



HAL
open science

Towards emotion analysis for Alsatian theater

Qinyue Liu, Pablo Ruiz Fabo, Delphine Bernhard

► **To cite this version:**

Qinyue Liu, Pablo Ruiz Fabo, Delphine Bernhard. Towards emotion analysis for Alsatian theater. Computational Humanities Research (CHR) 2023, Dec 2023, Paris, France. 2023, 10.5281/zenodo.8404252 . hal-04213017

HAL Id: hal-04213017

<https://hal.science/hal-04213017v1>

Submitted on 5 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Emotion analysis

- Complex task often carried out in Computational Literary Studies (Kim & Klinger, 2021)
- **Approaches**
 - Rule and lexicon-based (e.g. NRC, Mohammad & Turney, 2013)
 - Supervised learning, transfer learning with contextual LM (e.g. Grisot et al., 2022; Dennerlein et al., 2023)
- **Applications**
 - Emotion arc computation (Teodorescu et al., 2023)
 - Subgenre characterization
- In the case of Alsatian, **low-resource** scenario

The MeThAL Alsatian theater corpus

- **Alsatian?**
 - Germanic varieties from Alsace (Eastern France)
- 77 **plays** (> 600K tokens), TEI-encoded
 - 51 without prior electronic text → OCR + corrections
- 27 **authors**
 - Minor and major ones
 - Female and male
 - Variety of origins
- Variety of dramatic **subgenres**
- Rich bibliographic and **character metadata**
 - **Social variables**: gender, socioprofessional group, age, ...



Emotion analysis for Alsatian

ELAL: Emotion Lexicon for Alsatian

Bernhard & Ruiz (2022)

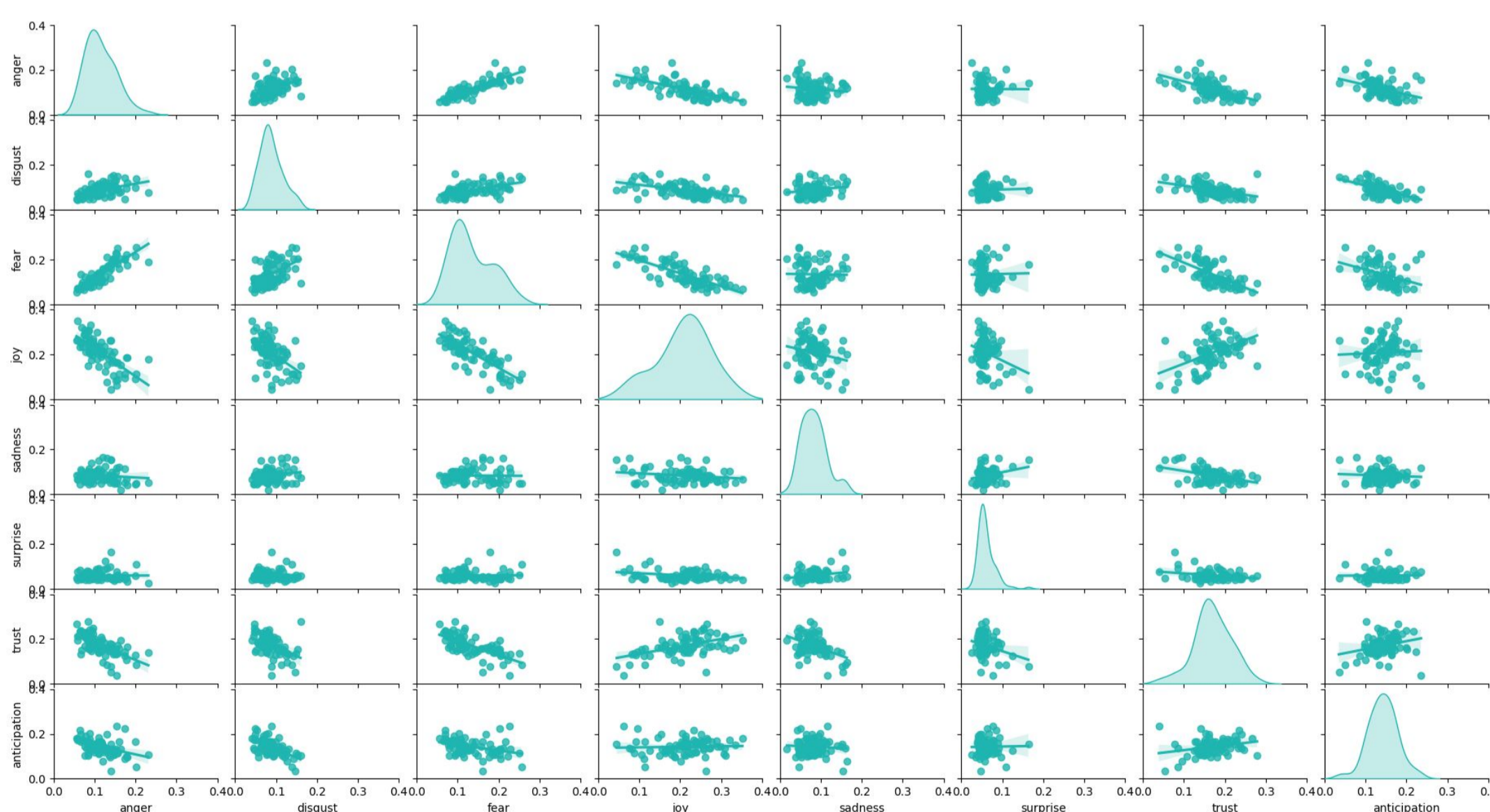
- Based on French-Alsatian **bilingual lexica** and **emotion lexica** for French and German (NRC, FEEL) and German (NRC)
 - Terms are aggregated into a large **network**
 - Emotion scores for the large-resource languages are projected onto Alsatian terms
- Handles Alsatian's huge **orthographic variability**, matching variants with graphic similarity measures (+ manual correction)
 - 3,273 entries
 - 11,920 Alsatian forms

EDYTHA: Emotion Dynamics for Theater in Alsatian

- Inspired by TED (Twitter Emotion Dynamics) tool (Vishnubhotla & Mohammad, 2022) with important modifications
- Accepts **TEI**-encoded input
- Integrates Alsatian **tokenizer** (Bernhard, 2018). Applies ELAL lexicon to compute **emotion scores** averaged over corpus chunks
 - Creates **subcorpora** based on plays' and character **metadata**
 - **Orthographic variant detection** thanks to ELAL
- Emotion term **weighting schemes** based on tf-idf:
 - Prevents frequent terms from skewing scores
- **tf-idf configuration** options:
 - idf taking each play as a document
 - idf taking each character speech turn as a document
- Compares emotions by **character social variables** (e.g. do socially disfavoured characters show more negative emotion?)

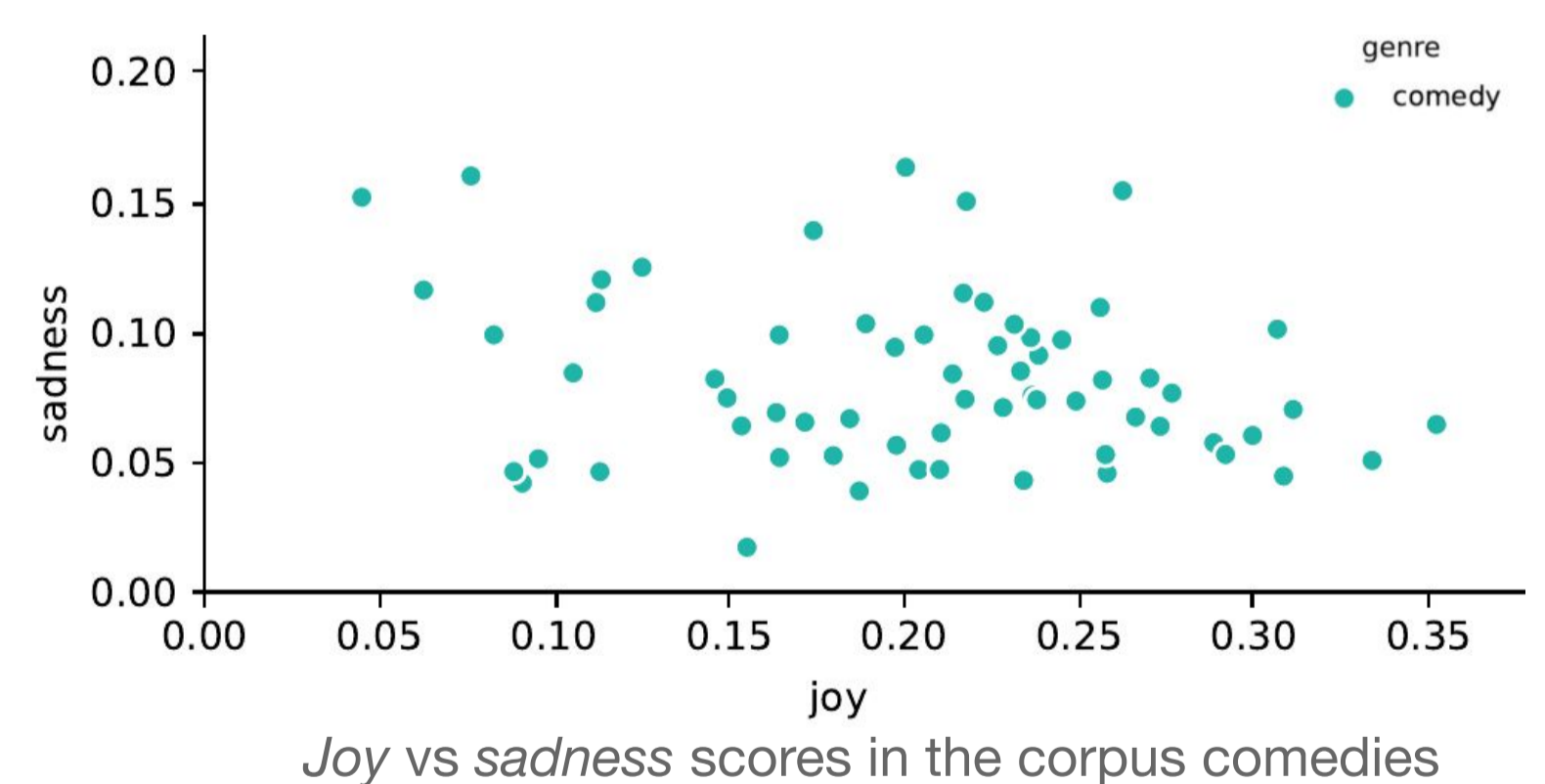
EDYTHA: Outputs

EDYTHA **aggregates scores** and produces **visualizations**, partly inspired by the Katharsis tool (Schmidt et al., 2019)

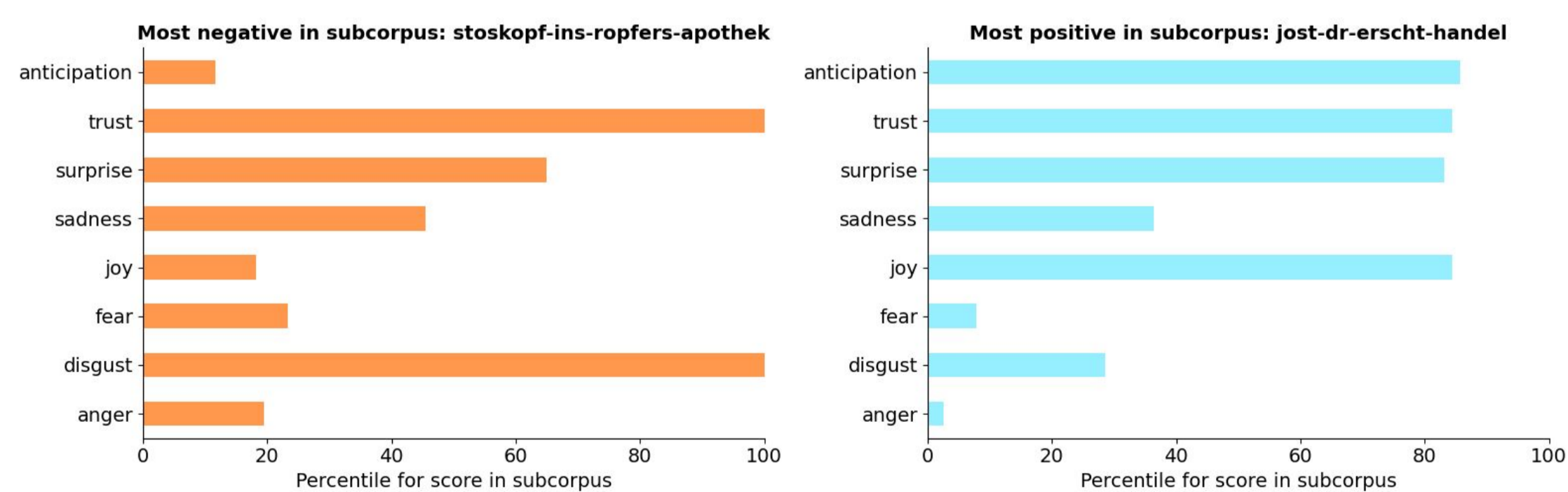


Correlation between different emotion scores in the entire MeThAL corpus

<https://git.unistra.fr/methal/edytha>



Joy vs sadness scores in the corpus comedies



Emotion scores in plays with the lowest and highest polarity (valence): For each emotion type, percentile with respect to all plays in the selected subcorpus

References

- Bernhard, D. & Ruiz, P. (2022). ELAL: An Emotion Lexicon for the Analysis of Alsatian Theatre Plays. In *LREC 2022*. Dennerlein, K. Schmidt, T & Wolff, C. (2023). Computational emotion classification for genre corpora of German tragedies and comedies from 17th to early 19th century, *Digital Scholarship in the Humanities*. Grisot, G. Rebor, S., Herrmann, J. B. & Pennino, F. (2022). Sentiment lexicons or BERT? A comparison of sentiment analysis approaches and their performance. In *Digital Humanities 2022*. Kim, & Klinger (2021). A Survey on Sentiment and Emotion Analysis for Computational Literary Studies, *Zeitschrift für digitale Geisteswissenschaften*. Mohammad, S., & Turney, P. (2013). Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29. Ruiz, P., Bernhard, D., Briand, A. & Werner, C. (2024). Computational drama analysis from almost zero electronic text. In M. Andresen & N. Reiter (eds). *Computational Drama Analysis: Reflecting Methods and Interpretations*. Schmidt, T. Burghardt, M., Dennerlein, K. Wolff, C. (2019). Katharsis – A Tool for Computational Drametrics. In *Digital Humanities 2019*. Teodorescu, D., Mohammad, S. M. (2023). Generating High-Quality Emotion Arcs For Low-Resource Languages Using Emotion Lexicons. Vishnubhotla, K., & Mohammad, S. M. (2022). Tweet Emotion Dynamics: Emotion Word Usage in Tweets from US and Canada. In *LREC 2022*.