# Reconstructing musical experiences from historical texts

## (Brigitte Mathiak, Claes Neuefeind)

13th Symposium and Summer School On Service-Oriented Computing,
June 17 – June 23, 2019, Crete, Greece (DH Session)

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Context

- DFG research unit on „Historical Music Psychology"

- Establish a new approach in musicology:
  - How did people experience music in its historical context?
  - Which (compositional) patterns triggered these experiences?
  - What can we learn from it?


→ Basic idea: Access to these experiences gives access to the relation between compositional practice and expressive qualities

# Examples

- **Ludwig van Beethoven, Symphony No. 3, *Eroica* (1803)**
  - Anonymous (1811), quoted in Sipe, *Beethoven*, p. 78: "The composer has called it 'Sinfonia Eroica' and seems to have portrayed in it the picture of a battle. One hears the courageous assaults, the wild rage, the unremitting attacks and confused anger, the loud, groaning lamentation, and finally, exultation and **jubilation**."

- **Felix Mendelssohn, Symphony No. 5, Reformation (1830)**
  - Anonymous. "New Philharmonic concerts." *The Musical World*, vol. 46, no. 19 (May 9, 1868), p. 318: "The performance of this **noble** composition [...] was for the most part very fine."

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Examples

- **Richard Wagner, Overture to *Tannhäuser* (1845)**
  - Anonymous. "First Philharmonic concert." *Dwight's Journal of Music*, vol. 20, no. 16 (January 18, 1862), pp. 334–5, here p. 334: "It [sc. the overture] is imposing, startling in its effect, contrasting solemn religious tranquillity and **triumph** with delirious, despairing rapture of the senses [...]."

- **Johannes Brahms, Symphony No. 1 (1876)**
  - Anonymous. "The C-minor Symphony by Brahms." *Dwight's Journal of Music*, vol. 37, no. 19 (December 22, 1877), pp. 149–50: "But these defects are richly compensated by the bold, **majestic** way in which the final movement builds itself up."

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Methodology

SummerSoc 2019, Crete

19.06.2019

# Objectives

- Find historical textual evidence:
  - develop a procedure (a ‚tool') that allows to search for particular expressive qualities (e.g. ‚sad' ‚majestic' or ‚noble') in a corpus of 19th century music periodicals and journals to determine which works are mentioned most frequently in those contexts
- On this basis, analyze the characteristics of the works in question for similarities

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Hentschel: „Das Unheimliche" (the uncanny)

- Find historical textual evidence for the reception of music as being ‚uncanny' and analyze the works described in those texts

  1. Define a ‚semantic field'
  2. Search for all word forms of this field in the RIPM database
  3. Manual filtering based on direct contexts (skip obviously irrelevant ones)
  4. Select "suitable documents" (containing references to works/pieces)
  5. Musicological analysis of works mentioned in the documents

# Methodology: operationalization

- The „DH perspective": Which parts can be automatized or at least be supported by computational techniques?

  1. Define a ‚semantic field'
  2. Search for all word forms of this field in the RIPM database
  3. Manual filtering based on direct contexts (skip obviously irrelevant ones)
  4. Select "suitable documents" (containing references to works/pieces)
  5. Musicological analysis of works mentioned in the documents

# Methodology: operationalization

- The „DH perspective": Which parts can be automatized or at least be supported by computational techniques?

1. Definition of ‚semantic fields'
2. Search for all word forms of these fields
3. ~~Manual filtering based on direct contexts (skip obviously irrelevant ones)~~
4. Select "suitable documents" (containing references to works/pieces)
5. ~~Musicological analysis of works mentioned in the documents~~

SummerSoc 2019, Crete    19.06.2019

# Methodology (refined)

- The „DH perspective": Which parts can be automatized or at least be supported by computational techniques?

  - Generate exemplary **semantic fields** and refine them manually
  - Recognize/annotate **work mentions** (relevant documents)
  - **Search** for common contexts of semantic fields and work mentions

# Data

Towards a corpus of musical experiences

SummerSoc 2019, Crete

19.06.2019

# Data

- Approx. 300 music periodicals, 1.18 Mio pages

- Selection:
  - 1800-1914
  - German, Italian, French, English (among others)

# Data

- ***Allgemeine musikalische Zeitung*** (1798-1848, 1863-1882, Leipzig)
- ***Wiener Hof-Theater-Taschenbuch*** (1804-1816, Wien)
- ***Protokoll der Schweizerischen Musik-Gesellschaft*** (1808-1856, Luzern/Zürich/ Schaffhausen)
- ***The Quarterly Musical Magazine and Review*** (1818-1828, London)
- ***The Musical World*** (1836-1891, London)
- ***Revue et Gazette musicale de Paris*** (1835-1880, Paris)
- ***La France Musicale*** (1837-1870, Paris)
- ***Annales de la Société royale des Mélomanes*** (1838-1888, Gent)
- ***Gazzetta musicale di Milano*** (1842-1862, 1866-1902 [nach Fellinger bis 1912], Mailand)
- ***L'Italia musicale*** (1847-1859, Mailand)
- ...
- + many more!

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Data: Corpus Enrichment

- Gather more source files:
  - BSB (Bavarian State Library)
  - ZDB (Zeitschriftendatenbank, State Library Berlin)
  - ÖNB (Austrian National Library)
  - Wikisource
  - …

- OCR (tesseract) + Postcorrection

- Data model (Metadata, text structure)

# Step 1: Semantic Fields

19.06.2019

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Semantic Fields

- Semantic field != „Wortfeld" (Trier 1931)

- Hentschel (2016): sets of manually collected words, cf. the semantic field for ‚uncanny':

  - Erschauern, geisterhaft, gespenstig, gespenstisch, Grauen, grauenerregend, grauenhaft, grauenvoll, Graus, Grausen, grausig, gruselig, horribel, mysteriös, schauderhaft, Schauer, schauerlich, schauervoll, schaurig, unheimlich; abominable, demoniacal, demonic, dismay, dread, dreadful, eerie, eery, ghastly, ghostly, gruesome, horrible, horrid, horrific, mysterious, scary, shiver, shudder, spookish, spooky, uneasy, uncanny; angoissant, épouvante, frémir, frisson, horreur, horrible, inquiétant, lugubre, sinistre, téné- breux; misterioso, orrendo, orribile, orridezza, orrore, sinistro, spaventoso, mistico, inquietante.

# Semantic Fields

Options for compiling such fields:

- Samples based on corpus (cf. ‚the uncanny')
- Contemporary dictionaries, encyclopedias, etc. (e.g. Dornseiff)
- Dedicated resources like NRC Emolex, WordNet
- Use word embeddings to find sets of similar words

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Embeddings: Linguistic background

**The distributional hypothesis**

*Words with similar distributional properties have similar meanings*

(cf. Harris 1954, Rubenstein/Goodenough 1965, Miller/Charles 1991)

# Distributional Methodology

The lorry *driver* swerved on the road. As well as causing *pollution*, a lorry also has large *wheels*. A lorry requires *diesel* to work. A lorry might carry *sweet* apples and bananas. Bananas are easier to *peel* than apples but apples have nicer *trees*. Bananas are cheaper than apples in a *shop*.

|  | lorry | apples | bananas |
|---|---|---|---|
| *sweet* | 1 | 1 | 2 |
| *trees* | 0 | 2 | 2 |
| *shop* | 0 | 0 | 1 |
| *eat* | 0 | 0 | 0 |
| *peel* | 0 | 2 | 2 |
| *driver* | 1 | 0 | 0 |
| *road* | 1 | 0 | 0 |
| *diesel* | 2 | 0 | 0 |
| *pollution* | 1 | 0 | 0 |
| *wheels* | 2 | 0 | 0 |

Universität zu Köln

DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften
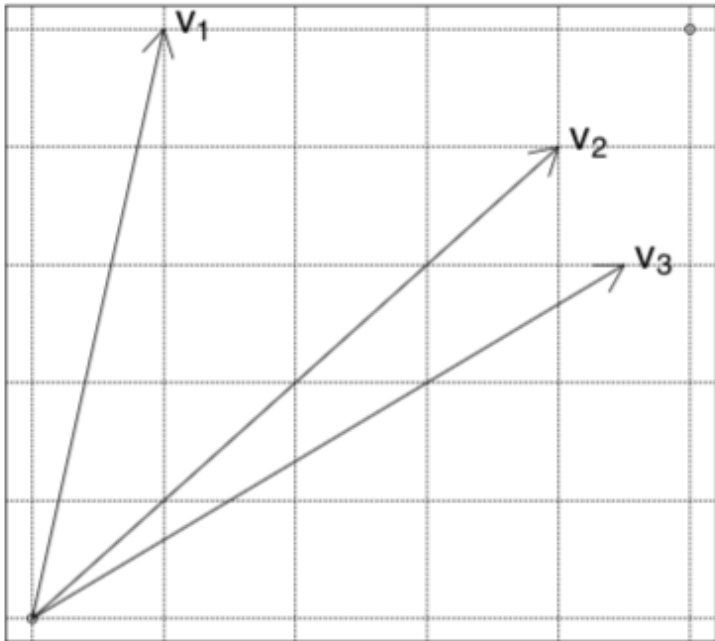
# Embeddings: Linguistic background

**The geometric metaphor of meaning**

*Meanings are locations in a semantic space, and semantic similarity is proximity between the locations*

(cf. Lakoff/Johnson 1980, 1999)

# Vector Space Models



- **Model of representation**: meaning as vectors in a high-dimensional space

- **Model of acquisition**: methodology for building the representation

SummerSoc 2019, Crete

19.06.2019

# Distributional Semantic Models (DSM)
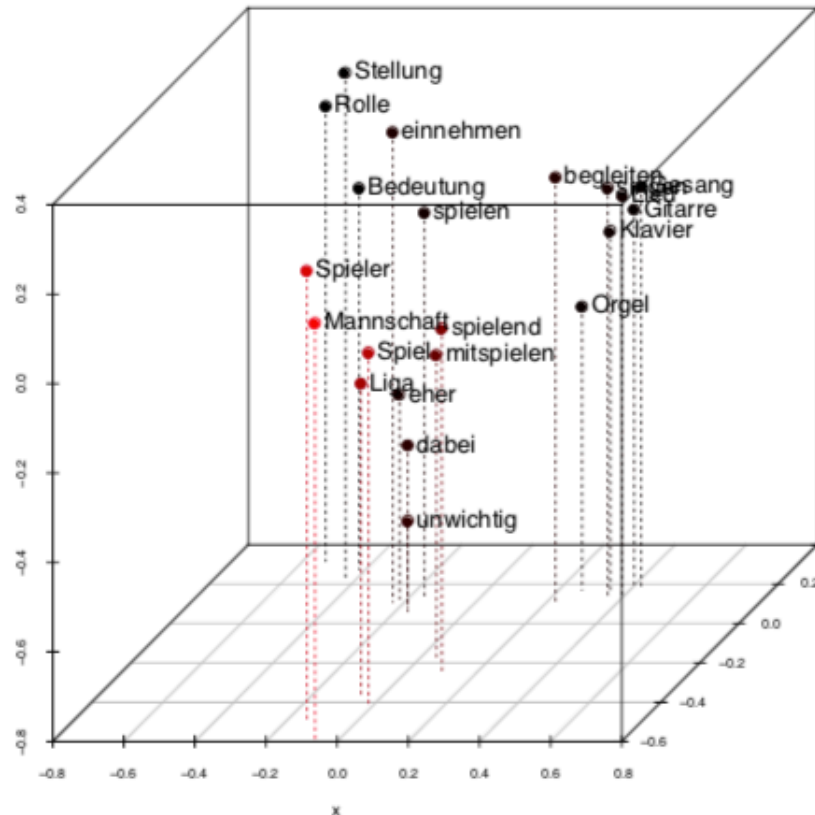


- **Benefits:**
  - Numerical representation
  - Straightforward: no additional (meta-)vocabulary

- **Shortcomings:**
  - Bag of words -> word order?
  - Sparse data -> dimensionality reduction

# Feature learning

If dimensionality is a problem, why not simply learn the features?

→ Predictive method (instead of count-based)

- Popular approach:
  - word2vec (Mikolov 2014)
- Basic idea: predict words from its neighbors, and then Deep Learning

SummerSoc 2019, Crete      19.06.2019

# word2vec – Input/Output



INPUT    PROJECTION    OUTPUT

w(t-2)

w(t-1)

SUM

w(t)

w(t+1)

w(t+2)

**CBOW**

INPUT    PROJECTION    OUTPUT

w(t)

w(t-2)

w(t-1)

w(t+1)

w(t+2)

**Skip-gram**

SummerSoc 2019, Crete                    19.06.2019

# word2vec – Deep Learning

SummerSoc 2019, Crete                    19.06.2019
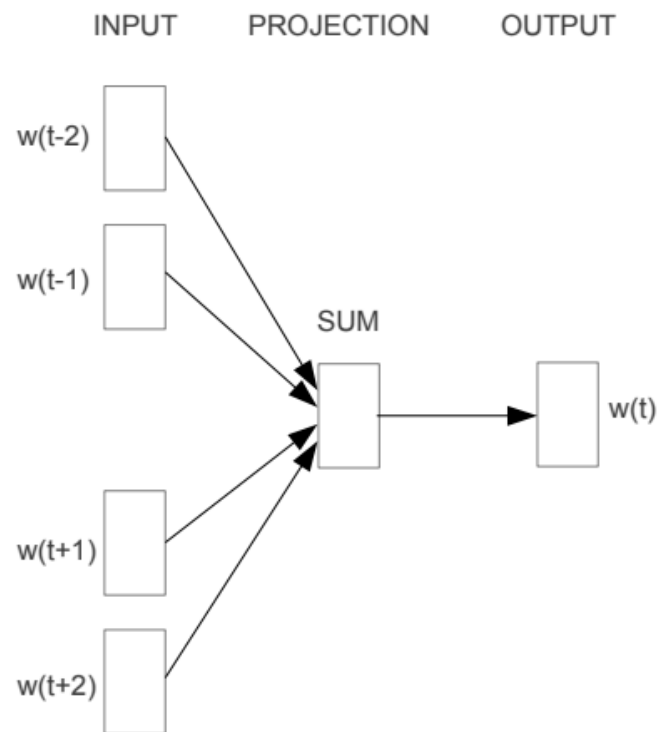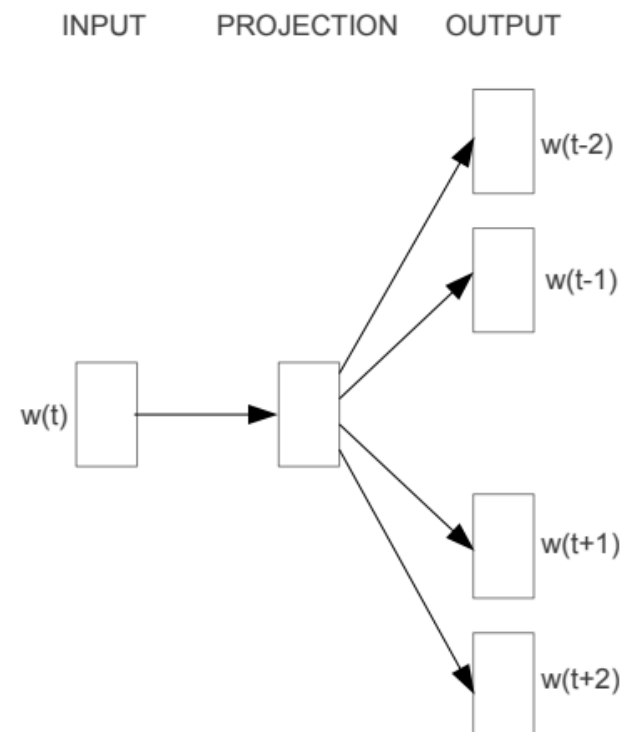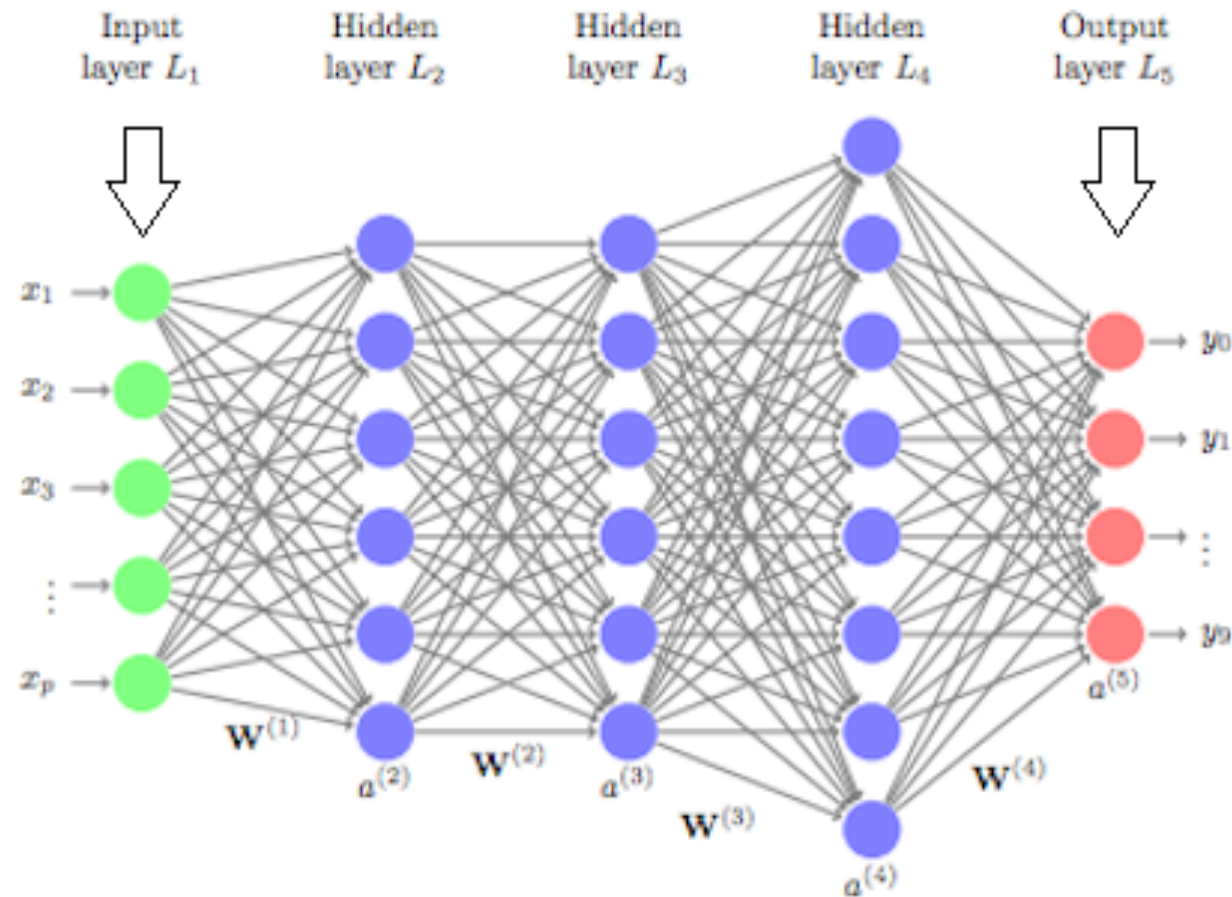
# Word2vec: features
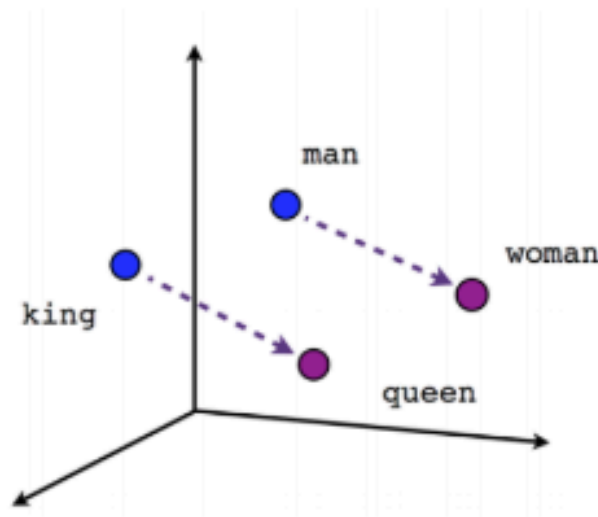


Male-Female
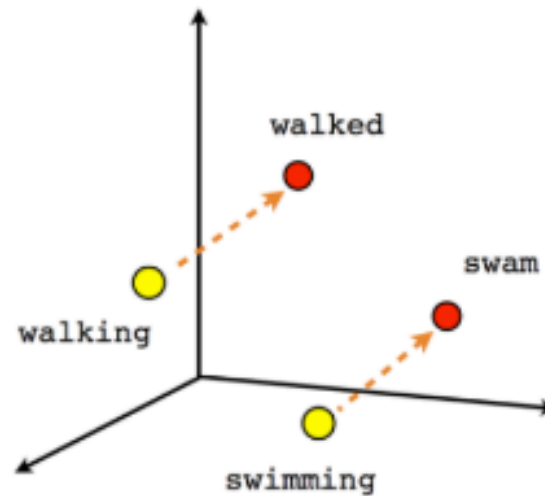
Verb tense

Country-Capital

SummerSoc 2019, Crete

19.06.2019
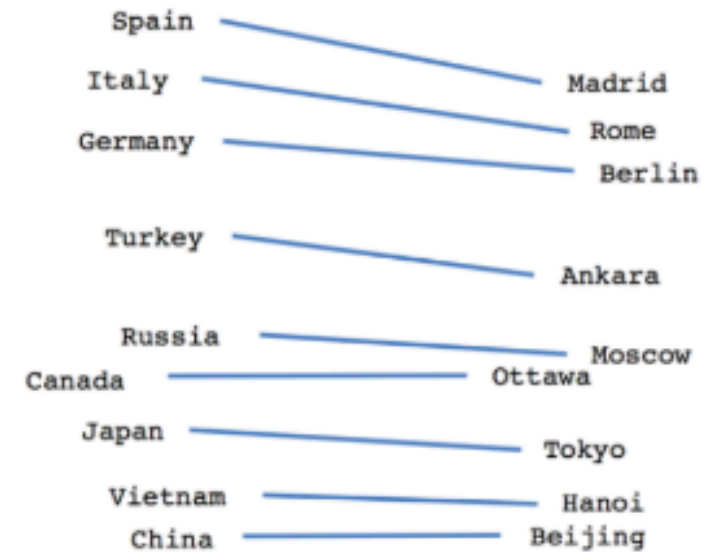
# Semantic Fields: Operationalization

- Build word2vec model based on raw text corpus of as much 19th century language we can find

- Query for similar words (here: word forms)


- However, „similar" means, may replace each other, so „good" is close to „bad"

- → …we need additional manual input

SummerSoc 2019, Crete    19.06.2019

# Refinement

- To actually work with semantic fields in the context of the project, we need to focus on specific examples
  - → Need for selection and refinement!

- Manual procedure: select, test, refine
  - Contemporary dictionaries, encyclopedias, etc. (e.g. Dornseiff)
  - Corpus-based samples

# Step 2: Detecting work mentions

# Work mentions

- **Ludwig van Beeth[...]** [...] *[...]roica* **(1803)**
  - Anonymous (1811[...] [...] p. 78: "The composer has called it '**Sinfonia Eroica**' a[...] [...] it the picture of a battle. One hears the courageous assaults, the wild rage, the unremitting [...] [...] the loud, groaning lamentation, and finally, exultation a[...]

- **Felix Mendelssohn, Symphony No. 5, Reformation** [...]
  - Anonymous. "New Philharmonic concerts." *The Musical World*, vol. 46, no. 19 (May 9, 1868), p. 318: "The performance of this noble **composition** [...] was for the most part very fine."

Simple, can be found with a list of relevant work titles

Title is ambiguous and can only be attributed through co-reverence resolution

Universität zu Köln

DCH Data Center for the Humanities Kölner Datenzentrum für die Geisteswissenschaften

# Work mentions

- Task: Entity recognition
- Options:
  - Knowledge-driven (lists of names and pieces)
  - Pattern induction (based on linguistic patterns)
  - ML-based (CRFs, Embeddings)
- Stand-off annotation

SummerSoc 2019, Crete                                      19.06.2019

# Step 3: Search

# Semantic Search

- Query expansion



```
Original terms:      Replaced by:
_____

jump              → (jump,leap,hop)
leap              → (jump,leap,hop)
hop               → (jump,leap,hop)
```

"jump,leap,hop" →

SummerSoc 2019, Crete                19.06.2019

# Synonym Token Filter

```
{
    "settings": {
        "index" : {
            "analysis" : {
                "analyzer" : {
                    "synonym" : {
                        "tokenizer" : "whitespace",
                        "filter" : ["synonym"]
                    }
                },
                "filter" : {
                    "synonym" : {
                        "type" : "synonym",
                        "synonyms_path" : "analysis/synonym.txt"
                    }
                }
            }
        }
    }
}
```

```
"u s a,united states,united states of america => usa"
"g b,gb,great britain => britain,england,scotland,wales"
```

Original terms:     Replaced by:
_____

u s a              → (usa)
united states      → (usa)
great britain      → (britain,england,scotland,wales)

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Scoring

- Use annotations to filter/sort search results:
    - weights determined by occurrence and position of work mentions
    - prioritize results in which such annotations occur close(r) to search terms
    - But: keep the rest (in case work mentions weren't properly annotated)

- Still: manual inspection by musicologist needed

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Step 4: Musicological analysis

# Musicological Analysis

- In-depth analysis of search results (type of expressive quality etc.)
- Analysis of compositions referred to in the texts

19.06.2019

# Summary and Outlook

SummerSoc 2019, Crete

19.06.2019

# Summary

- Workflow
    - Generate (and refine) semantic fields
    - Identify and annotate work mentions
    - Semantic search
    - Musicological analysis

→Gives access to the relation between compositional patterns and specific expressive qualities

- Benefits of „digitally enhanced" methodology: access, quantification

Universität zu Köln

DCH
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

# Thank you

**Work Group**

Frank Hentschel (Historical Musicology, Cologne)
Arnold Jacobshagen (Historical Musicology, Cologne
Gunter Kreutz (Systematic Musicology, Oldenburg)
Michael Oehler (Systematic Musicology, Osnabrück)
Frank Leymann (IAAS, U Stuttgart)
Brigitte Mathiak/Claes Neuefeind (IDH, U Cologne)

Universität zu Köln

**DCH**
Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften