



# The „Computational“ Turn in Journalism Studies

A Review of  
“Text as Data” Approaches

Valerie Hase (University of Zurich)

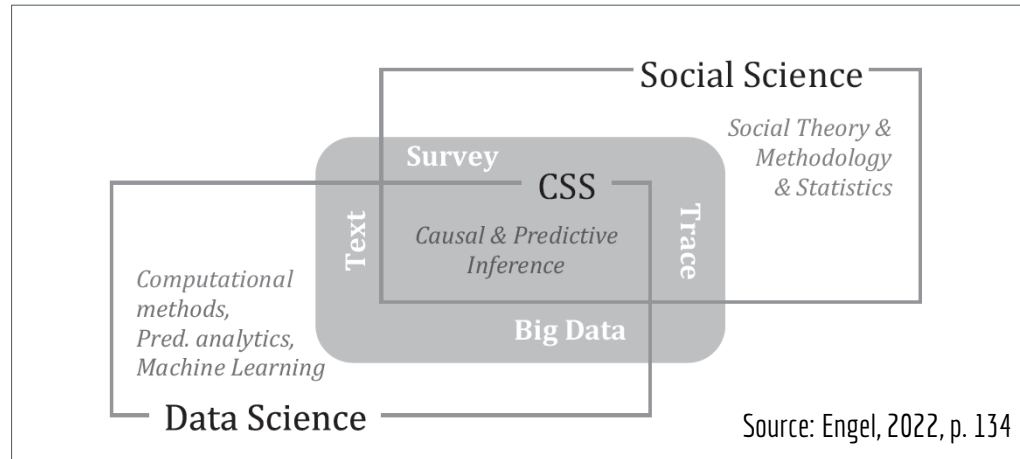


**Universität  
Zürich** UZH

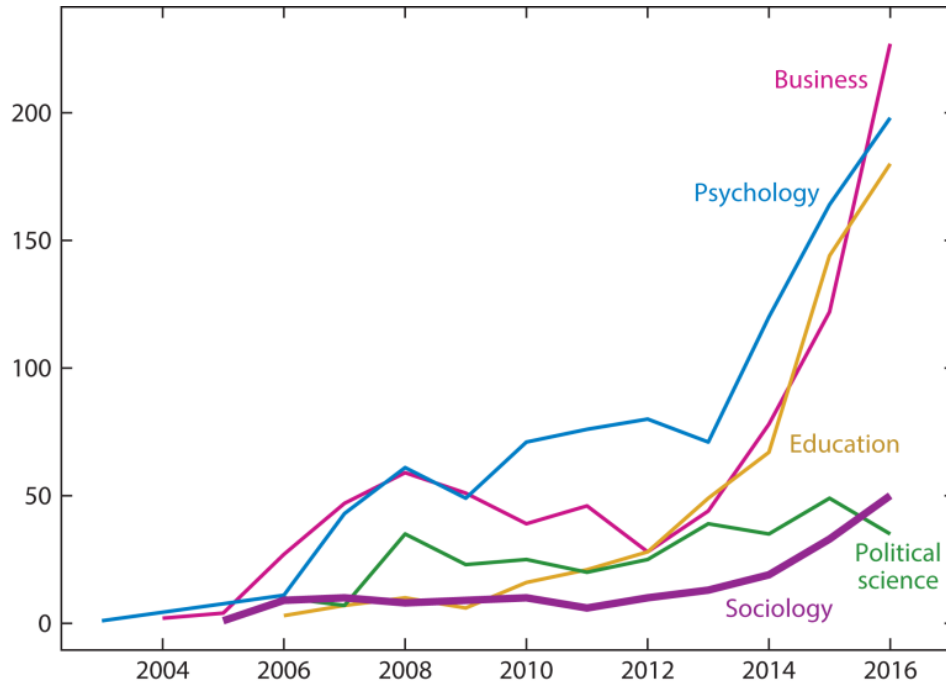
# Defining Computational Social Science (CSS)

(van Atteveldt & Peng, 2018; Hilbert et al., 2019; Lazer et al., 2009, 2020; Shah et al., 2015)

Applying Algorithmic Solutions to Analyze Unstructured, Complex, and Often Naturally Occurring Data Related to Questions & Theories from the Social Sciences



# Why Should We Care about CSS?



Number of studies mentioning CSS/big data,  
2000-2016

(Edelmann et al., 2020, p. 64)

On the one side...

"BIG DATA ANALYTICS [...] THREATENS TO COLONIZE THE SOCIAL SCIENCES AND HUMANITIES BY TURNING THESE FIELDS INTO COMPUTER SCIENCE."

(FUCHS & QIU, 2011, P. 222)

... and on the other side:

“COMPUTATIONAL SOCIAL SCIENCE INVOLVES  
INTERDISCIPLINARY FIELDS THAT LEVERAGE  
CAPABILITIES TO COLLECT AND ANALYZE DATA WITH AN  
UNPRECEDENTED BREADTH, DEPTH, AND SCALE.”

(CHANG ET AL., 2020, P. 68)

# How Is CSS Transforming the Social Sciences?

- Our data & methods
- Our theories & questions
- The way we work



# Transforming Our Data & Methods

(van Atteveldt et al., 2021; Baden et al., 2021; boyd & Crawford, 2011; Chan et al., 2021; Dobbrick et al., 2021; Engel, 2022; Hilbert et al., 2019)

---

Introducing (new) methods

vs.

Introducing assumptions we feel free to ignore & „researcher degrees of freedom“?

Introducing new, more accurate data sources

vs.

Analyzing more of the same data & introducing new biases



# Transforming Our Theories & Questions

(Anderson, 2008; Grimmer et al., 2021; Hofman et al., 2021; Kitchin, 2014; di Maggio, 2015; Nelson, 2020; Waldherr et al., 2021)

---

Introducing new theories &  
fostering theory-building

vs.

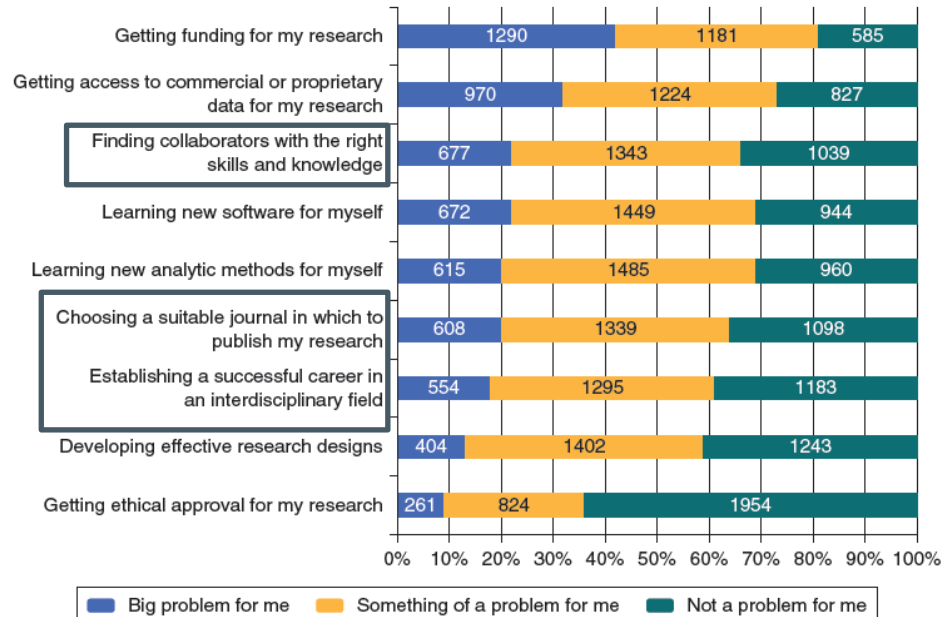
„The end of theory“  
(Anderson, 2008)

„Empiricism reborn“  
(Kitchin, 2014)



# Transforming the Way We Work

Figure 15 Challenges facing big data researchers (n = 2273)

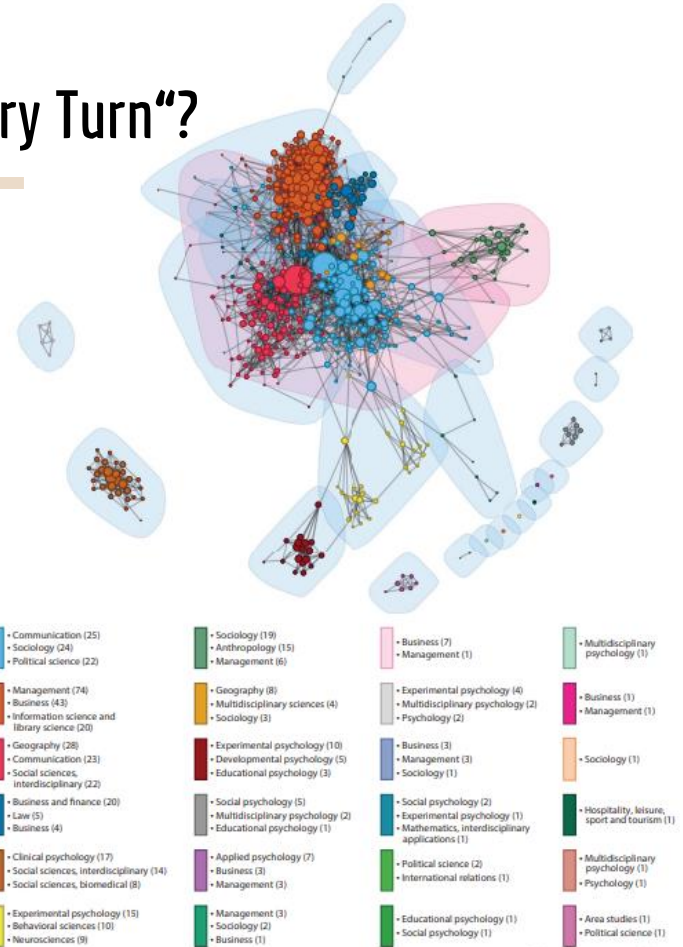


Source: Metzler et al., 2016 p. 12

# The Computational Turn: an „Interdisciplinary Turn“?

«tacking these kinds of questions requires an interdisciplinary approach»

(Wallach, 2018, p. 42)



Source: Edelmann et al., 2020, p. 65

# How Is CSS Transforming Interdisciplinarity within the Social Sciences?

- Our data & methods
- Our theories & questions
- The way we work



# On Interdisciplinarity

(Klein, 2017; von Nordheim et al., 2021; Wagner, 2011)

---

## Methodological interdisciplinarity

Use of methods from other disciplines (e.g., “text as data” approaches)

## Theoretical interdisciplinarity

Use of theories from other disciplines (e.g., complexity theory)

## Practical interdisciplinarity

Dissolution of disciplinary boundaries (e.g., cross-disciplinary teams, publishing outside of one’s field)



# CSS Does Not Guarantee Interdisciplinarity

(Chang et al., 2020; Lazer et al., 2020; National Science Foundation, 2020; Theocharis & Jungherr, 2021; Windsor, 2021)

---

«Interdisciplinary research faces a **daunting combination** of higher difficulty of attracting funding, lower likelihood of being published in top journals, and less recognition by tenure and promotions committees even once published»

(National Science Foundation, 2020, p. 13)

# The study: One method, one research field

---

## CONTENT ANALYSIS

Manual CA „genuine“ method of  
communication research (Loosen & Scholl, 2012)

Automated CA: Interdisciplinary use and  
development (Baden, et al., 2021; di Maggio, 2015)

## JOURNALISM STUDIES

Manual/automated CA  
frequently employed  
(Baden et al., 2021; Boczek & Hase, 2020; de Grove et al.,  
2020; Löffelholz & Rothenberger, 2011)

# Research Questions

---

**RQ1:** How are “text as data” approaches employed in the field of journalism studies?

**RQ2:** Are „text as data“ approaches related to (more) interdisciplinarity in communication science?

Search-term based retrieval of studies via Scopus (until 2020)

*(computer assisted OR automated OR automatic OR computational) AND (content analysis OR text analysis OR visual analysis) AND (journalis\* OR news\*)*

Out of all retrieved studies, manual inclusion ( $\alpha_{min} = .9$ ) to identify CSS sample

**CSS Sample (N = 262):**

*empirical + methodological studies using or developing automated CA to study journalistic communication*

**Benchmark Sample (N = 262):**

*empirical + methodological studies using mostly manual CA to study journalistic communication*

# Method: Systematic Literature Review



Manual coding of variables ( $\alpha_{min} = .75$ ), including:

### **Bibliographic-formal aspects**

e.g., disciplinary affiliation of each author; discipline of publication according to SSCI

### **Theories/concepts**

e.g., reliance on theories/concepts

### **Data/Methods**

e.g., focus on text vs. visual content, methods, variables, validation

# Method: Systematic Literature Review



## **RQ1:** How are “text as data” approaches employed in journalism studies?

---

### **Theories & concepts in empirical studies**

---

Majority of studies (56.9%) follows more explorative approach

Roughly a quarter of studies (23.8%) does not reference any theoretical concept

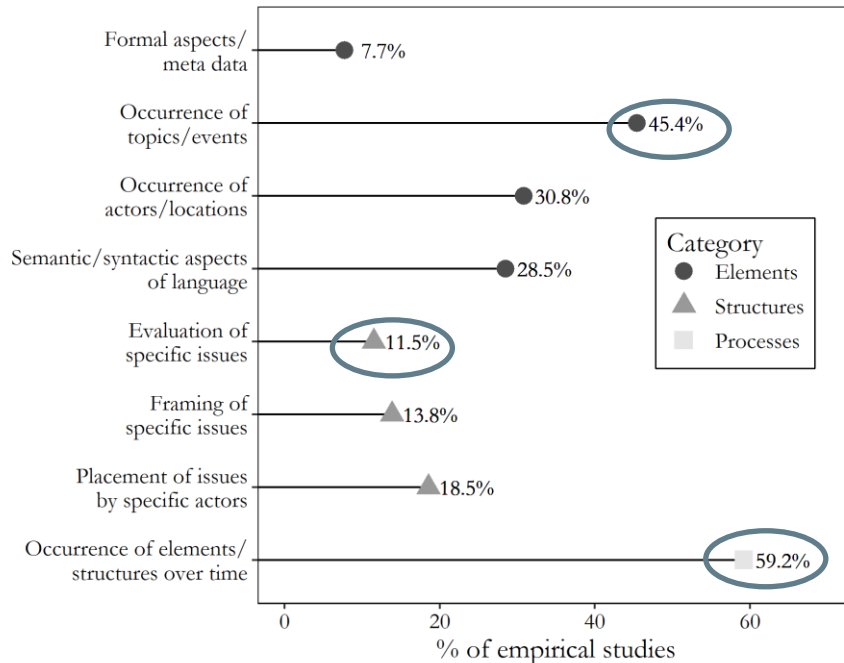
### **Data in empirical studies**

---

Overwhelming focus on written text (98.5%) over visuals/spoken language

# RQ1: How are “text as data” approaches employed in journalism studies?

## Variables analyzed in empirical studies



# RQ1: How are “text as data” approaches employed in journalism studies?

## Methods in empirical studies

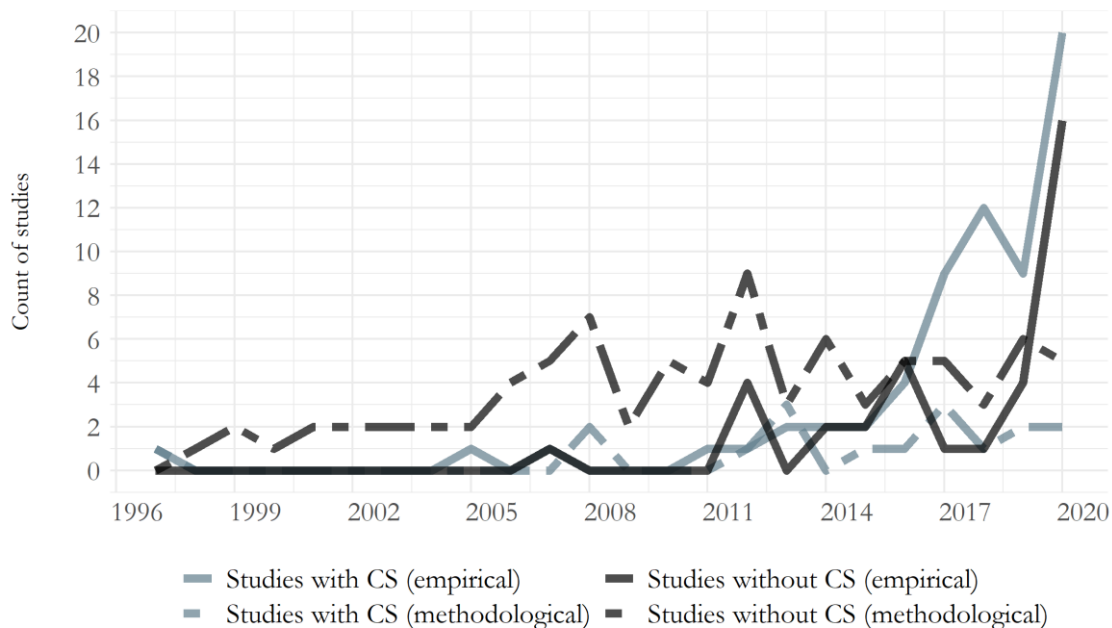
**46.9%** organic dict.  
**33.8%** rule-based  
**30.0%** unsupervised ML  
**20.0%** off-the-shelf dict.  
**18.5%** supervised ML  
(*N* = 130)

**40.8%**

of studies in CSS sample  
validate results against a  
manual gold standard

(*N* = 130)

## RQ2: Are text-as-data approaches related to more **methodological** interdisciplinarity?

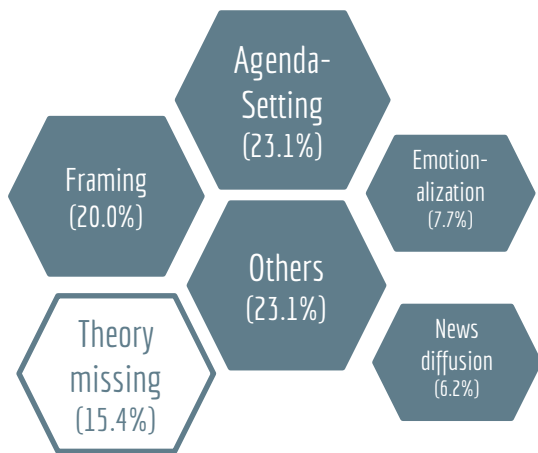


**Yes, but  
83.5%**

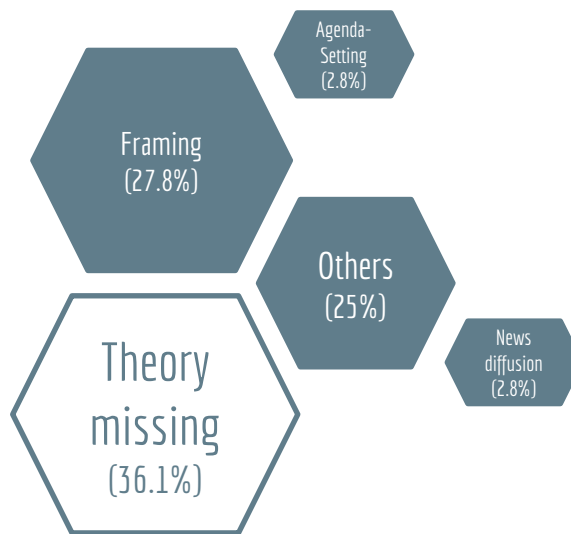
of methodological studies in  
CSS sample **without**  
involvement of communication  
science ( $N = 204$ )

## RQ2: Are text-as-data approaches related to more **theoretical** interdisciplinarity?

### Studies involving Communication Scientists



### Studies not involving Communication Scientists

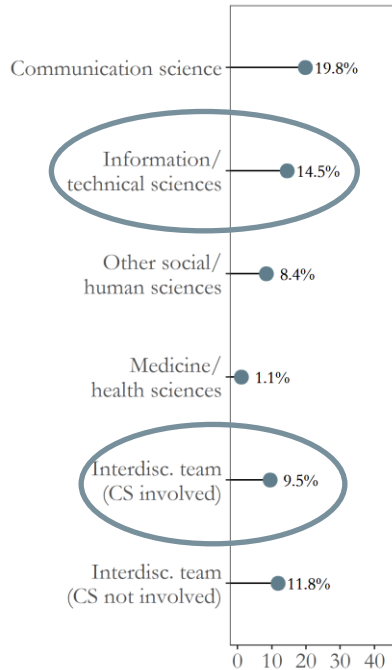


**No,**

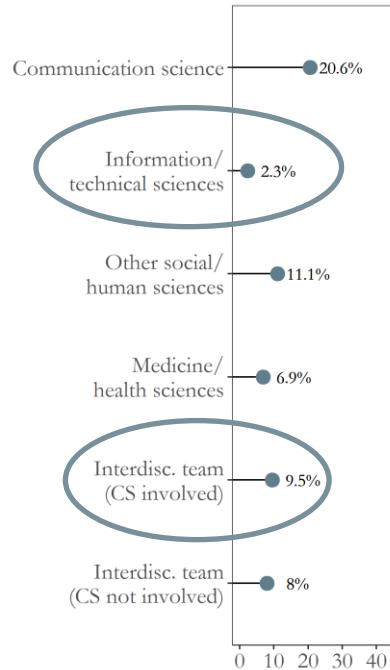
most empirical studies in CSS sample refer to „**classic**“ theories/concepts (if at all)  
( $N = 101$ )

## RQ2: Are text-as-data approaches related to more **practical** interdisciplinarity?

CSS SAMPLE



BENCHMARK SAMPLE



**No,**

similar level of of interdisciplinary cooperations & fewer publications «outside» our field

**but: shift**

towards the information/technical sciences (in cooperation & alone)

(N = 262 each)

# (How) is CSS transforming the social sciences?



## Transforming our data/methods

---

- New measurements & variables
- But: Lack of methodological standards, more of the same data

## Transforming our theories

---

- No indication of more interdisciplinarity in terms of theories: „same old, same old“
- Little theory-building; „banalization“ of existing theories

## Transforming how we work

---

- Not more, but other interdisciplinary cooperations
- But: stronger focus on „our“ publication outlets, little importance of „conference proceedings“



# To Get the Best of Both Worlds, We Need To...

- Make computational methods „our own“,  
„glass-box machine learning“ (Dobbrick et al., 2021)  
„hybrid/semi-supervised“ approaches (Baden et al., 2020; Watanabe & Zhou, 2021)
- Integrate social science in computing (Baden et al., 2021; Connolly, 2020)
- Reward interdisciplinary career paths (Lazer et al., 2020; Windsor, 2021)

Thanks for allowing me to steal your lunch break!  
**Any questions?**



Valerie Hase  
IKMZ, University of Zurich  
v.hase@ikmz.uzh.ch

