





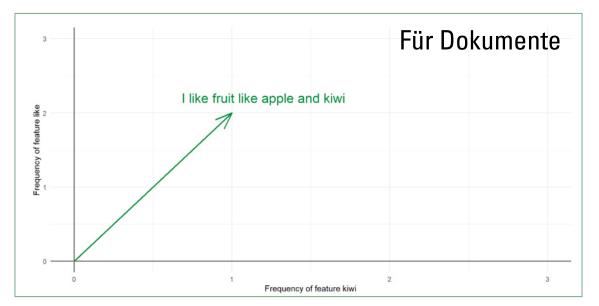
Agenda

- 1. Weitere Verfahren
- 2. Ausblick

1. Weitere Verfahren

Trend I: Mehr als Bag-of-Words

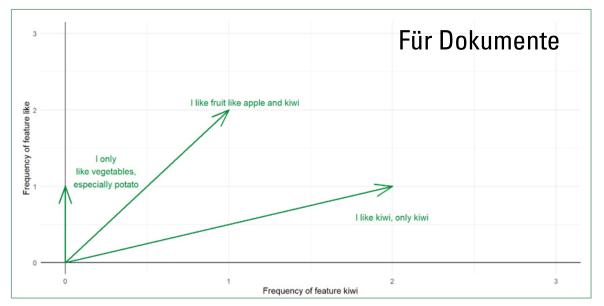
 Vector Space Modelle (VSM): Repräsentieren Wörter/Dokumente als Vektoren in einem Vektorraum. Vektoren, die einander nahe sind, sind sich semantisch nahe.



Weitere Verfahren | Evaluation

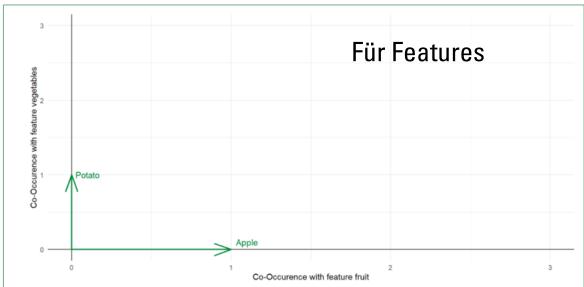
Trend I: Mehr als Bag-of-Words

 Vector Space Modelle (VSM): Repräsentieren Wörter/Dokumente als Vektoren in einem Vektorraum. Vektoren, die einander nahe sind, sind sich semantisch nahe.





 Word Embeddings: Repräsentieren z. B. Features als Vektoren in einem ndimensionalen Raum. Wir erklären die Bedeutung eines Features über die Features in dessen Kontext!





Trend I: Mehr als Bag-of-Words

- Word Embeddings...
 - "predict a focal word as a function of the other words that appear within a small window of that focal word" (<u>Rodriguez et al., 2023, p. 3</u>)
 - "represent a word as a dense vector in a low-dimensional space learned from unlabeled data" (<u>Grimmer et al., 2022, p. 79</u>)
- Wir erklären die Bedeutung eines Wortes durch Wörter um dieses herum!
- Beispiel <u>hier</u>



Trend II: Mehr als Text

- Andere Formen von Inhalt analysieren, z.B.
 - Bilder
 - Videos
 - Audio

Trend III: Mehr als Englisch

Zunehmend Verfahren für multilinguale Analysen (Licht & Lind, 2023)



COMPUTATIONAL COMMUNICATION RESEARCH 5.2 (2023) 1–31 HTTPS://DOLORG/10.5117/CCR2023.2.2.LICH

Going cross-lingual: A guide to multilingual text analysis

Hauke Licht

University of Cologne, Cologne Center for Comparative Politics

Fabienne Lind

University of Vienna, Department of Communication

Abstract

Text-as-data methods have revolutionized the study of political behavior and communication, and the increasing availability of multilingual text collections promises exciting new applications of these methods in comparative research. To encourage researchers to seize these opportunities, we provide a guide to multilingual quantitative text analysis. Responding to the unique challenges research faces in multilingual analysis, we provide a systematic overview of multilingual text analysis methods developed for political and communication science research. To structure this overview, we distinguish between separate analysis, input alignment, and anchoring approaches to cross-lingual text analysis. We then compare these approaches' resource intensiveness and discuss the strategies they offer for approaching measurement equivalence. We argue that to ensure valid measurement across languages and contexts, researchers should reflect on these aspects when choosing between approaches. We conclude with an outlook on future directions for method development and potential fields of applications. Overall, our contribution helps political and communication scientists to navigate the field of multilingual text analysis and gives impulses for their wider adoption and further development.

Trend IV: Mehr als Automatisierung

- Semi-automatisierte Verfahren,
 - z.B. Active Learning
 - z.B. Seeded Sentimentanalyse (Watanabe, <u>2021</u>)
- Automatisierte Verfahren mit geringen Mengen an Trainingsdaten, z.B. Few-Shot Learning (Rieger et al., 2024)
- Gesundes Misstrauen in menschliche Codierungen (Baden et al., 2024; Oschatz et al., 2024)

2. Evaluation



Evaluation: The good & the bad

- In 1 Satz, was hat euch am Workshop gefallen/geholfen/wovon hätte es mehr Inhalt sein dürfen
- In 1 Satz, was hat euch am Workshop nicht gefallen/nicht geholfen/wovon hätte es weniger Inhalt sein dürfen

Danke! Fragen?



Dr. Valerie Hase IfKW, LMU Munich

valeriehase

valerie-hase.com



Luisa Kutlar IfKW, LMU Munich



Iuisakutlar