TexTiles: Exploring Patterns in Historical Discourse

Robert Roessler, Caiseen Kelly, Michael Behrisch, Johanna Beyer.
4th Workshop on Visualization for the Digital Humanities (at IEEE Vis 19), 2019.
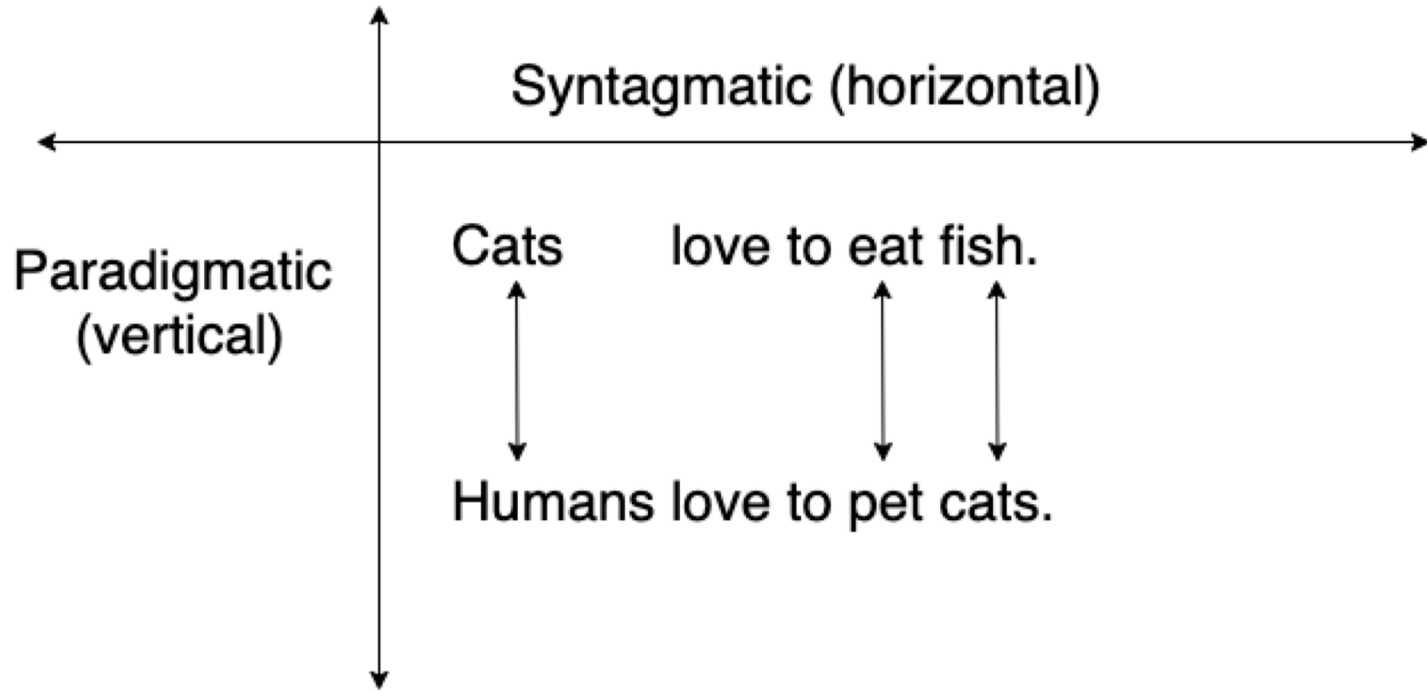
Background

- Discourse
    - A social boundary that defines what statements can be said about a topic
- Discourse analysis
    - A method in the humanities to analyze historical trends and knowledge figurations
- Corpus (pl. corpora)
    - a collection of written texts

User requirements

- Synchronic analysis
  - Explore a system of discourse at one point in time
- Diachronic analysis
  - Analyze discourse over time
- Syntagmatic analysis
  - Combination and connection of words surrounding a previously defined keyword
- Paradigmatic analysis
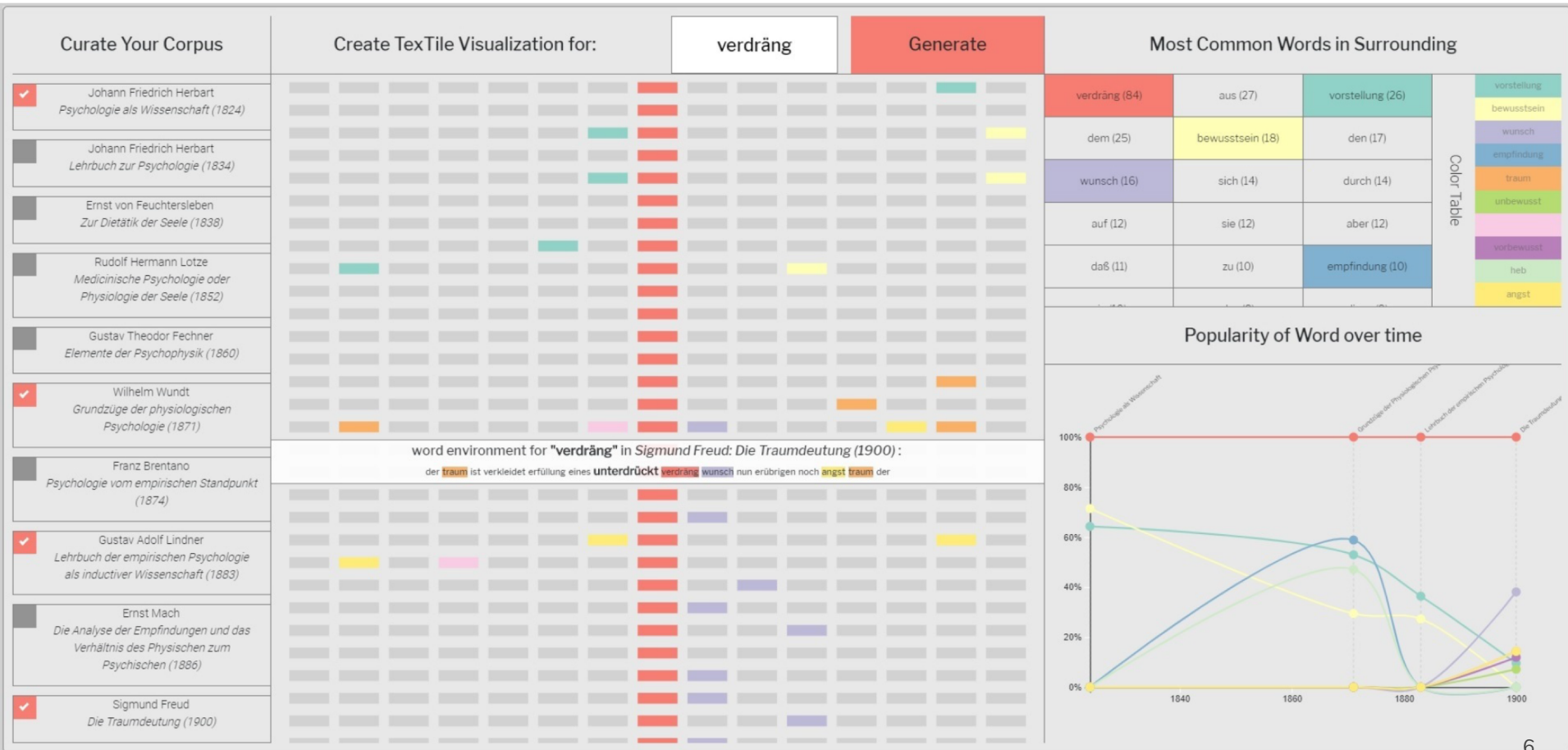 - Interchangeability in certain word slots
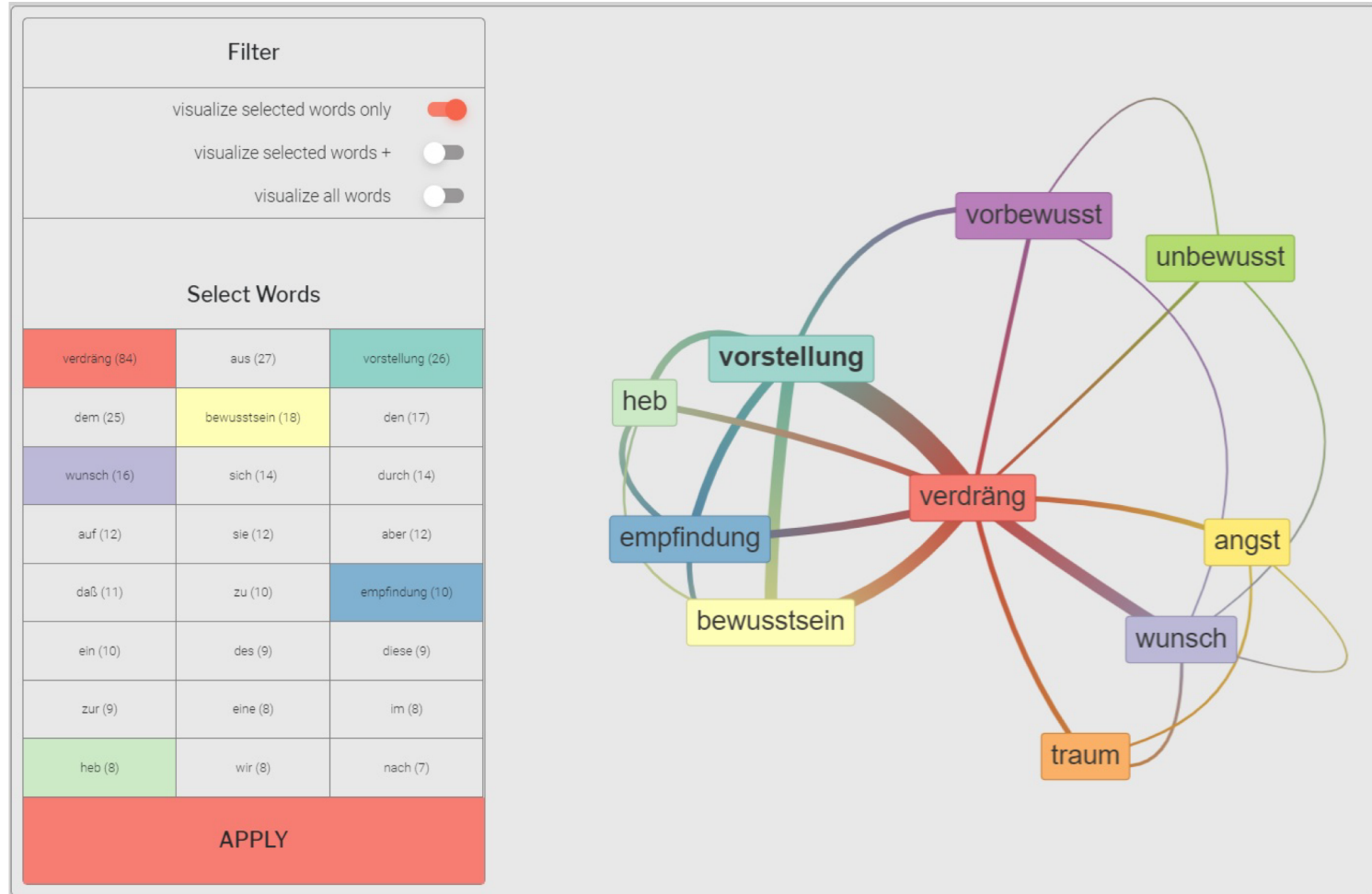
User requirements

Abstract tasks

- Synchronic analysis
  - Co-occurrence of words
- Diachronic analysis
  - Relative popularity of the surrounding words

Word-Net view

Word-Net view

Comparative Word-Matrix view

Analysis summary

- What: data
  - Text
- What: derived
  - Frequency of co-occurrence of words
- What: derived
  - Co-occurrence network

- How: encode
  - Time-series line charts
  - Force-directed placement
  - Matrix views of networks
- Scale
  - Thousands of pages of text, dozens of words

Critique

- Strengths
  - Two-dimensional layout augments keyword search
  - Matrix views of networks with color encoding
- Limitation
  - Paradigmatic analysis
  - Definition of co-occurrence
  - Line charts for too scarce and possibly categorical data

Thank you!