# TOWARDS A **GOLDENHYMNS DATASET** FOR STUDYING **DIACHRONIC TRENDS** IN THE 19TH CENTURY **DANISH RELIGIOUS HYMNS**

**EA LINDHARDT OVERGAARD, PASCALE FELDKAMP & YURI BIZZONI**

INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

CHC
CENTER FOR HUMANITIES COMPUTING

SOLIDUM PETIT IN PROFUNDIS · UNIVERSITAS ARHUSIENSIS

# INTRODUCTION

Overview:

1. Background information on Danish Religious Hymns

2. Methods

   • We present GoldenHymns, a new dataset of Danish Religious Hymns

   • Valence annotation strategies

3. Results and discussion

INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

LCHANGE | EA LINDHARDT OVERGAARD
15. AUGUST 2024 | PH.D-STUDERENDE

CHC
CENTER FOR HUMANITIES COMPUTING

# HYMNS ARE CULTURAL ARTEFACTS

- Hymnal books are the most distributed poetry books in Denmark
- Has historically been a school book, church book and book for private use
- Role of communal singing in Danish culture - a part of the national identity
- Has been widelye distributed and used since 1569
- Under-studied for historical changes

INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

LCHANGE

15. AUGUST 2024

EA LINDHARDT OVERGAARD

PH.D-STUDERENDE

CHC
CENTER FOR HUMANITIES COMPUTING

# COMMON UNDERSTANDINGS OF HYMN LITERATURE

- Older hymns depict a dualistic view: negative earthly vs. positive religious instances
- Hypothesis: polarity shifts in religious concepts and general sentiment over time

**17th century**

| Jorderiigs Guld | Earthly gold |
| Er prægtig Muld, | Is splendid soil |
| Himlen er Ene af Salighed fuld. | Heaven alone is full of bliss |
| | |
| Alle Ting maa | Everything must |
| Enden opnaa, | Reach its end, |
| Himmelens Salighed | Heaven's bliss |
| Ene skal staa! | alone shall stand |

**19th century**

| Et helligt Liv, en salig Død | A holy life, a blessed death |
| Hinanden kjærlig møde | meet each other lovingly |
| Som Fuglesangen sagte sød | Like the gentle sweet song of birds |
| | |
| En liflig Aftenrøde: | A pleasant evening glow |
| De to maa ikke skilles ad | the two must not be seperated |
| Den ene gjør den anden glad | One makes the other happy |
| Og meget liflig favner. | and embraces very sweetly |

INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

LCHANGE
15. AUGUST 2024

EA LINDHARDT OVERGAARD
PH.D-STUDERENDE

CHC
CENTER FOR HUMANITIES COMPUTING

SOLIDUM PETIT IN PROFUNDIS · UNIVERSITAS ARHUSIENSIS

# SA FOR HISTORICAL AND DANISH RESOURCES

**Danish XIXc. hymns datasets**

- Few existing resources

- Grundtvig dataset (CGR, 2019) one author

- Reformation dataset (DSLDK, 2021), represents a limited period, but are in medieval Danish highly affected by German - heavily orthographical deviations

**Sentiment Analysis for historical Danish**

- Rare SA for historical Danish

  - Allaith et al (2023) finds multilingual transformer models outperform models trained on modern Danish based on Danish lexical resources

  - Danish dictionary-based tools show comparable performance to English (Schneidermann and Petersen, 2022)

  - Danish tool Sentida shows robust performance for fiction (Lauridsen et al., 2019)

- No comprehensive SA benchmark for Danish

INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

LCHANGE | EA LINDHARDT OVERGAARD
15. AUGUST 2024 | PH.D-STUDERENDE

CHC
CENTER FOR HUMANITIES COMPUTING

SOLIDUM PETIT IN PROFUNDIS · UNIVERSITAS ARHUSIENSIS

# METHODS: CORPUS

**Verses** of 65 **Danish religious hymns** were annotated for valence (0-10) by 2 annotators. The hymns are from three focal years: 1798, 1857, 1873.

| Hymns | Verses | Words | x̄ Verses pr. hymn | Period |
|---|---|---|---|---|
| 65 | 1,914 | 10,303 | 32.9 | 1798-1873 |

Three text-settings:

- Original text
- Modernized version with modern spelling norms
- Validated English translations

Valence scores per verse line

- Mean human-annotated (n=2) valence score
- Automatic annotation scores (dictionary- and transformer based tools)

# METHODS: MODERNIZATION AND TRANSLATION

**Modernized Hymns Section**

- Modernized by prompting ChatGPT 3.5
- Manually revised and edited by two scholars
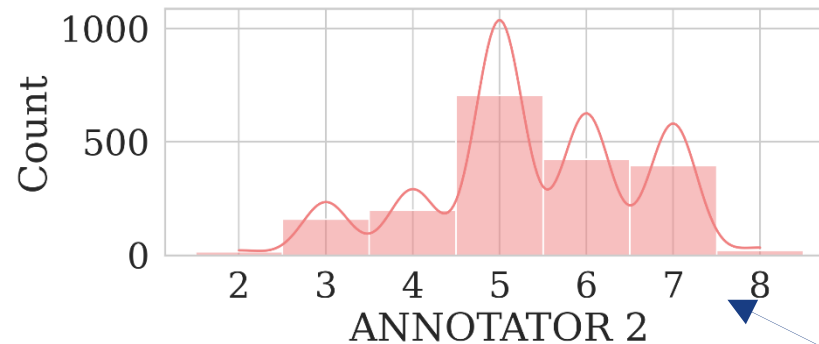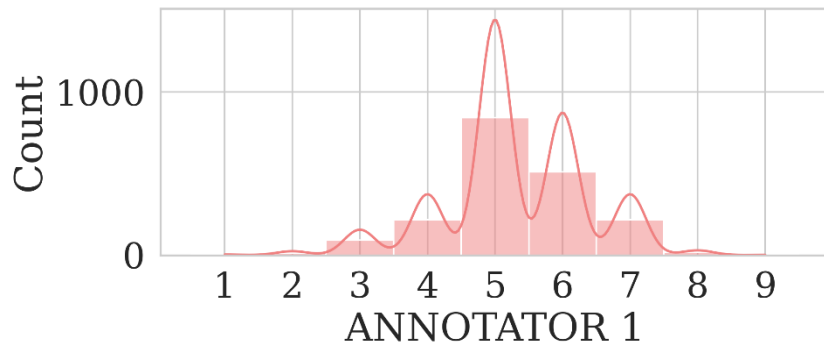
**Translated Hymns Section**

- Translated with deep translator package in python to retrieve google translated sentences.
- Manually revised and edited to:
  - keep verse lines and syntactical structure
  - keep semantic

INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

LCHANGE

15. AUGUST 2024

EA LINDHARDT OVERGAARD

PH.D-STUDERENDE

# METHODS: HUMAN ANNOTATION

- 2 scholars annotate 1914 verses
- 0 (most negative) to 10 (most positive)
- Fragmentary nature, challenging task
- Spearman's r between them 0.73

| Sentence | Score |
|---|---|
| J Mistvivl, Angest, Smerte<br>(M) I mistvivl, angst og smerte<br>(En) In doubt, anxiety, and pain | 2.0 |
| Ungdomsliv i Morgenrøden<br>(M) Ungdomsliv i morgenrøden<br>(En) Youthful life in the morning's red glow | 7.0 |
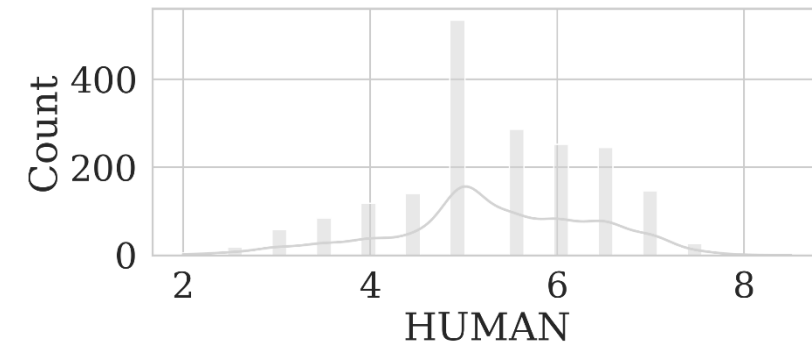


Slight positive 'bias' in the hymns

INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

LCHANGE
15. AUGUST 2024

EA LINDHARDT OVERGAARD
PH.D-STUDERENDE

# METHODS: MODELS

Score verses in the hymns for valence (no contexts)

(We did not use LLMs)

Sentiment models with wide application in Danish:

- Afinn
- Sentida
- Asent

Dictionary- and rule-based

- Senda
- Alexandra Institute sentiment base
- RoBERTa clm multilingual base sentiment
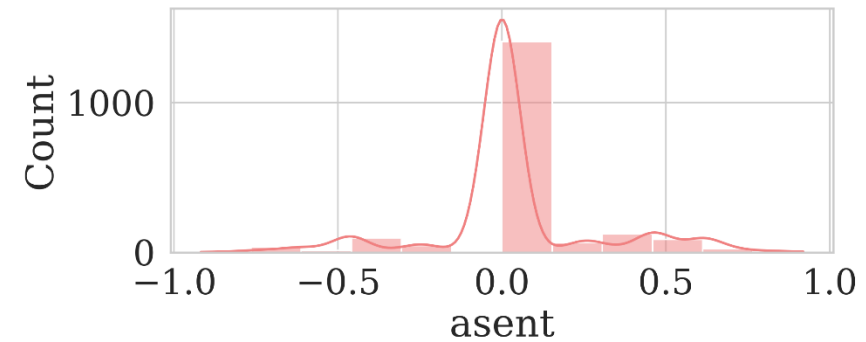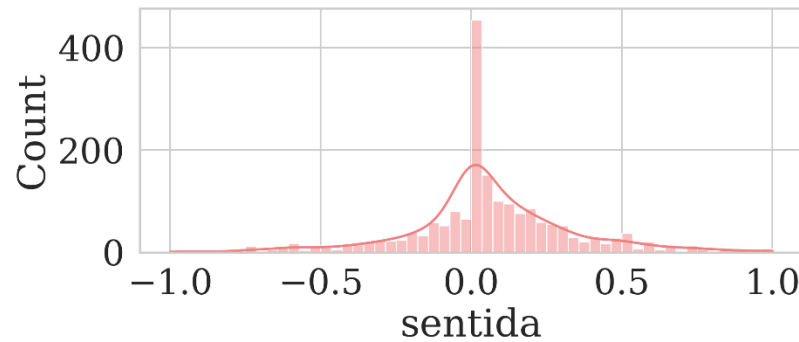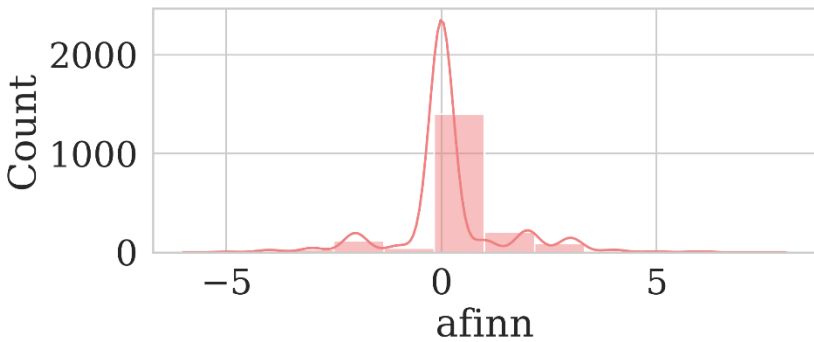
Transformer-based

# METHODS: CONVERTING SCORES AND DISTRIBUTIONS

- Converted the categorical outputs to continuous SA scores by using the confidence scores
- -1 (most negative) to 1 (most positive), 0 being neutral
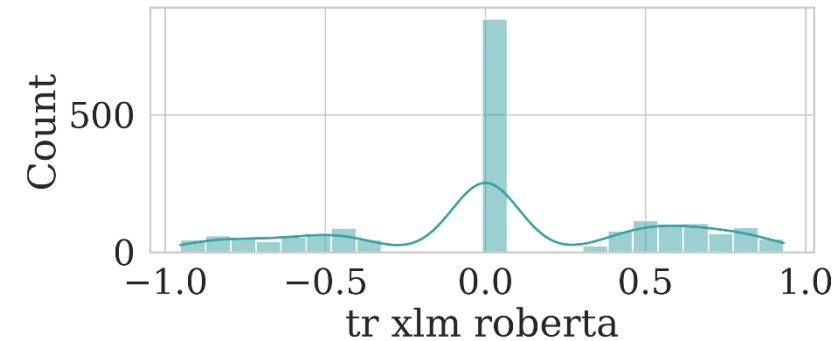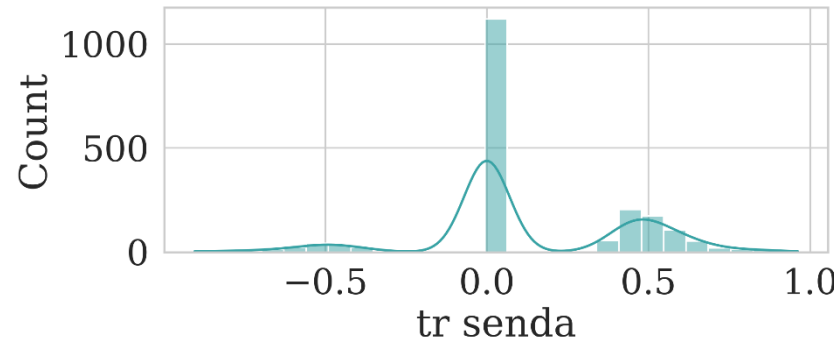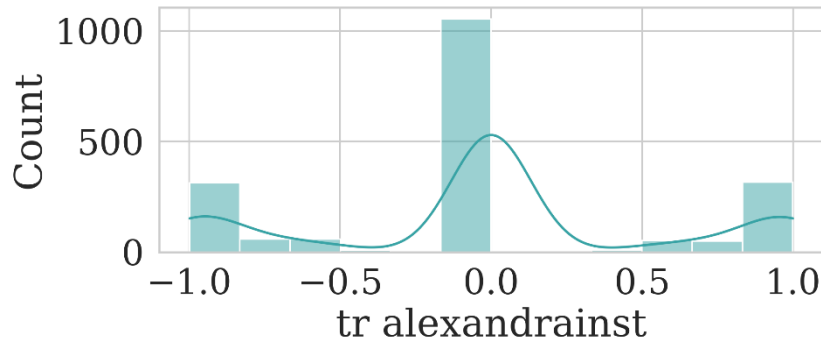


CLASS: Positive, CONF.: 0.78      +0.78
CLASS: Negative, CONF.: -0.89     -0.89
CLASS: Neutral, CONF: .77          0.0

# RESULTS: AUTHENTIC VS. MODERNIZED HYMNS

|  | Alex.inst. | Senda | RoBERTa | Asent | Afinn | Sentida |
|---|---|---|---|---|---|---|
| **Hymns original** | 0.39 | 0.32 | 0.39 | 0.40 | 0.39 | **0.49** |
| **Hymns modernized** | 0.42 | 0.35 | 0.46 | 0.41 | 0.40 | **0.53** |

- Dictionary-based tools perform comparable to or better than transformer-based methods
- All methods improve when applied on modernized versions
- Sentida performs best in both cases
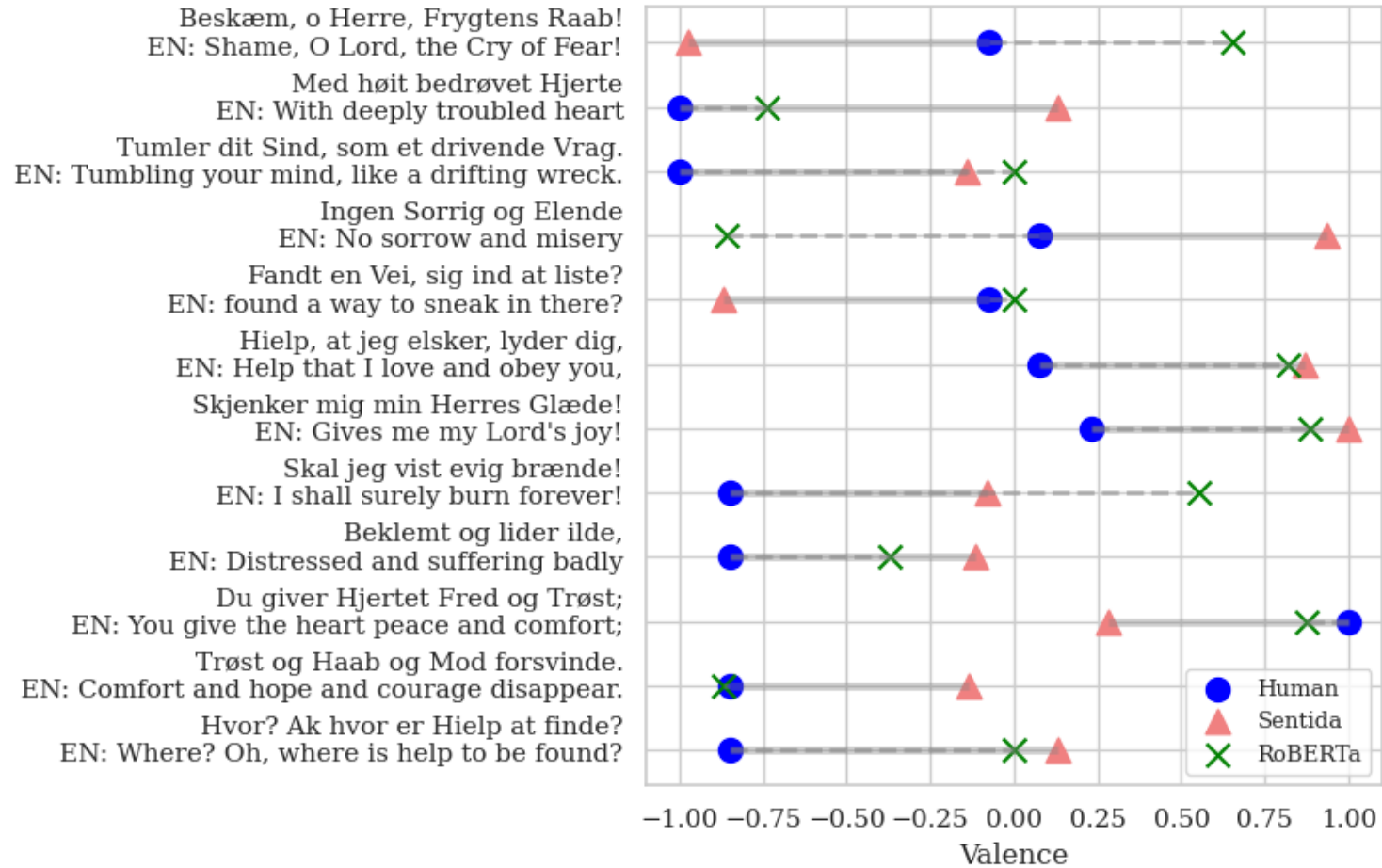
# RESULTS: AUTHENTIC VS. MODERNIZED HYMNS

|  | Alex.inst. | Senda | RoBERTa | Asent | Afinn | Sentida |
|---|---|---|---|---|---|---|
| **1798** | 0.29 (0.36) | 0.30 (0.36) | 0.36 (0.40) | 0.35 (0.33) | 0.36 (0.34) | **0.43 (0.44)** |
| **1857** | 0.39 (0.36) | 0.36 ( 0.38) | 0.43 (0.51) | 0.41 (0.42) | 0.38 (0.39) | **0.49 (0.53)** |
| **1873** | 0.43 (0.48) | 0.30 (0.33) | 0.37 (0.46) | 0.40 (0.44) | 0.40 (0.44) | **0.51 (0.56)** |

- Models generally improve over time
- This trend is the same, though weakened, when applied on modernizations
- Sentida performs best in all cases

# WHERE HUMANS AND MODELS DISAGREE



INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

LCHANGE

15. AUGUST 2024

EA LINDHARDT OVERGAARD

PH.D-STUDERENDE

# CONCLUSION

- Dictionary-based methods outperform transformer-based models, contrary previous studies, where multilingual transformer-based models performs best
  - this method might have advantages in certain domains, such as hymns - maybe poetic genres in general

- We report an overall improvement when applying models on modernized hymns
  - non-modern spelling and archaic vocabulary are obstacles

- The models generally improves over time both for original and modernized hymns.
  - The latter suggests that hymns may invoke sentiment differently through time

# LITERATURE

CGR. 2019. Salmer Dataset. Centre for Grundtvig Research (CGR). Github Repository: https://github.com/centre-for-humanities-computing/grundtvig-data

DSLDK. 2021. Salmer Dataset. Society for Danish Language and Literature (DSLDK). Github Repository: https://github.com/dsldk/salmer

Nina Schneidermann and Bolette Pedersen. 2022. Evaluating a New Danish Sentiment Resource: the Danish Sentiment Lexicon, DSL. In Proceedings of the 2nd Workshop on Sentiment Analysis and Linguistic Linked Data, pages 19–24, Marseille, France. European Language Resources Association.

Gustav Aarup Lauridsen, Jacob Aarup Dalsgaard, and Lars Kjartan Bacher Svendsen. 2019. SENTIDA: A New Tool for Sentiment Analysis in Danish.

Journal of Language Works - Sprogvidenskabeligt Studenter-tidsskrift, 4(1):38–53. Number: 1.

Ali Allaith, Kirstine Degn, Alexander Conroy, Bolette Pedersen, Jens Bjerring-Hansen, and Daniel Hersh-covich. 2023. Sentiment Classification of Historical Danish and Norwegian Literary Texts. In Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa), pages 324–334, Tórshavn, Faroe Islands. University of Tartu Library

INSTITUT FOR KOMMUNIKATION OG KULTUR

AARHUS UNIVERSITET

LCHANGE

15. AUGUST 2024

EA LINDHARDT OVERGAARD

PH.D-STUDERENDE

AARHUS UNIVERSITET