A Transnational Analysis of News and Tweets about *Nuclear Phase-Out* in the Aftermath of the Fukushima Incident

Philipp Heinrich\(^1\), Christoph Adrian\(^3\), Olena Kalashnikova\(^2\), Fabian Schäfer\(^2\), and Stefan Evert\(^1\)

\(^1\)Chair of CCL, \(^2\)Chair of Japanese Studies, \(^3\)Chair of Communication Science

Friedrich-Alexander University of Erlangen-Nuremberg

May 8, 2018
Introduction
Background

- Exploring the *Fukushima Effect*
  - identification and analysis of the tempo-spatial propagation of *discourses* in the *transnational algorithmic public sphere*
  - case study: Fukushima Effect
  - data: mass and social media (German, Japanese)

- **Emerging Fields Initiative** of FAU

- **Team:**
  - Chair of Corpus and Computational Linguistics
    Prof. Dr. Stefan Evert, Philipp Heinrich
  - Chair of Japanese Studies
    Prof. Dr. Fabian Schäfer, Olena Kalashnikova
  - Chair of Communication Science
    Prof. Dr. Christina Holtz-Bacha, Christoph Adrian
  - Chair of Computer Graphics
    Prof. Dr.-Ing. Marc Stamminger, Jonas Müller
Introduction

Methodology
Keywords, Collocations, and Discourse
Corpora

Case Study: Fukushima Effect
Mass Media
Social Media

Instead of a Conclusion: Work in Progress
Methodology
From Text to Discourse

- Foucault (1969): statements in conversation
- interpretation of text means categorizing
  - utterances
  - sentences
  - paragraphs
  - tweets
- categories
  - not known *a priori*
  - must be made up *on the fly* by the hermeneutic interpreter

Case Study “Exploring the Fukushima Effect”

- discourse(s): attitudes and opinions towards energy sources
- intra- and transmedial and -national
- “edited mass communication” vs. “mass self-communication”
From Text to Discourse

- Foucault (1969): statements in conversation
- interpretation of text means categorizing
  - utterances
  - sentences
  - paragraphs
  - tweets
- categories
  - not known \textit{a priori}
  - must be made up \textit{on the fly} by the hermeneutic interpreter

Case Study “Exploring the Fukushima Effect”

✓ discourse(s): attitudes and opinions towards energy sources
✓ intra- and transmedial and -national
✓ “edited mass communication” vs. “mass self-communication”
Corpus-Based Discourse Analysis (CDA)

- CDA means analyzing and deconstructing concordance lines (Baker, 2006)
  - concordances are the essence of discourses
- finding **discourses**: **nodes** + **attitudes**
  - (topic) nodes: defined by *keywords* or (more generally) *corpus queries*
  - attitudes: *collocates* that are retrieved by statistical methods
- examples
  - “refugees as victims” (Baker, 2006)
  - “Fukushima as worst case scenario”

**in practice:**

- look at (*n* best) collocates of topic node
- think of categories
- categorize manually
Corpus-Based Discourse Analysis (CDA)

- CDA means analyzing and deconstructing concordance lines (Baker, 2006)
  - concordances are the essence of discourses
- finding discourses: nodes + attitudes
  - (topic) nodes: defined by keywords or (more generally) corpus queries
  - attitudes: collocates that are retrieved by statistical methods
- examples
  - “refugees as victims” (Baker, 2006)
  - “Fukushima as worst case scenario”

in practice:

- look at \( n \) best collocates of topic node
- think of categories
- categorize manually
Collocates and Keywords

keywords

- given two frequency lists of lexical items
- perform statistical tests
  - likelihood
  - log ratio
  - frequency filter

collocates

- given a definition of a subcorpus
- rate lexical items according to association strength
  - window size
  - segment-based
  - association measure
Corpora – mass media

Frankfurter Allgemeine Zeitung (2011–2014)

- statistics:
  - 306,580 articles, 1,656,372 paragraphs
  - 145,055,523 tokens (1,981,726 types)
- linguistic annotation:
  - TreeTagger (tokenization, POS-tagging, lemmatization)

Yomiuri Shimbun (2011–2015)

- statistics:
  - 1,688,435 articles, 12,757,433 paragraphs
  - 580,518,367 tokens (392,971 types)
- linguistic annotation:
  - MeCab (SUWs)
Corpora – social media (Twitter)

German Twitter

- 10,266,835 original posts
- linguistic annotation:
  - tokenization: SoMaJo (Proisl and Uhrig, 2016)
  - POS-tagging: SoMeWeTa (Proisl, 2018)
  - lemmatization: work in progress

Japanese Twitter

- 275,689,087 original posts
- linguistic annotation:
  - special dictionary: ipadic-neologd (Sato et al., 2017)
  + identification and removal of noise (Schäfer et al., 2017)
Corpora – social media (Twitter)

German Twitter

- 10,266,835 original posts
- linguistic annotation:
  - tokenization: SoMaJo (Proisl and Uhrig, 2016)
  - POS-tagging: SoMeWeTa (Proisl, 2018)
  - lemmatization: work in progress

Japanese Twitter

- 275,689,087 original posts
- linguistic annotation:
  - special dictionary: ipadic-neologd (Sato et al., 2017)
  - identification and removal of noise (Schäfer et al., 2017)
Case Study: Fukushima Effect
Mass media in the aftermath of the Fukushima incident

German (FAZ)

- salience of energy transition discourse relatively stable (2011–2014)
- nuclear phase-out (Atomausstieg) as part of this discourse: sparked shortly after 3/11
  - political actors and issues (Ethikkommission, electricity supply)
  - economic actors (RWE)
  - technological issues (Stromnetz)

Japanese (Yomiuri)

- nuclear phase-out (脱原発) in 2011:
  - political actors (菅, 野田, 首相)
  - economic issues (発電, 稼働, 復興)
  - technological aspects (安全, 燃料)
- nuclear phase-out in 2014:
  - elections and politics (演説, as used in 街頭演説)
  - fewer words regarding economics (note アベノミクス)
German Twitter

Monthly signal strength of tweets

Monthly salience of topics
Japanese Twitter

monthly signal strength of tweets

Fukushima incident

monthly salience of topics

relative frequency [%]

topic
elections
#elections
#nuclear phase-out
nuclear phase-out
Instead of a Conclusion: Work in Progress
Second-Order Collocates

Figure: Paragraph-collocates of Germany in the FAZ corpus.
Second-Order Collocates

Figure: Collocates of *Germany* in energy-transition paragraphs.
Discourse Embeddings
Towards Mixed-Methods Discourse Analysis
Thanks for listening.

Questions?
References


