



# The Course of Emotion in Three Centuries of German Text

## A Methodological Framework



Sven Buechel<sup>1</sup>, Johannes Hellrich<sup>1,2</sup> and Udo Hahn<sup>1</sup>

<sup>1</sup> Jena University Language & Information Engineering  
(JULIE) Lab

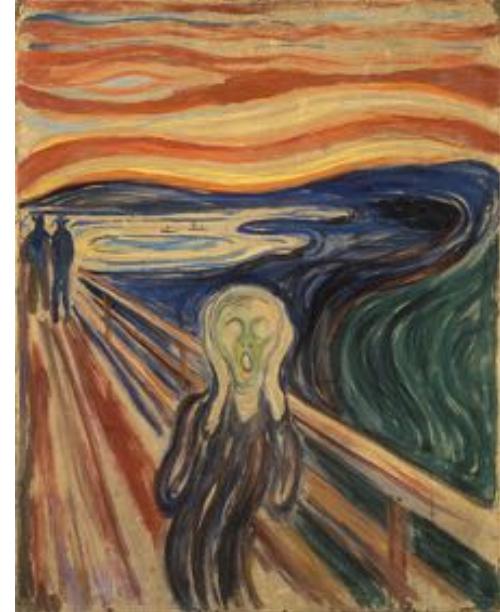
<http://www.julielab.de>

<sup>2</sup> Graduate School  
'The Romantic Model'

<http://www.modellromantik.uni-jena.de>

Friedrich Schiller University Jena, Jena, Germany

# Motivation: Emotion in the Humanities



<https://de.wikipedia.org/wiki/Laokoon-Gruppe>

[https://de.wikipedia.org/wiki/Der\\_Schrei](https://de.wikipedia.org/wiki/Der_Schrei)

<http://www.br.de/telekolleg/faecher/deutsch/literatur/goethe-weimarer-klassik-100.html>

# Resource Problem in Sentiment Analysis

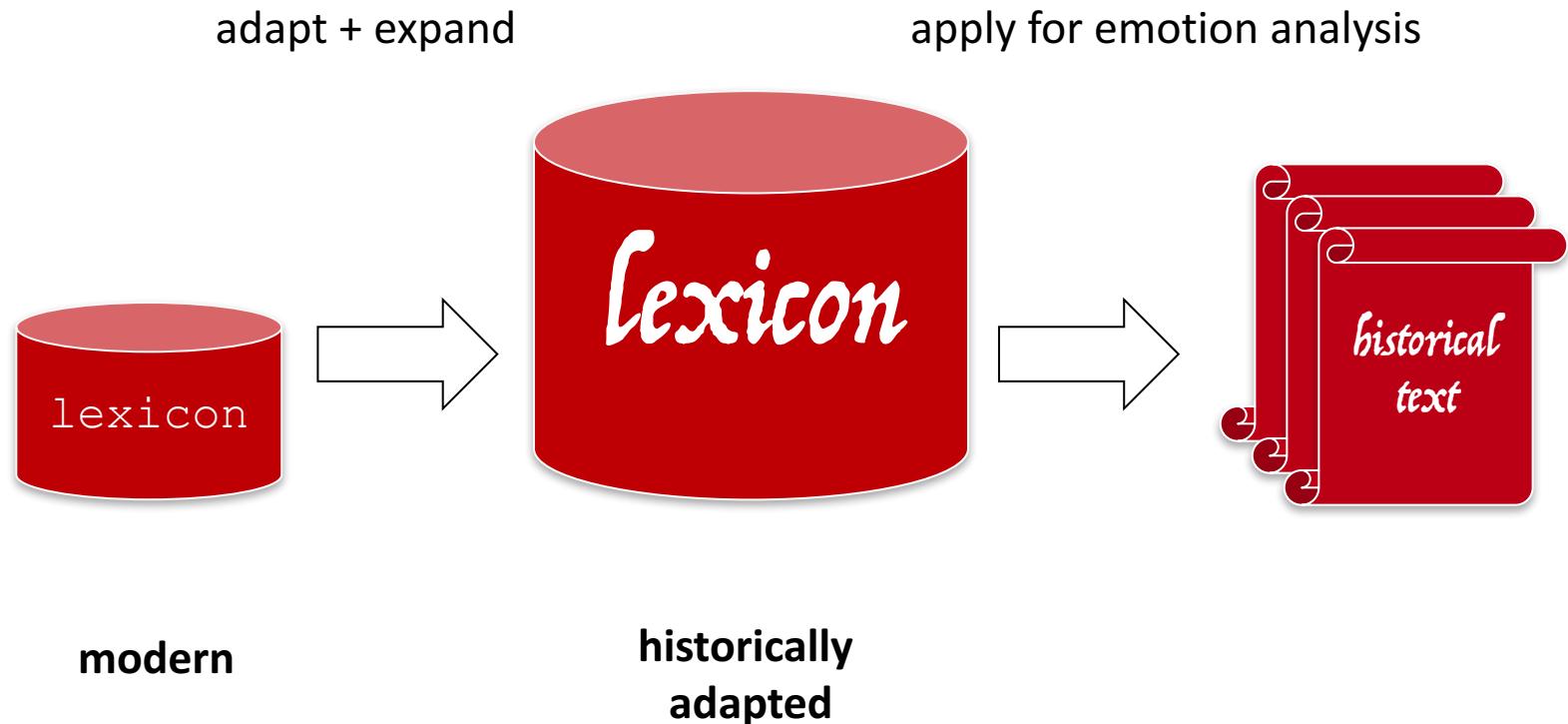
## Contemporary Resources



## Historical Resources

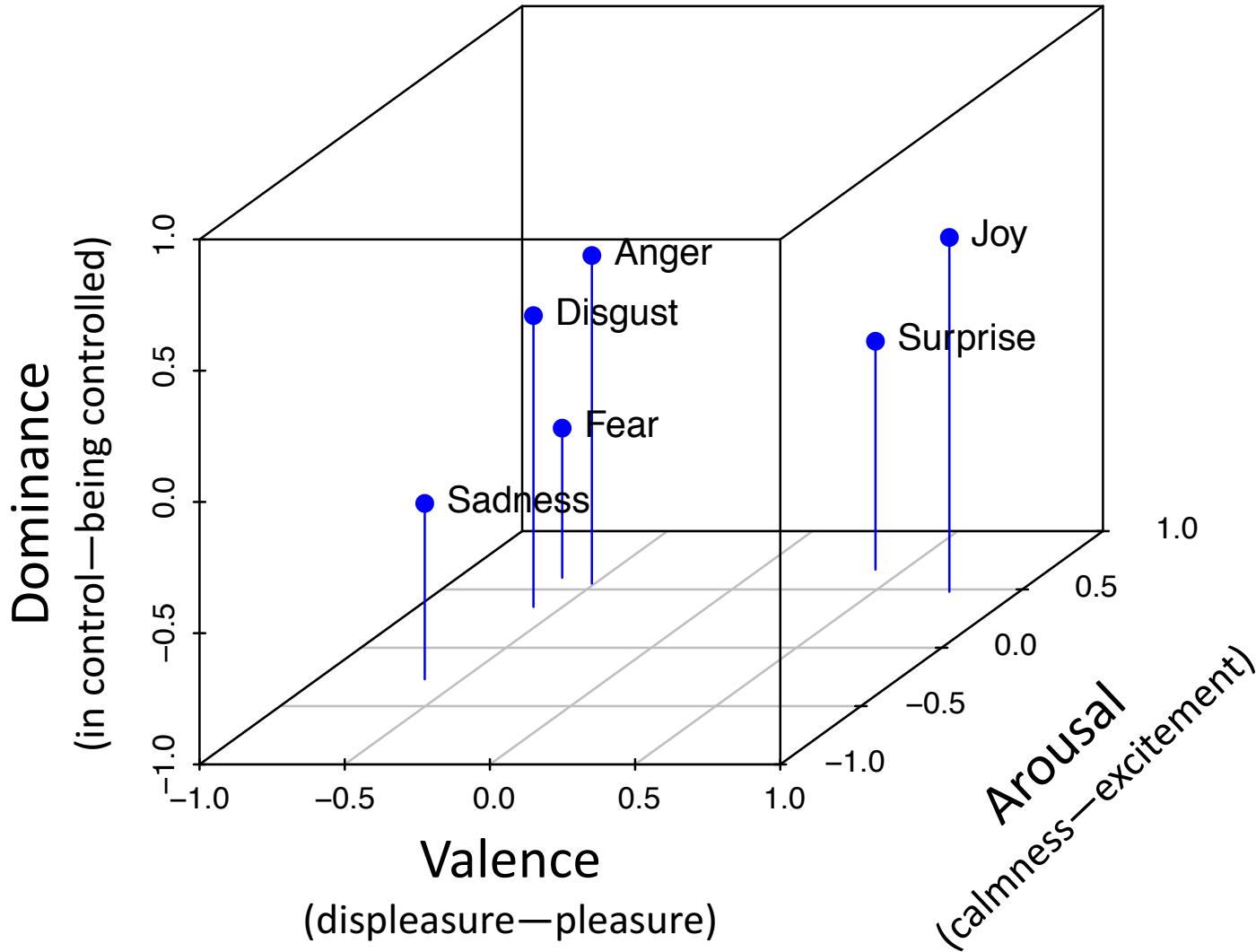


# Methodological Framework



# Methods

# Representing Emotion: VAD



# Emotion Lexicons

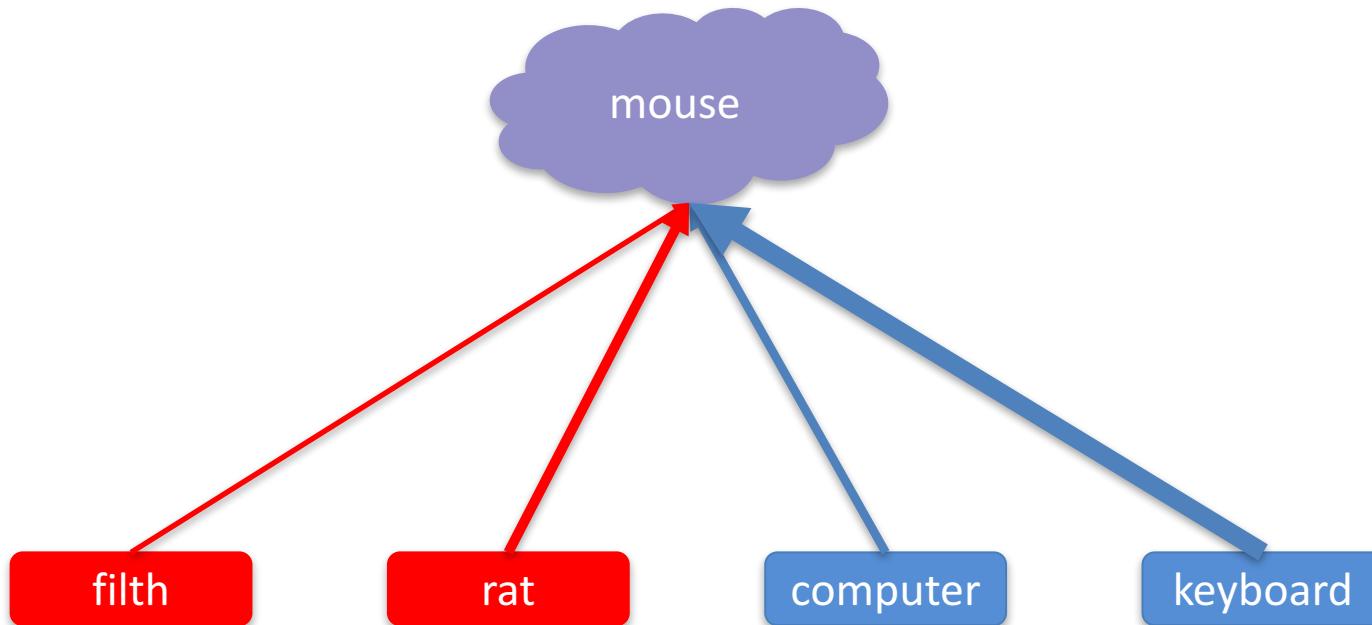
- Lexical resources
- Store emotion values of individual words
- Many lexicons available from psychology

Lemma	Valence	Arousal	Dominance
sunshine	8.1	5.3	5.4
terrorism	1.6	7.4	2.7
calm	6.9	1.7	7.4

(Warriner et al., 2013)  
(1-to-9 scales)

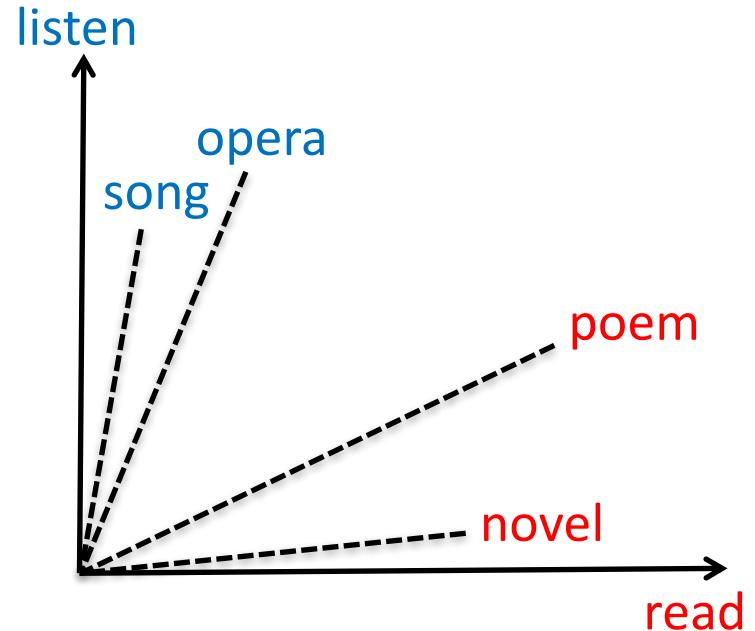
# Lexicon Expansion

- Automatic expansion based on **Seed**
- Turney-Littman-Algorithm:

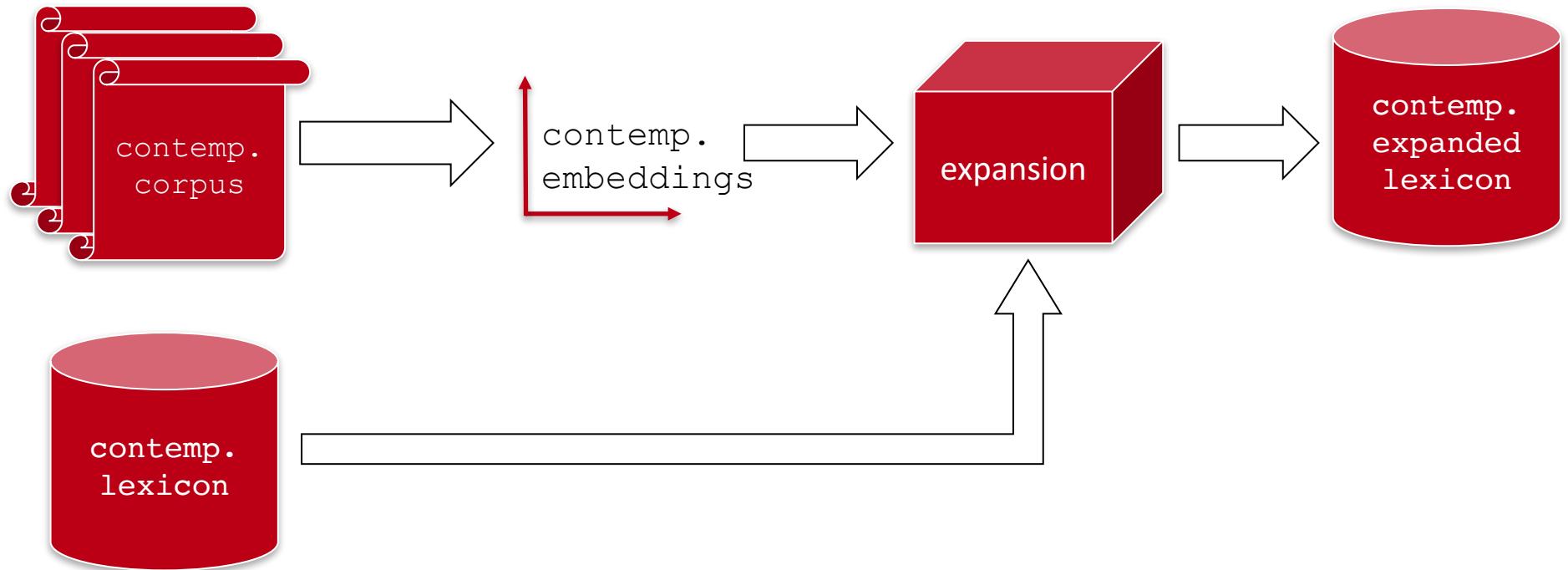


# Measuring Similarity: Word Embeddings

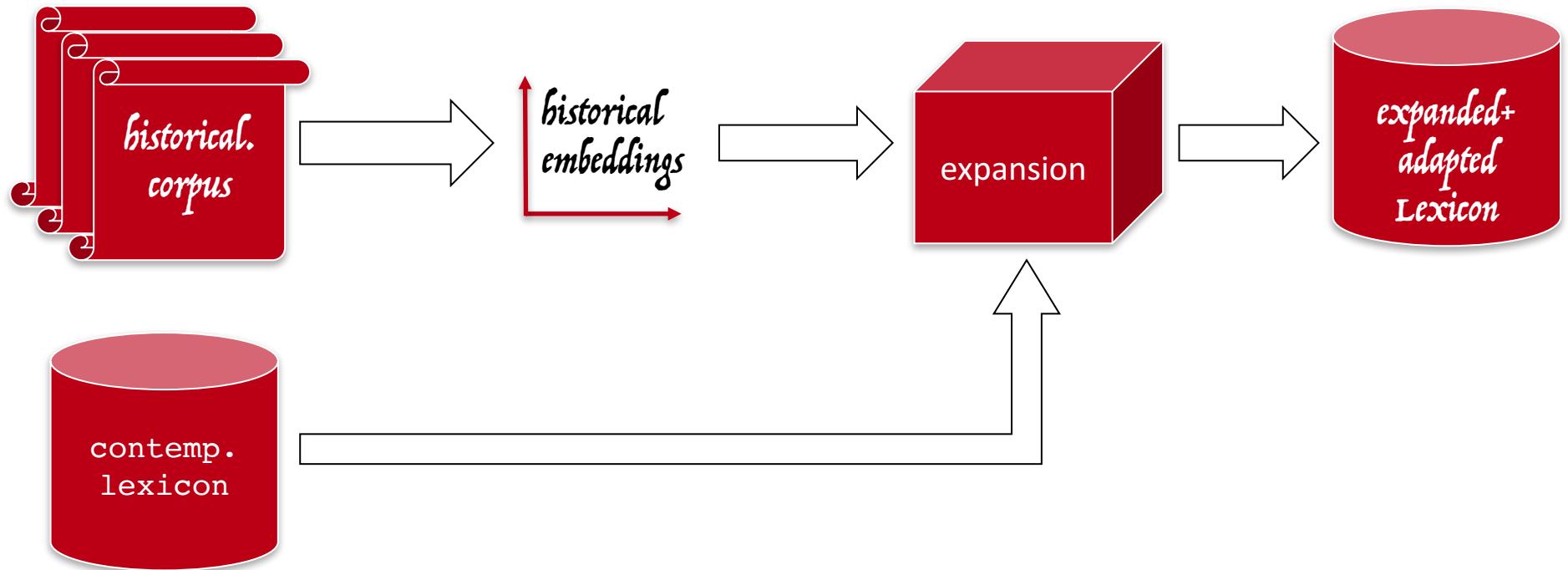
- Dominant approach to similarity
  - Dense, low-dimensional vectors as word representation
  - Most common: word2vec (SGNS) (Mikolov et al., 2013)
  - SVD<sub>PPMI</sub> supposedly better suited for humanities (Levy et al., 2015, Hellrich & Hahn, 2016)
- Johannes' talk for details!  
(Friday, 9am, LP-26)



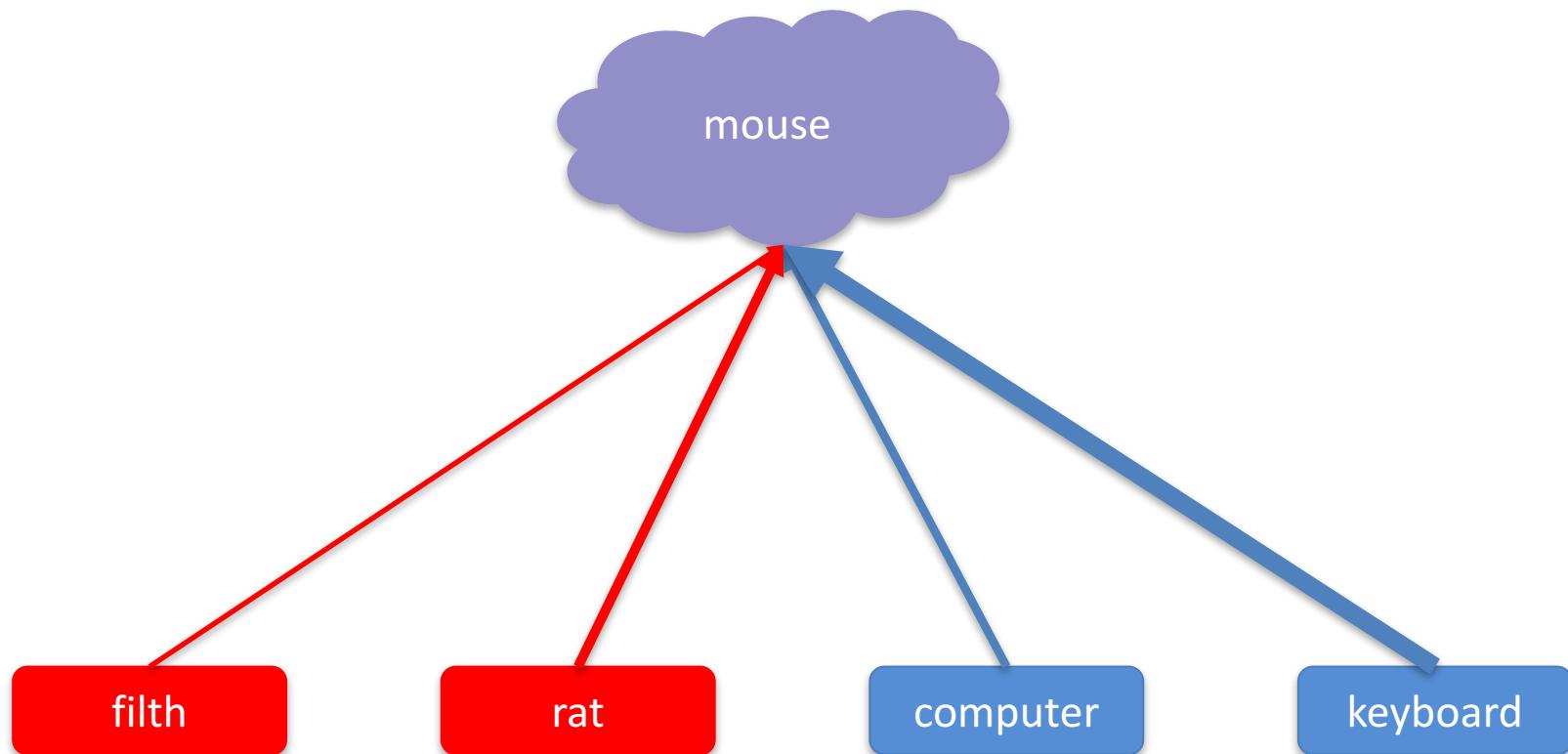
# Classical Expansion Methodology



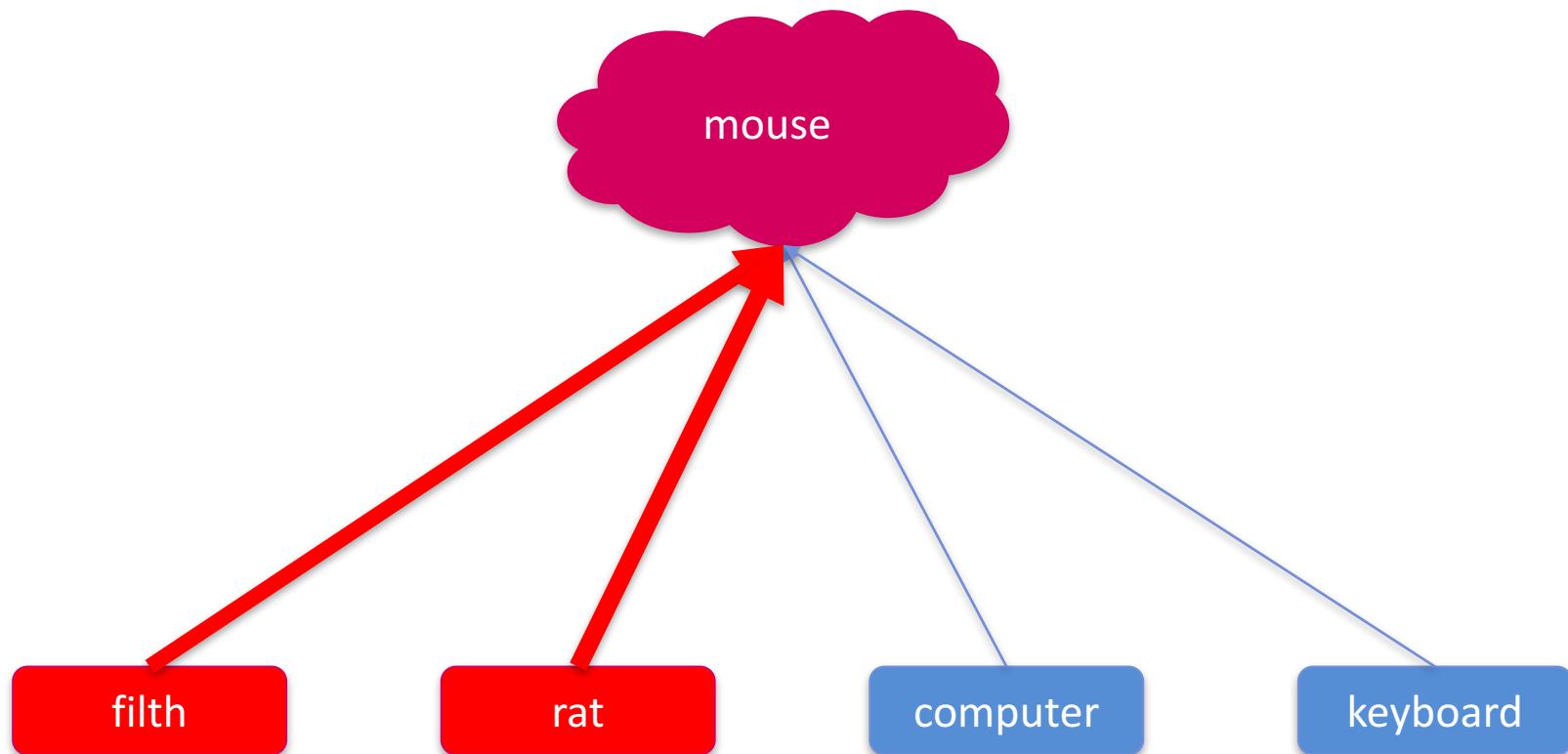
# Modifications to Expansion Methodology



# Intuition of Temporal Adaptation (Before)

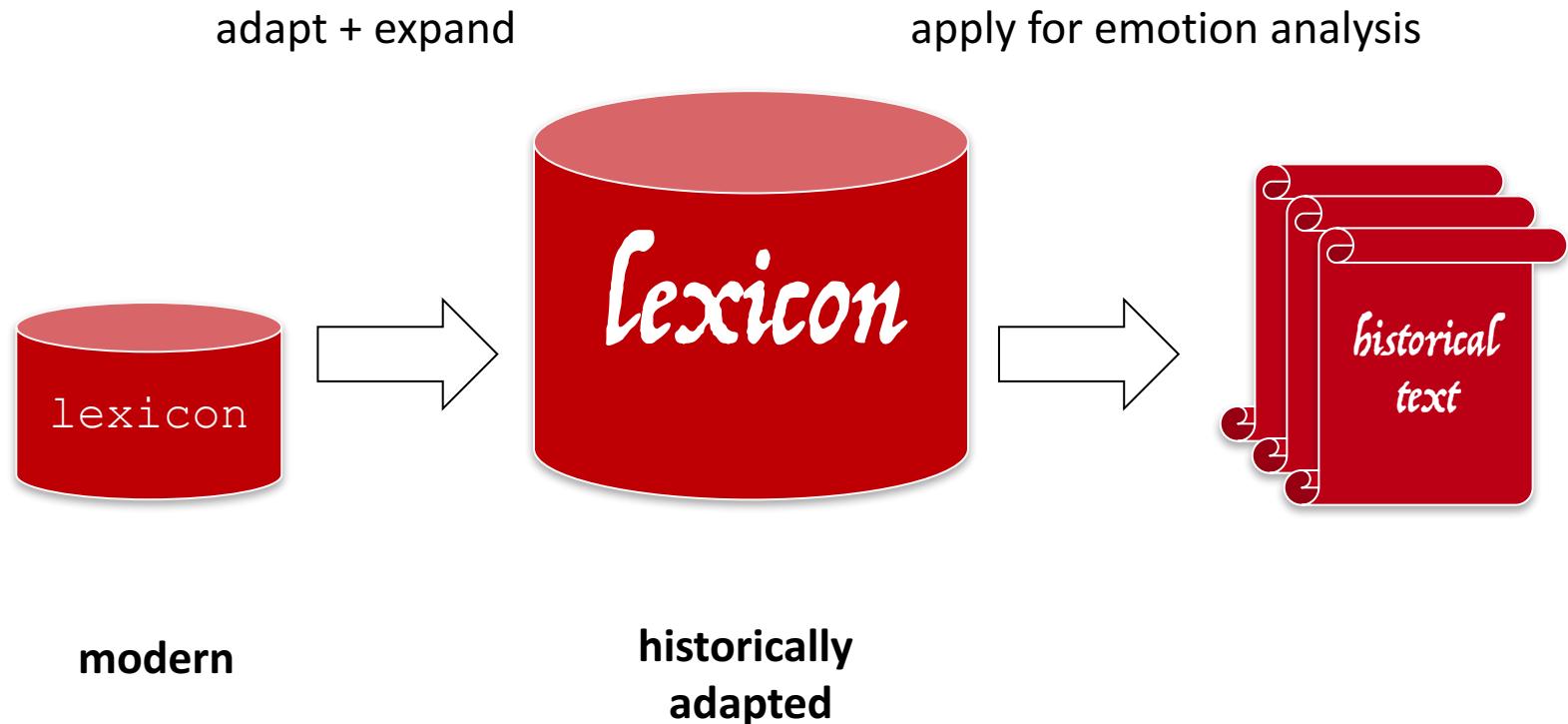


# Intuition of Temporal Adaptation (After)



# Experiments I (Lexicon Induction)

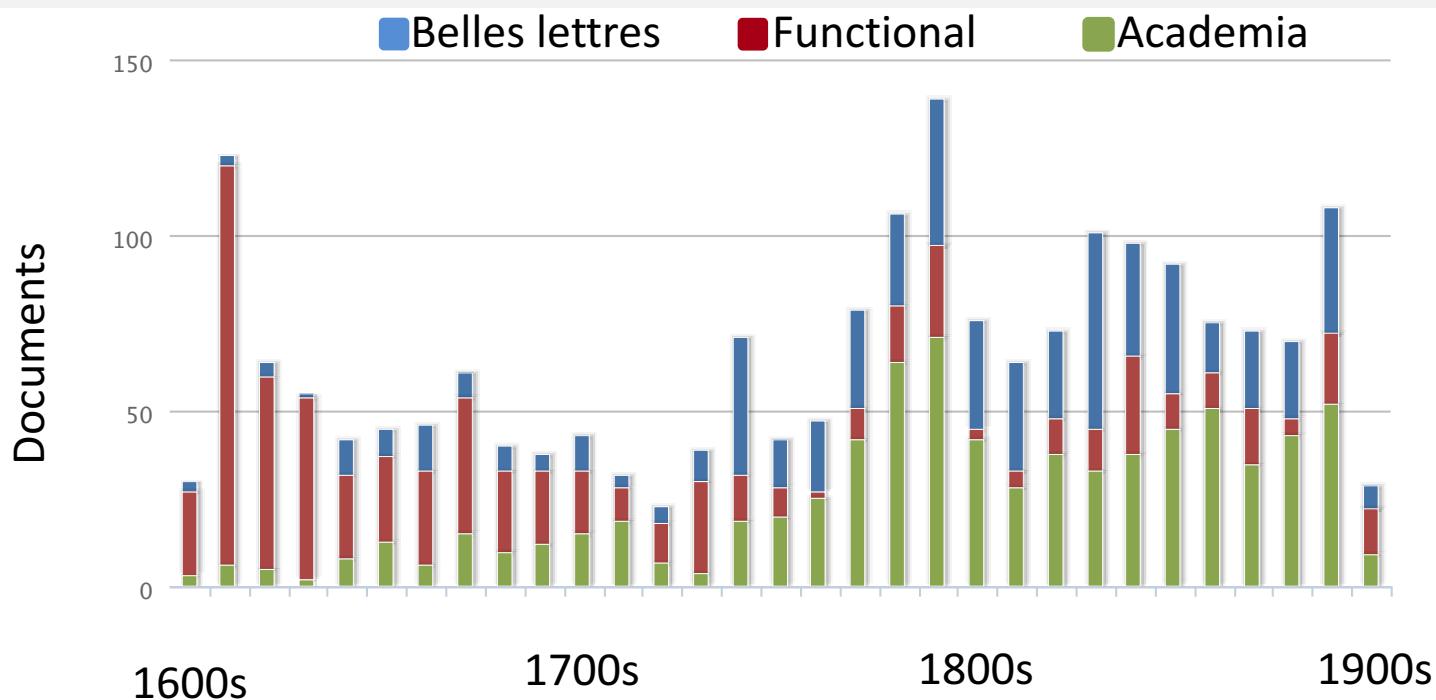
# Methodological Framework



# Resources

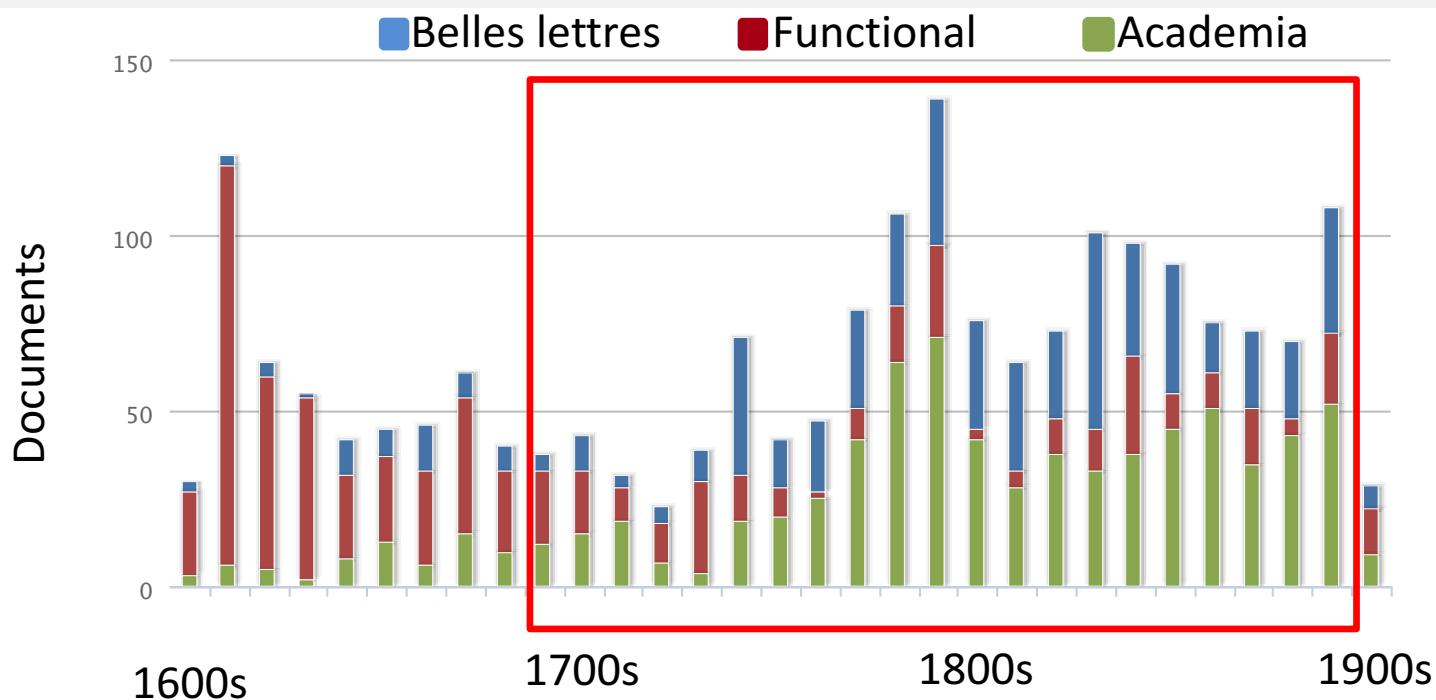
- Seed Lexicon:
  - ANGST (Schmidtke et al. (2014))
  - A thousand German word-emotion pairs
  - Well established acquisition methodology
- Target corpus:
  - Deutsches Textarchiv (DTA) [“German Text Archive”]
  - Balanced and manually transcribed
  - Rich meta-data (e.g., genre and subgenre)

# Target Corpus: DTA



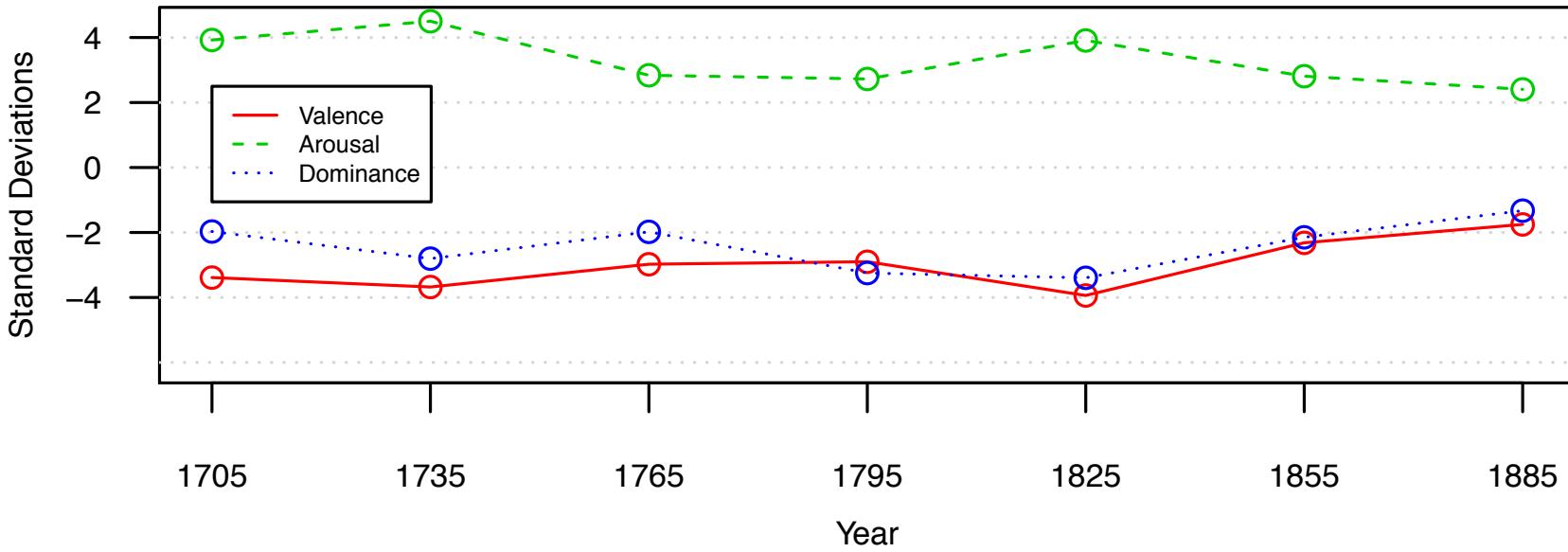
- 1<sup>st</sup> third shows different genre distribution
- Individual decades comprise too little text
  - Aggregate 30-years slices
  - Select 1690-1899 (~ 1k documents, 7 slices)

# Target Corpus: DTA



- 1<sup>st</sup> third shows different genre distribution
- Individual decades comprise too little text
  - Aggregate 30-years slices
  - Select 1690-1899 (~ 1k documents, 7 slices)

# Affective Course of *Sünde* [Sin]



- Strong increase in Valence since onset of Enlightenment
- Backing up our findings with corpus linguistics...

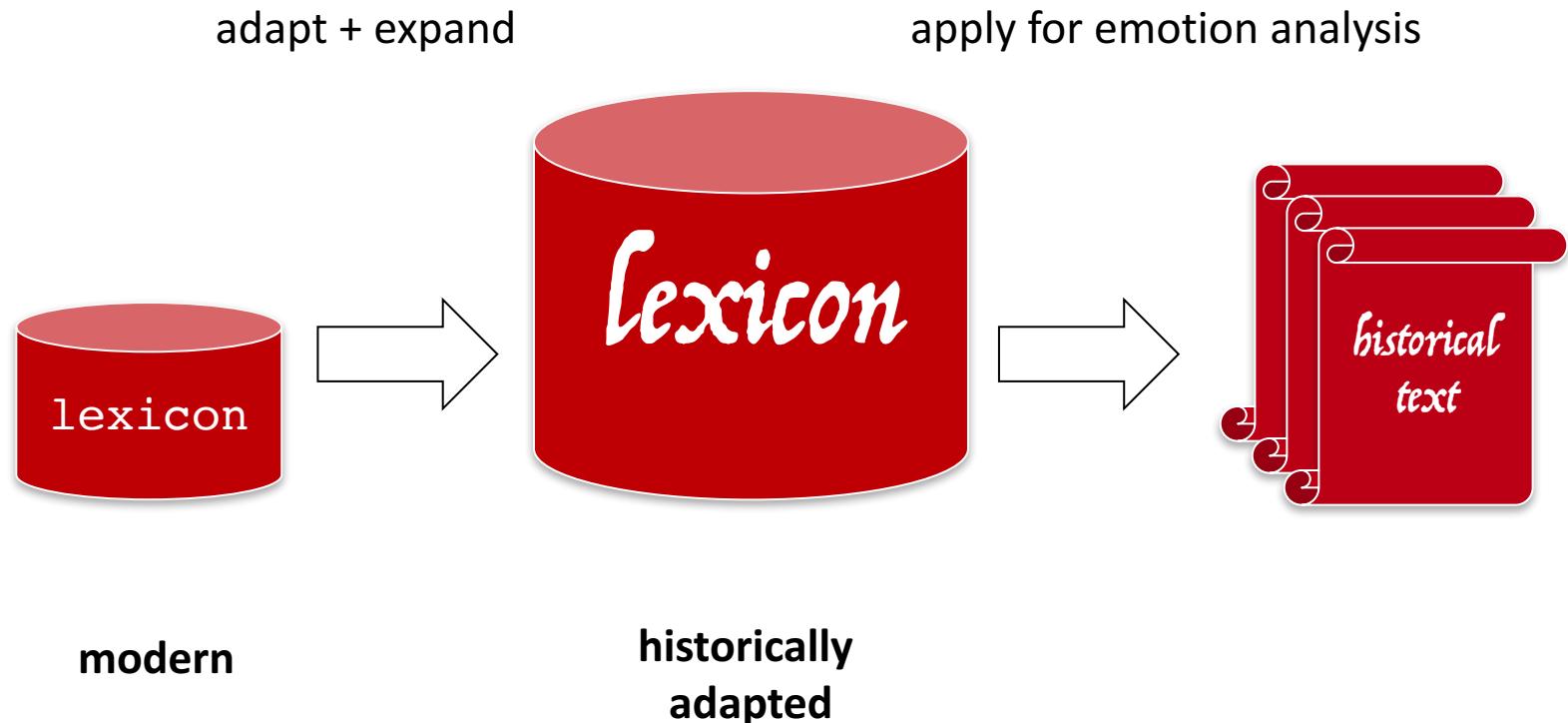
# Top 10 Collocations for *Sünde* [Sin]

Rank	Lemma and Translation			
	1690s		1890s	
1	todt-	death (prefix use)	Lamm	lamb
2	Erzürnung	infuriation	hinwegnehmen	to take away
3	läßlich	minor (clerical)	Verzeihung	forgiveness
4	beichten	to confess	Ausschweifung	excess
5	Nachlaß	abatement	Gotte	god
6	Grobheit	crudeness	Schande	shame
7	verschweigen	to conceal	Reue	repentance
8	beweinen	to beweep	Ärgernis	nuisance
9	pichen	to pitch	Laster	vice
10	beichten	to confess	aufrichtig	sincere

# Experiments II

## (Application to Historical Text)

# Methodological Framework



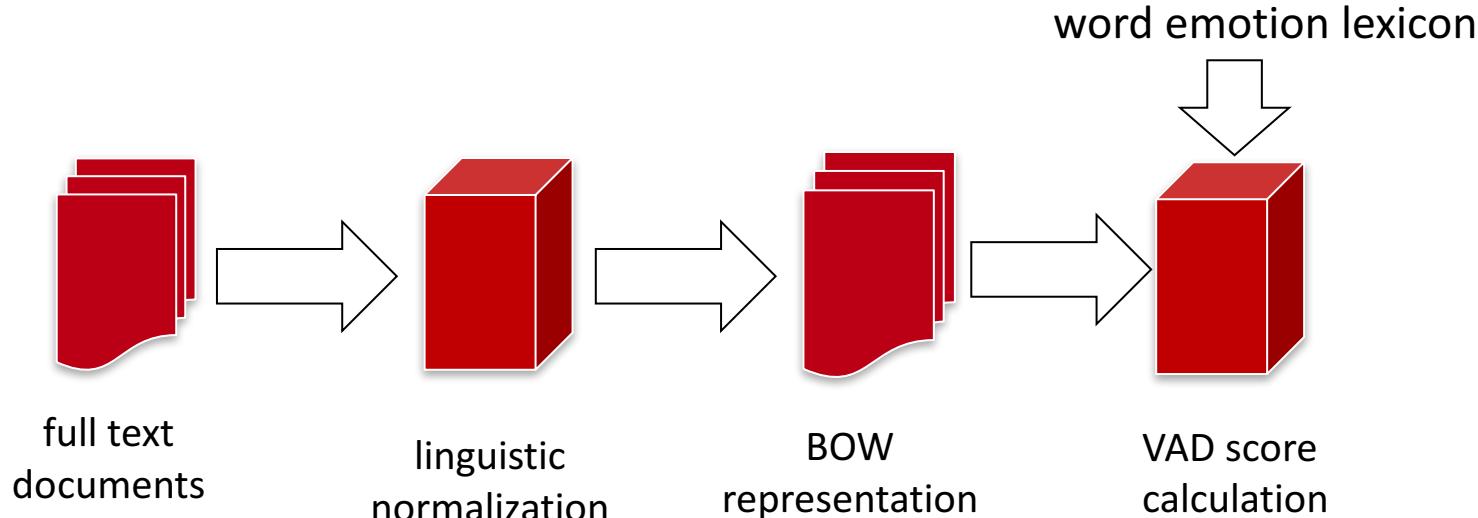
# Sentence-Level Emotion Prediction: JEmAS

(Buechel & Hahn, ECAI 2016)

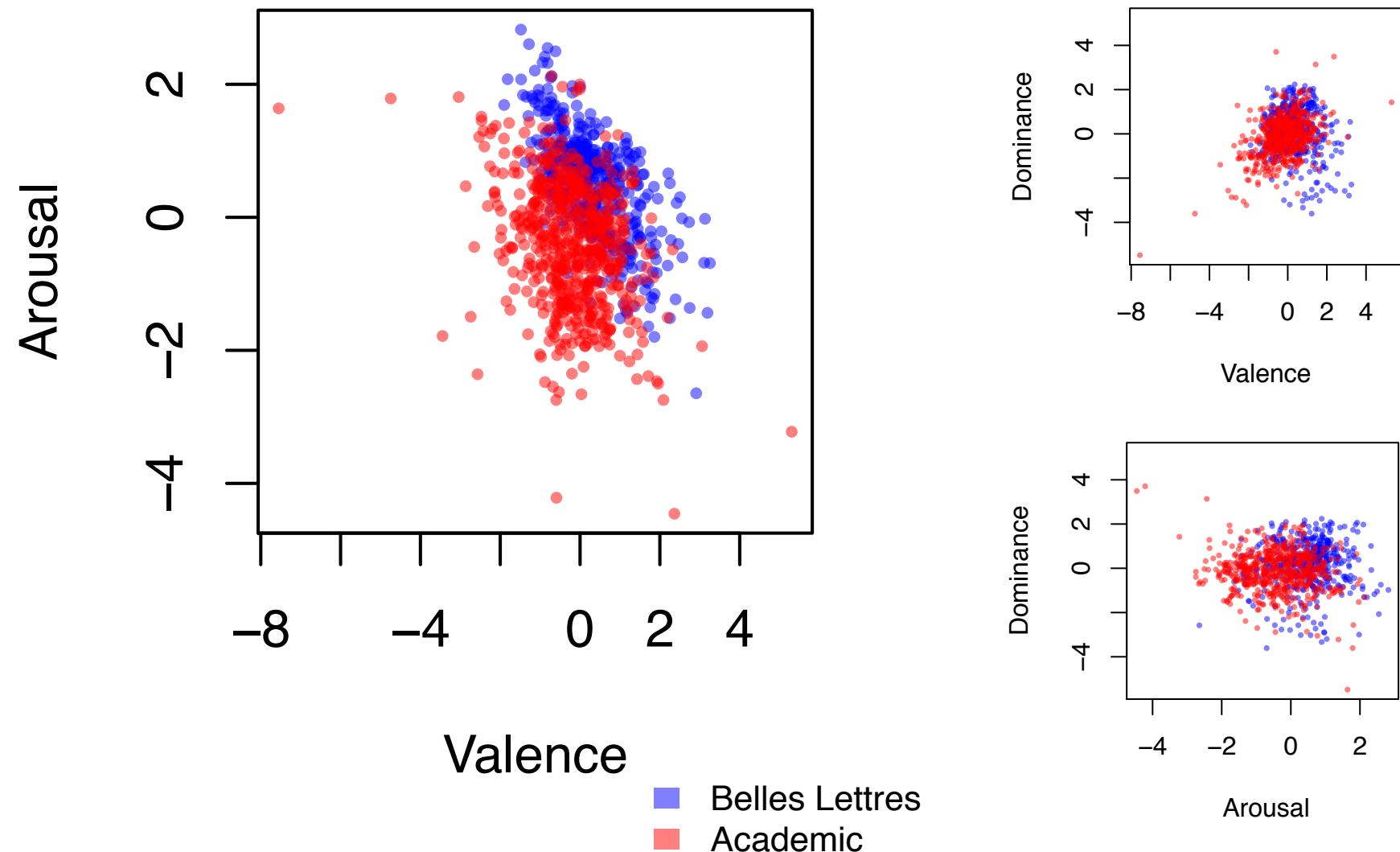
**Available:**

<https://github.com/JULIELab/JEmAS>

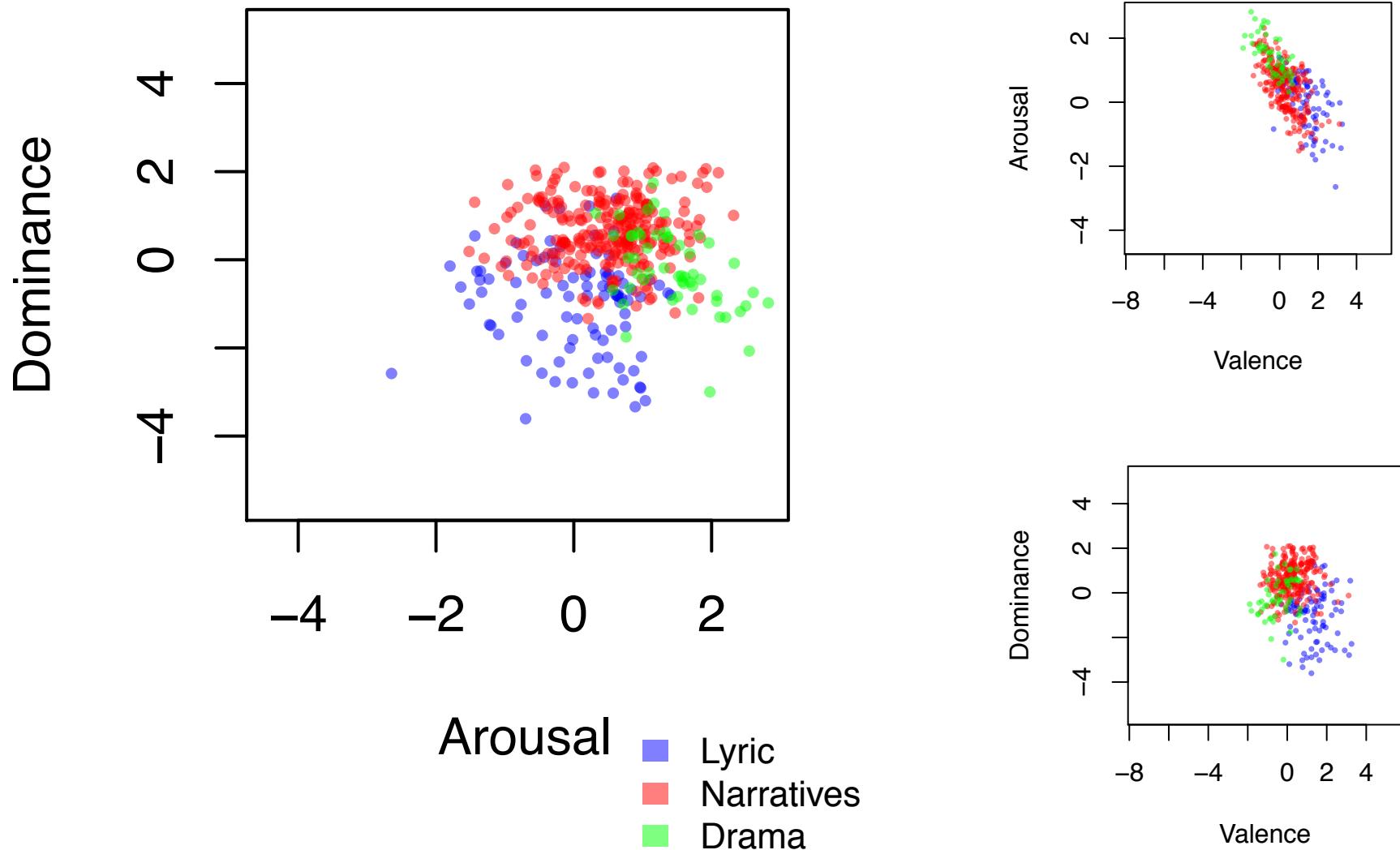
(sunshine, <8, 3, 5>)  
(terrorism, <2, 7, 3>)  
(calm, <7, 2, 7>)



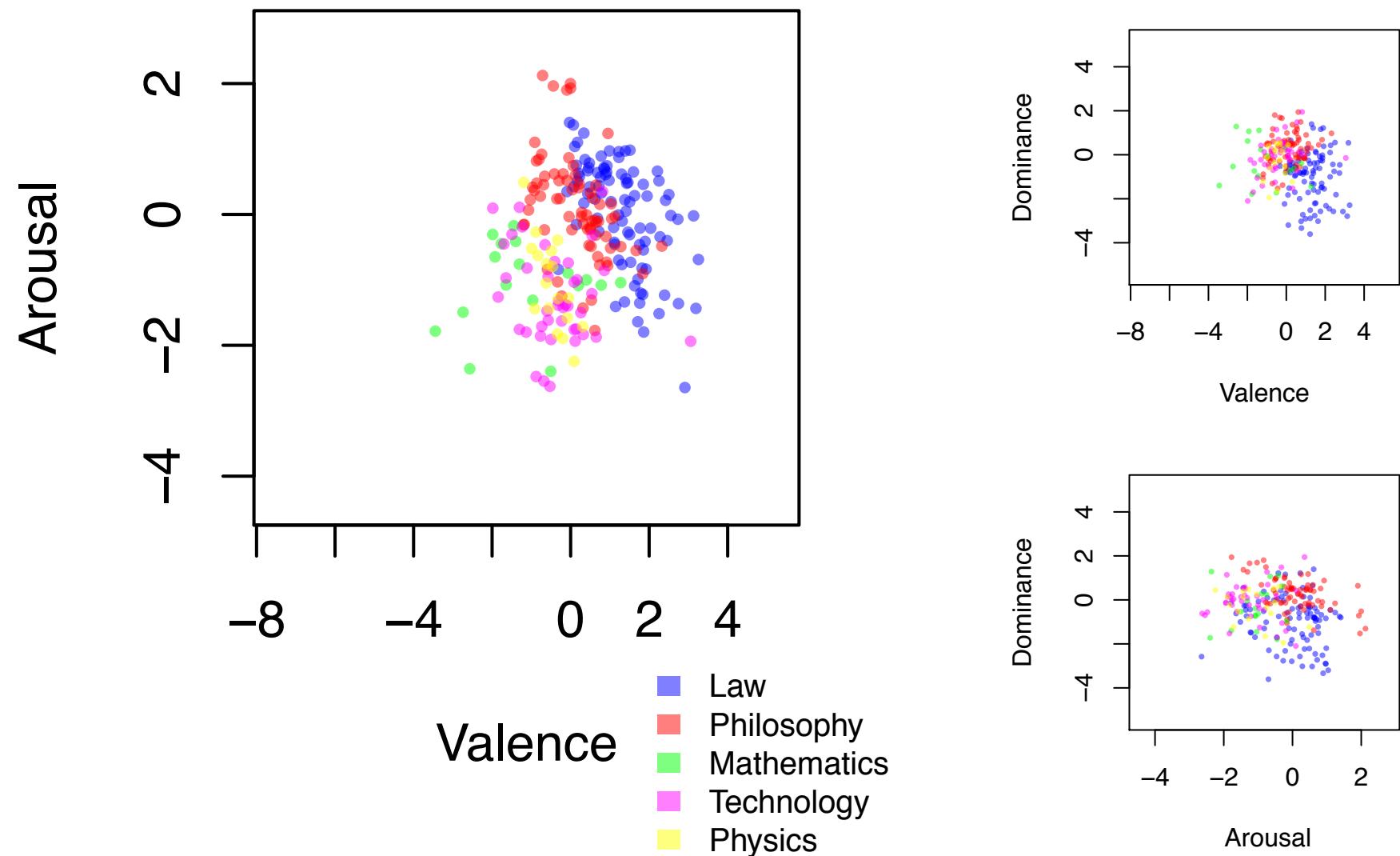
# Distinction of DTA Main Classes



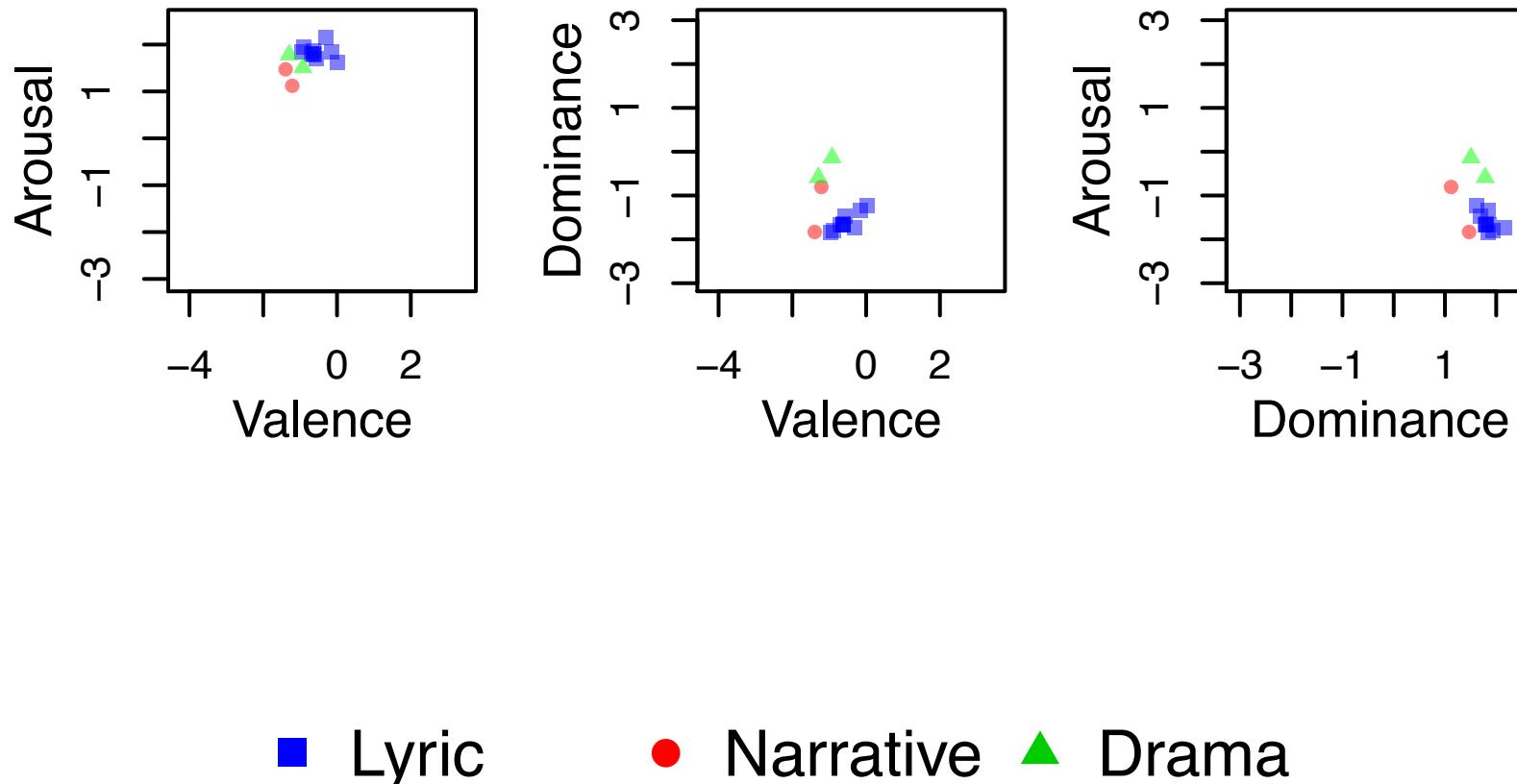
# Distinction between Principal Literary Forms



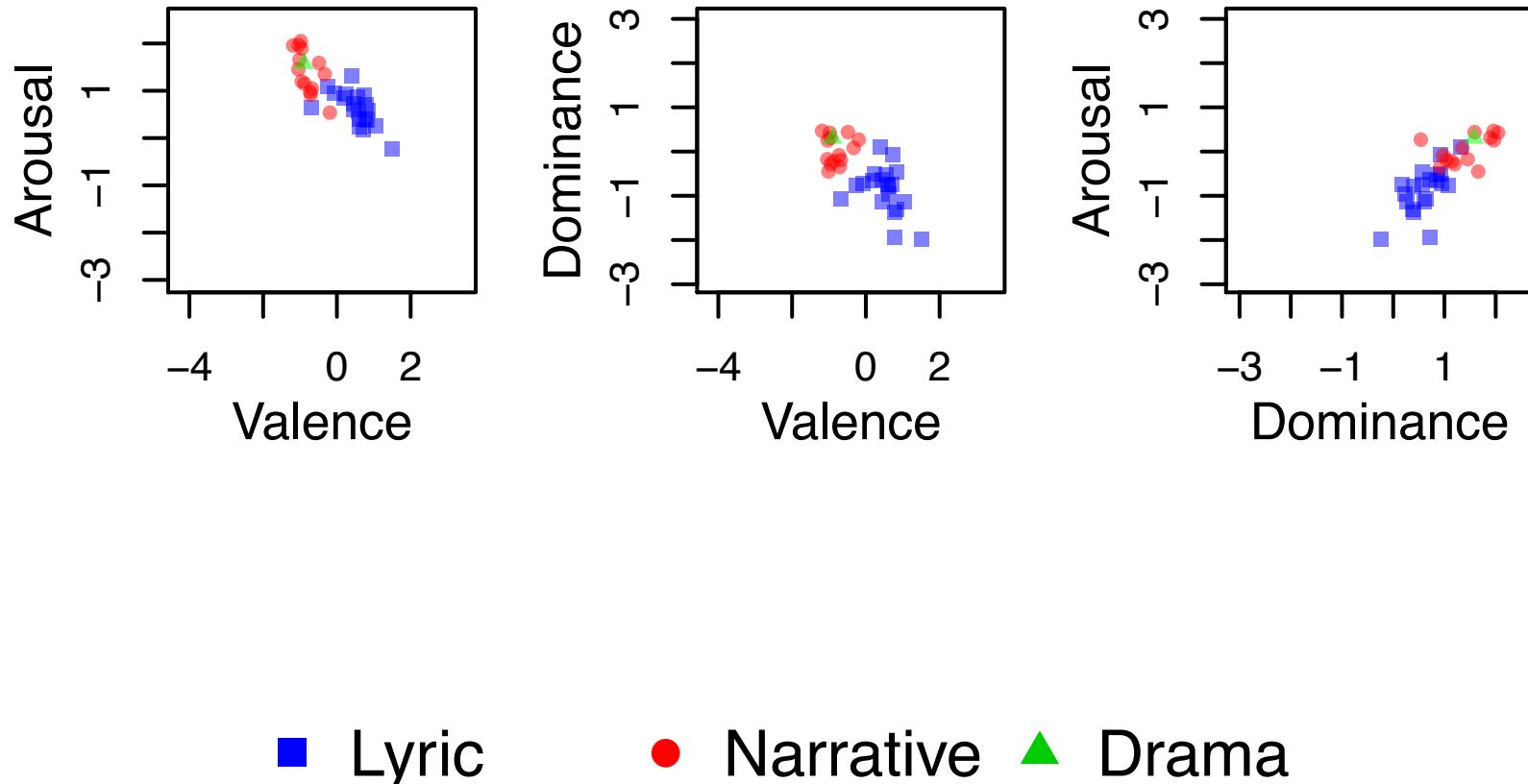
# Distinction of Academic Subclasses



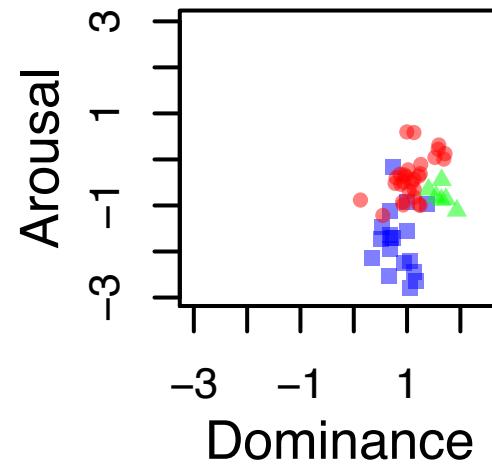
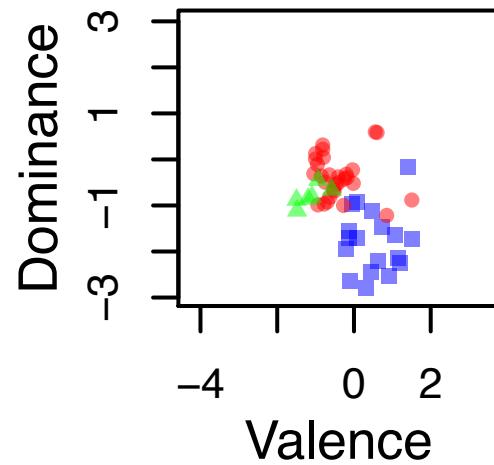
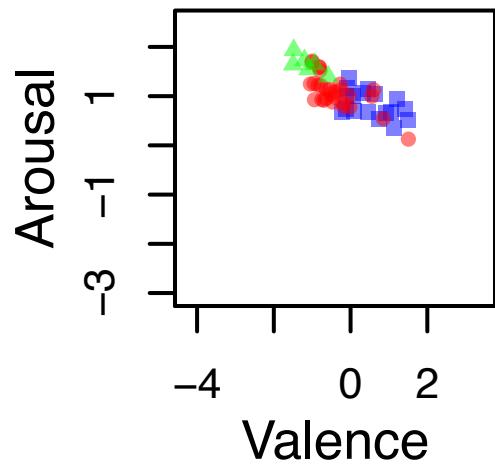
# Development of Literary Forms (1690-1719)



# Development of Literary Forms (1720-1749)



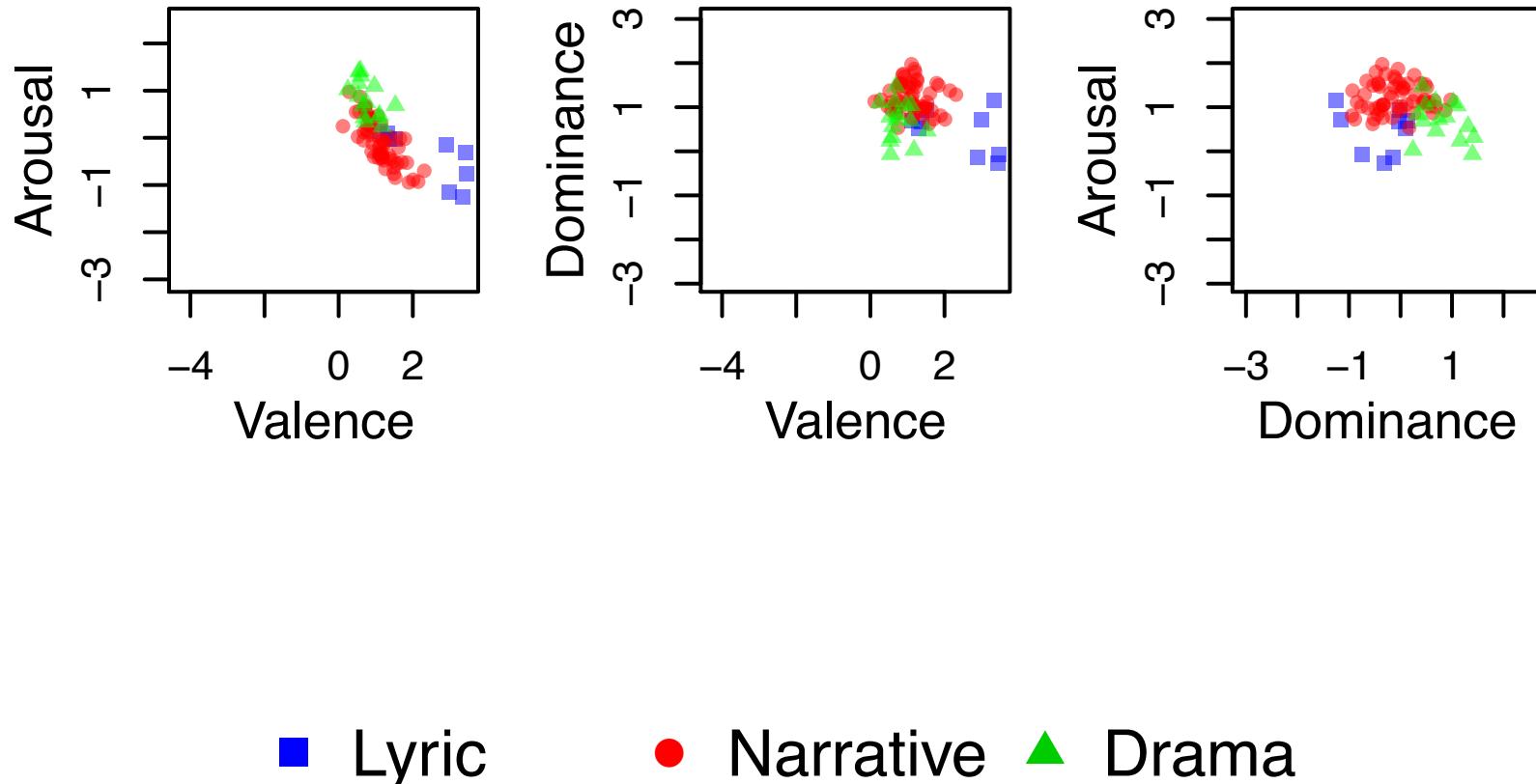
# Development of Literary Forms (1750-1779)



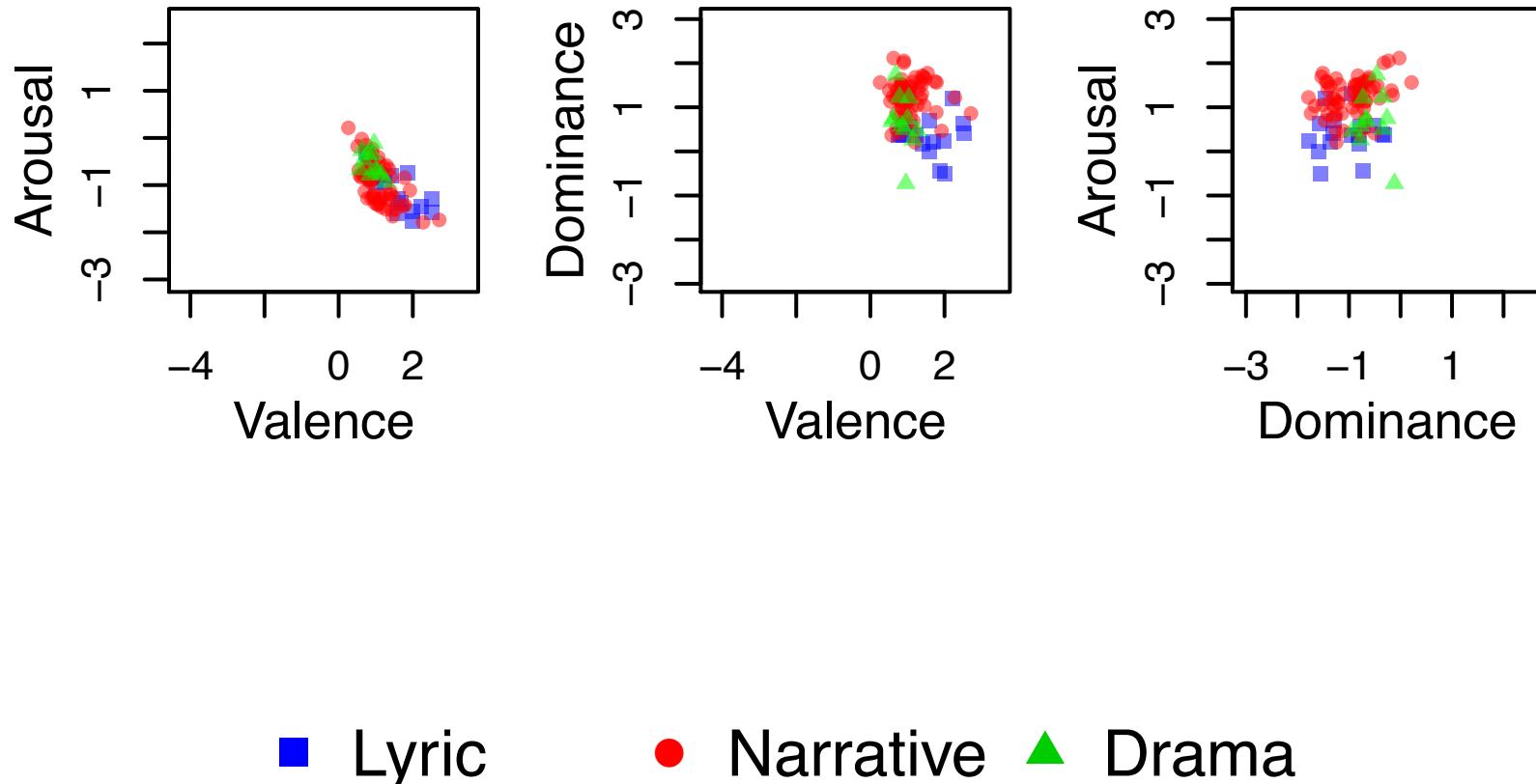
■ Lyric

● Narrative ▲ Drama

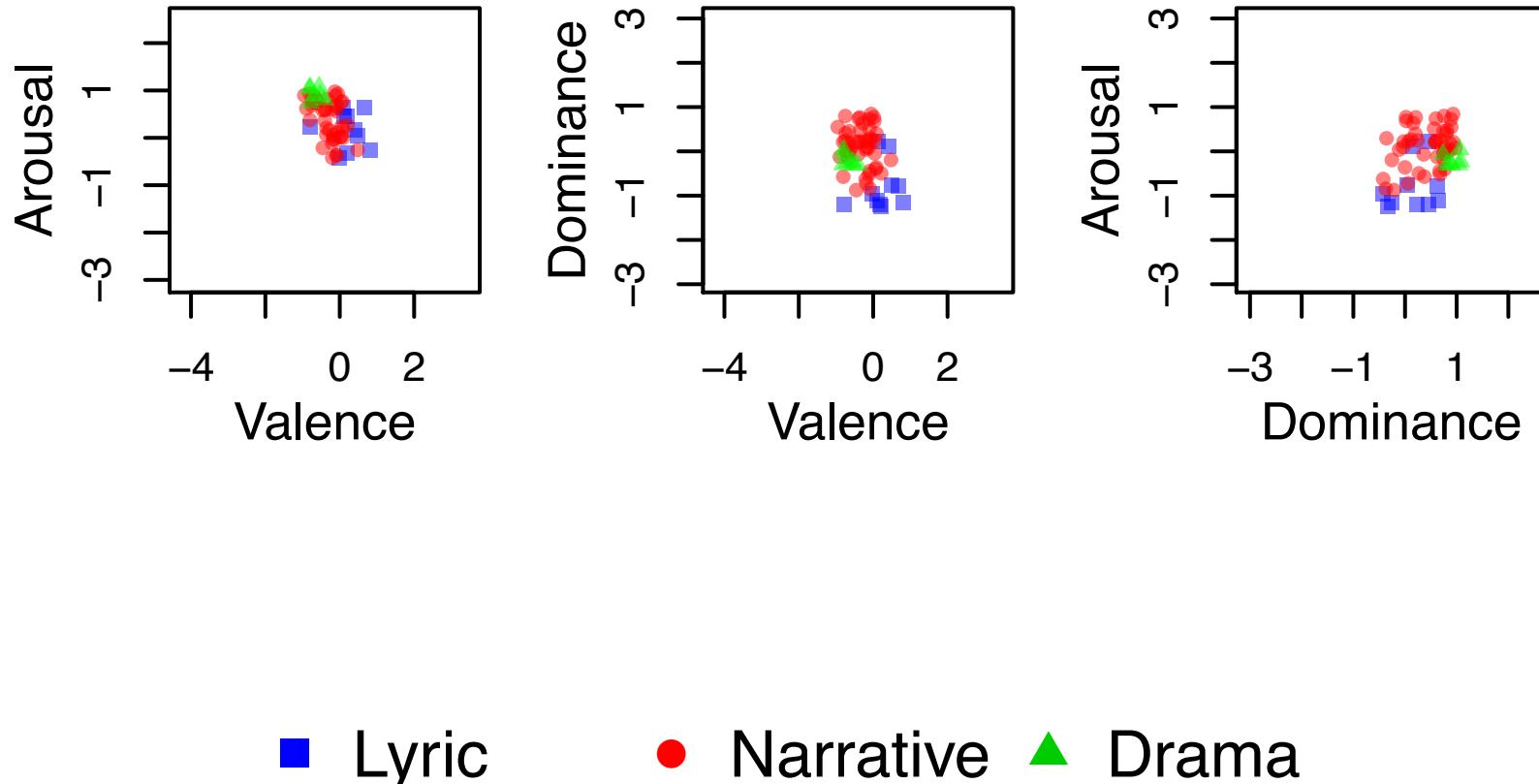
# Development of Literary Forms (1780-1809)



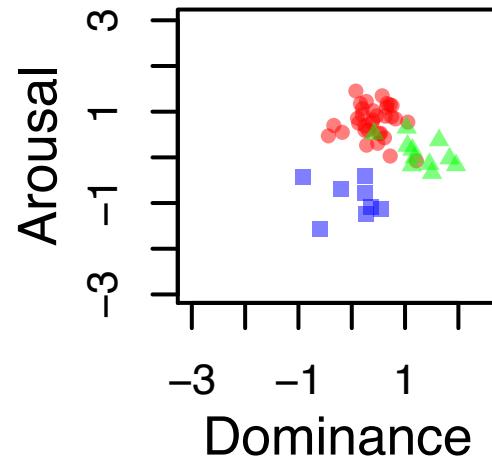
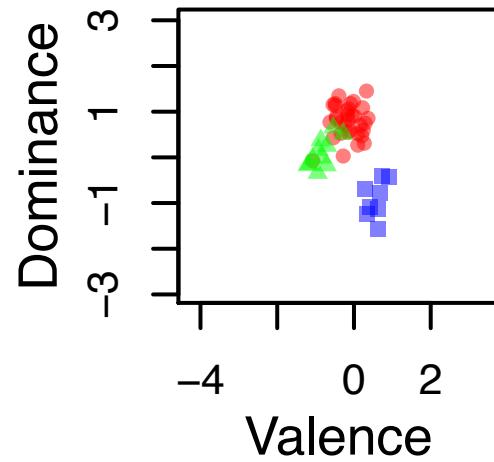
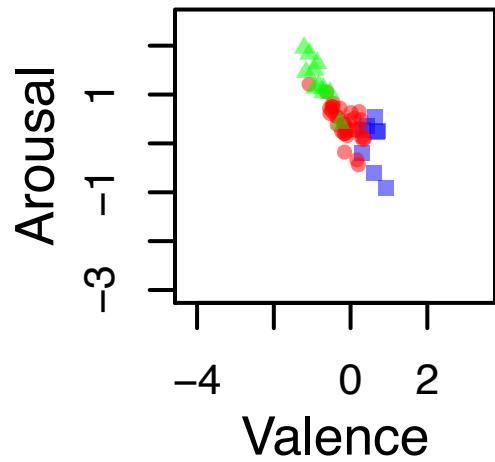
# Development of Literary Forms (1810-1839)



# Development of Literary Forms (1840-1869)



# Development of Literary Genres (1870-1899)



■ Lyric

● Narrative ▲ Drama

# Conclusion

# Conclusion

- Methodological framework for quantifying emotion in non-contemporary text
  - Combination of temporally-specific lexicon expansion and lexicon-based emotion analysis
  - Application to German historical corpus as proof of concept
  - Qualitative validation on word and text level
- Genres show distinct and dynamic emotional patterns!
- Resources available:  
<https://github.com/JULIELab/HistEmo>



# The Course of Emotion in Three Centuries of German Text

## A Methodological Framework



Sven Buechel<sup>1</sup>, Johannes Hellrich<sup>1,2</sup> and Udo Hahn<sup>1</sup>

<sup>1</sup> Jena University Language & Information Engineering  
(JULIE) Lab

<http://www.julielab.de>

<sup>2</sup> Graduate School  
'The Romantic Model'

<http://www.modellromantik.uni-jena.de>

Friedrich Schiller University Jena, Jena, Germany