Häufigkeiten, Verteilungen, Cluster, Vektoren oder Distanzen zwischen solchen Vektoren
  - Diese können Hinweise geben auf: 1) sprachstrukturelle Unterschiede und 2) diskursive Unterschiede
  - **Interpretationsschritt.** Unterschiede, die als diskursiv gedeutet werden, können ebenso gut auf Skalierungseffekte, Sampling-Entscheidungen, Referenzwahl, Aggregationsebene oder andere Modellierungsentscheidungen zurückgehen
- Die konkrete Umsetzung korpuslinguistischer Methoden implementiert spezifische Vergleichslogiken

# Korpuslinguistische Methoden in der Vergleichenden Diskursanalyse

- Keywords
  - Gemessen wird die Über- oder Unterrepräsentation eines Ausdrucks relativ zu einem Referenzkorpus; geben Hinweise auf 1) Thema, aber auch 2) Register, Genre, Stil (Cvrček & Berrocal, 2025)
  - Herausforderung in Bezug auf Vergleich:
    - Wahl / Äquivalenz der Referenzkorpora (Geluso & Roz, 2019)
    - Grösse des Referenzkorpus (Berber Sardinha 2000)
    - Wahl des statistischen Masses (Pojanapunya & Watson Todd 2018)
    - Interpretation der Keywords (Fokus Differenz? Fokus Gemeinsamkeit? Fokus Absenzen?)

# Zwischenfazit: Herausforderungen



- 1) Medientitel unter Berücksichtigung des Mediensystems UK und CH-D
- 2) Suchworte unter Einbezug von funktionalen Äquivalenten
- 3) Zeitraum unter Berücksichtigung länderspezifischer diskursiver Ereignisse (z.B. No-Billag in CH)

# Anforderungen an offene digitale Forschungsinfrastrukturen

Und welche Anforderungen ergeben sich für linguistische Infrastrukturen, wenn sie für vergleichende Diskursanalysen genutzt werden sollen?



Welche methodologischen Herausforderungen bestehen, wenn vergleichende Diskursanalysen korpusbasiert durchgeführt werden?

# Forschungsinfrastrukturen und Vergleichen

- Infrastrukturen können Kommensurabilität herstellen – aber wie?
- Herausforderungen:
  - Black-Box-Phänomene: Je stärker Infrastrukturen Vergleich automatisieren, desto größer ist das Risiko, dass Kommensurabilität *still* hergestellt wird (z.B. Setzungen bei Annotationen, Datenbereinigung, statistischen Massen)
  - Standardisierung als ungewollte Homogenisierung von Verfahren
  - Korpusanalysetools implementieren spezifische Vergleichslogiken

# Unter dem Aspekt von *findability* & *accessibility*

## 1. Dokumentation

- Auffindbarkeit und Zitierbarkeit von Modellparametern (z.B. in Repositorien)
- Semantische Präzisierung von Metadaten (Interpretationshilfen für Metadatenkategorien)
- Transparenz über Lücken und Unterrepräsentation (z.B. hinsichtlich Zeiträumen)
- End-to-end Pipeline-Transparenz

## 2. Zugang & Verfügbarkeit von Daten

- Infrastrukturanbieter als Ersteller von Korpora (vgl. Erfolg von Parlamint)
- Kooperationsfähigkeit mit Datenhaltenden (z.B. Mediendatenbanken, vgl. [Swissdox@LiRi](mailto:Swissdox@LiRi) in CH)

## 3. Rechtliche/wissenschaftspolitische Klärungen

- Infrastrukturen als Akteure in rechtlichen Debatten um Themen wie Urheberrecht, Datenschutz, Wissenschaftsschranken

# Unter dem Aspekt von *interoperability* und *reusability*

- 1. Subkorpora als Kernfunktion:** Tools zur Korpusanalyse müssen ermöglichen, Subkorpora nach nachvollziehbaren Kriterien zu bilden (Zeit, Medium, Texttyp, Akteursgruppe etc.) und diese Kriterien mitzuspeichern
- 2. Interoperabilität als toolübergreifende Vergleichbarkeit**
  - exportierbare Ergebnisprotokolle für Toolvergleich
  - Analyseparameter sind dokumentiert, nachvollziehbar und berichtbar (z.B. Frequenztypen, Statistiken, Schwellenwerte)
  - Unterstützung disziplinärer Workflows (Anschlussfähigkeit an andere Tools zur qualitativen und quantitativen Analyse von Sprachdaten)
- 3. Datenaufbereitung**
  - Evaluierbarkeit von Verarbeitungseffekten
  - Alternative/variierbare Verarbeitungspipeline

# Fazit

# Vergleichende Diskursanalyse: Herausforderungen und Chancen

- Politische Dimension
- Rechtliche Dimension
- Infrastrukturelle Dimension
- Gegenstandsbezogene Dimension
- ...

# Herzlichen Dank!

# Referenzen

- Borges Völker, E., Wendt, M., Hennig, F., & Köhn, A. (2019). HDT-UD: A very large Universal Dependencies Treebank for German. In A. Rademaker & F. Tyers (Hrsg.), *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)* (S. 46–57). Association for Computational Linguistics.
- Cvrček, V., & Berrocal, M. (2025). Sibling-texts keyword analysis: Exploring topic and register keywords. *Digital Scholarship in the Humanities*, 40(3), 762–778.
- Sardinha, T. B. (2000). Comparing corpora with WordSmith Tools: How large must the reference corpus be? *The Workshop on Comparing Corpora*, 7–13.
- Czachur, W. (2020). Kontrastive Diskurslinguistik. In T. Niehr, J. Kilian, & J. Schiewe (Hrsg.), *Handbuch Sprachkritik* (S. 204–210). J.B. Metzler.
- De Marneffe, M.-C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, 1–54.
- Dreesen, P., Suremann, I., Krasselt, J., & Stücheli-Herlach, P. (in Vorb.). It's a model, stupid!
- Epstein, J. (2008). Why Model? *Journal of Artificial Societies and Social Simulation* 11/4 12.
- Eysette, S. (2025). *Taboo Language and Incest in the French and British Press (2017-2022). A cross-linguistic corpus-assisted discourse analysis*. Dissertation.
- Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, 35(4), 665–668.
- Geluso, J., & Hirsch, R. (2019). The reference corpus matters: Comparing the effect of different reference corpora on keyword analysis. *Register Studies*, 1(2), 209–242.

- Müller-Eberstein, M., van der Goot, R., & Plank, B. (2021). How Universal is Genre in Universal Dependencies? In D. Dakota, K. Evang, & S. Kübler (Hrsg.), *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2021)* (S. 69–85). Association for Computational Linguistics.
- Pagano, A. S., Chiril, P., Chierchiello, E., & Bosco, C. (2025). TreEn: A Multilingual Treebank Project on Environmental Discourse. In G. Bouma & Ç. Çöltekin (Hrsg.), *Proceedings of the Eighth Workshop on Universal Dependencies (UDW, SyntaxFest 2025)* (S. 80–96). Association for Computational Linguistics.
- Pojanapunya, P., & Watson Todd, R. (2018). Log-likelihood and odds ratio: Keynes statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1), 133–167.
- Stachowiak, H. (1979). *Allgemeine Modelltheorie*. Springer.
- Storjohann, P. & Schröter, M. (2011). Die Ordnung des öffentlichen Diskurses der Wirtschaftskrise–und die (Un-) ordnung des Ausgeblendeten. *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 7/1. 32-53.
- Stücheli-Herlach, P., Dreesen, P., & Krasselt, J. (2022). Öffentliche Diskurse modellieren und simulieren. Wege der transdisziplinären Diskurslinguistik. *Zeitschrift Für Diskursforschung / Journal of Discourse Studies*, 2, 245–356.
- Taylor, C. (2014). Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International Journal of Corpus Linguistics* 19/3. 368-400.
- Truan, N. (2019). Möglichkeiten und Herausforderungen einer pragmatisch orientierten kontrastiven
- Diskursanalyse. Ein Vorschlag am Beispiel deutscher, französischer und britischer Parlamentsdebatten. *Diskurse – digital* 1 (3), 29-50.