

# CANONICAL STATUS AND LITERARY INFLUENCE

## A COMPARATIVE STUDY OF DANISH NOVELS FROM THE MODERN BREAKTHROUGH (1870-1900)

Pascale Feldkamp, Alie Lassche, Jan Kostkan, Márton Kardos,  
Kenneth Enevoldsen, Katrine Baunvig, Kristoffer L. Nielbo



AARHUS  
UNIVERSITY  
CENTER FOR HUMANITIES COMPUTING

November 16



# OUTLINE

---

## Danish Culture Canon

文 17 languages ▾

Article Talk

Read Edit View history Tools ▾

From Wikipedia, the free encyclopedia

The Danish Culture Canon (Danish: *Kulturkanonen*) consists of 108 works of cultural excellence in eight categories: architecture, visual arts, design and crafts, film, literature, music, performing arts, and children's culture. An initiative of Brian Mikkelsen in 2004, it was developed by a series of committees under the auspices of the Danish Ministry of Culture in 2006–2007 as "a collection and presentation of the greatest, most important works of Denmark's cultural heritage." Each category contains 12 works although music

### This talk:

"canonicity" — data: 838 novels — embeddings approach — canon/non-canon similarity in time



AARHUS  
UNIVERSITY  
CENTER FOR HUMANITIES COMPUTING

# IS THE CANON LABEL ARBITRARY?

---

The two polar positions on canon in literary criticism: **arbitrariness** vs **novelty & influence**

**H1:** canonical works are top-down and present day constructs – they do not stand out from their contemporaries textually

**H2:** canonical works are distinguished by novelty and influence – they stand out and show an impact on their literary field



# DATA: THE MODERN BREAKTHROUGH

---

## Why the Modern Breakthrough (1870-1900)?

The full population of novels

We expect literary innovation in this period

We expect a decline of the the popular, historical novel



# DATA: THE MODERN BREAKTHROUGH

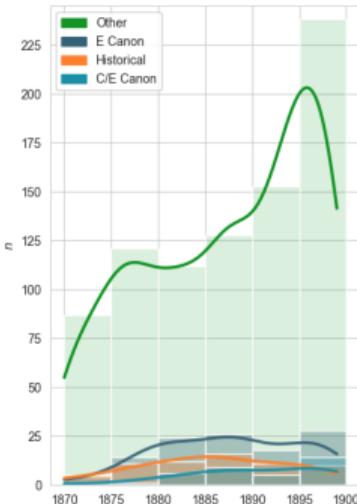
838 novels in Danish (1870-1900)  
tags: historical & canon novels

	titles	authors
Corpus	838	371
Cultural/Educational Canon	36	6
Lexicon Canon	110	19
Extended Canon	114	21
Historical Novels	65 (8)	19 (4)
Other	667	335

-> [mime-memo.github.io/](https://mime-memo.github.io/)

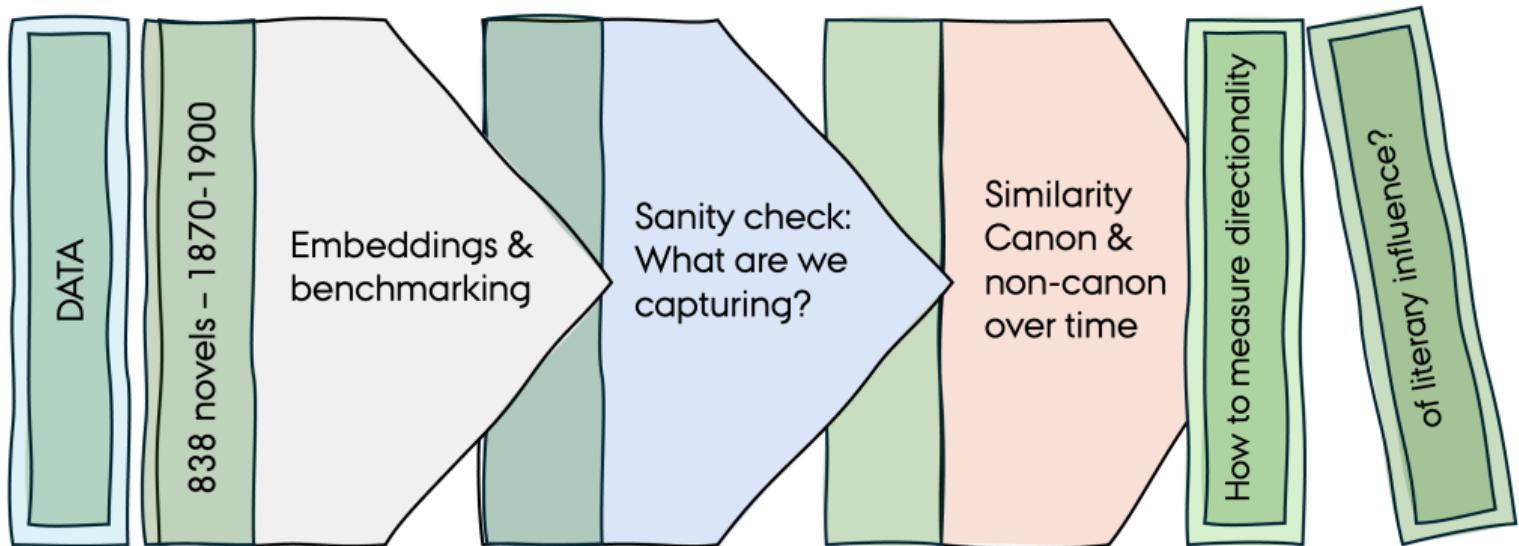


Cultural & Educational  
Canon / Danish Lexicon  
Lemma



# METHODOLOGICAL PIPELINE

---



# EMBEDDING BENCHMARKING

Document embeddings:  $\bar{X}$  slices  
4 models' performance on SEB & historical task

All	Danish	Norwegian	Swedish					
A comparison of sentence embeddings for Danish, Norwegian, and Swedish.								
<input type="button" value="Search in table"/>								
Model	Average Score	Average Rank	Open Source	Embedding Size	WPS (CPU)	Angry Tweets	Berholm Parallel	DKHate
1 multilingual-e5-large-instruct	66.7	4.8	✓	1024	182	64.6	55	67.1
2 text-embedding-3-large	65	6.4	✗	3072		57.8	43.3	70.2
3 embed-multilingual-v3.0	64.1	6.8	✗	1024		58.7	35.6	68.8
4 voyage-multilingual-2	61.8	9.8	✗	1024		55.8	42.5	65

Tasks: thematic clustering, sentiment classification, etc. & custom historical task

Dialect pairing	Dialog systems	Evidence retrieval	Hate speech classification	Language identification	Political classification	Question-answering	Sentiment classification	Thematic clustering
-----------------	----------------	--------------------	----------------------------	-------------------------	--------------------------	--------------------	--------------------------	---------------------

	$\bar{x}$ SEB	Historical	Sol
<b>Number of Datasets →</b>	24	1	25
<b>Models ↓</b>			
m-e5-large-instruct	<b>66.65</b>	<b>40.10</b>	<b>53.38</b>
m-e5-large	<u>60.69</u>	27.66	44.18
DFM-large	55.14	35.13	45.14
MeMo-BERT	36.85	<u>35.38</u>	36.12

Performance of encoder models on SEB tasks and custom historical task.  
Score of Interest (Sol) reflects the model's average score across tasks.



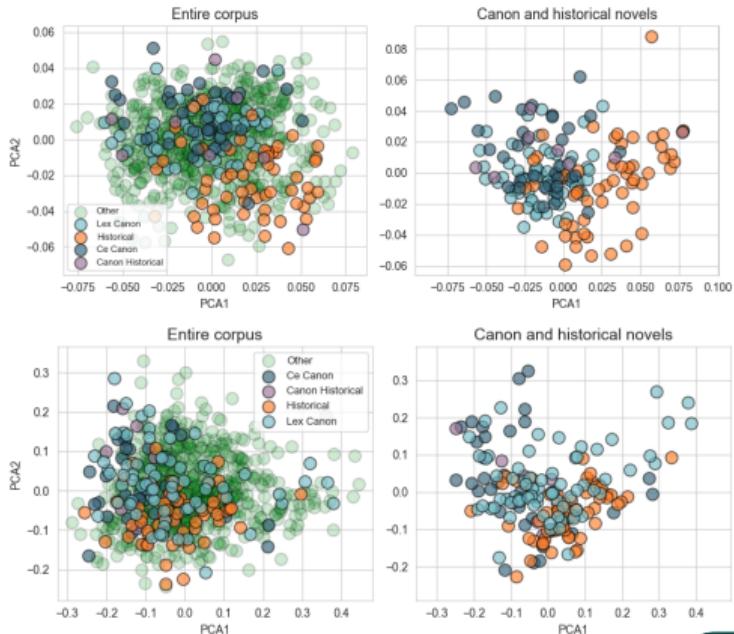
# SANITY CHECK

---

Are representations  
capturing  
semantics/stylistics?

**Can we distinguish  
between historical and  
canon novels?**

*Embeddings (upper)  
vs  
TF-IDF (lower)*



# SIMILARITY: DISTINGUISHING CANON

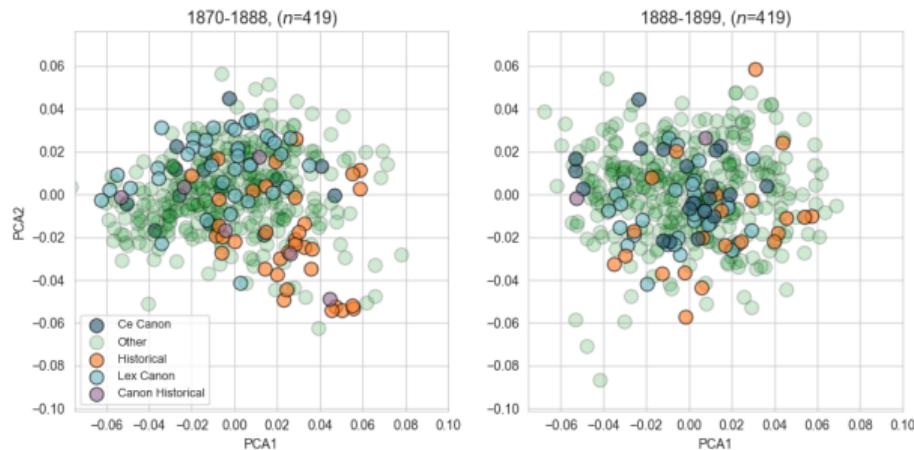
---

Canon stands out  
differently over the  
period

1870-1888 (left)

vs

1888-1899 (right)



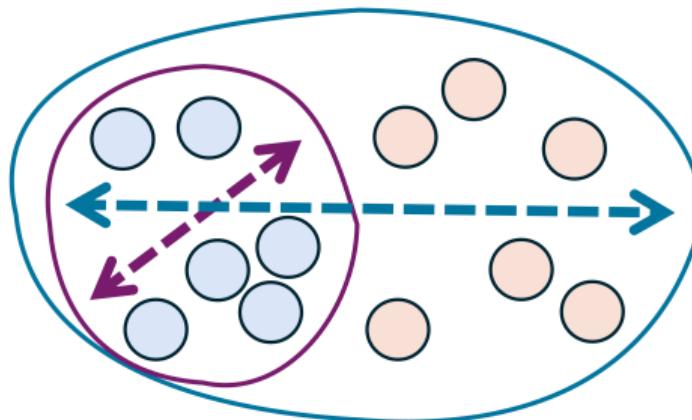
# SIMILARITY: DIACHRONIC CHANGE

---

Intra-group similarity  
(canon/canon)

&

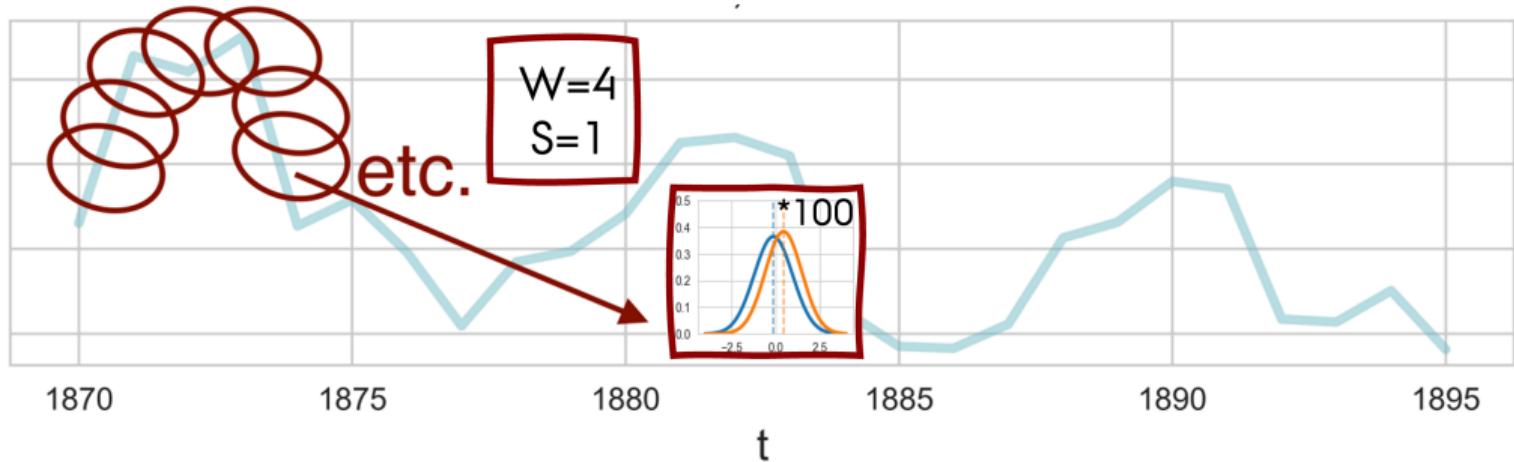
Inter-group similarity  
(canon/non-canon)



intra-/inter-group cosine similarity



# SIMILARITY: DIACHRONIC CHANGE

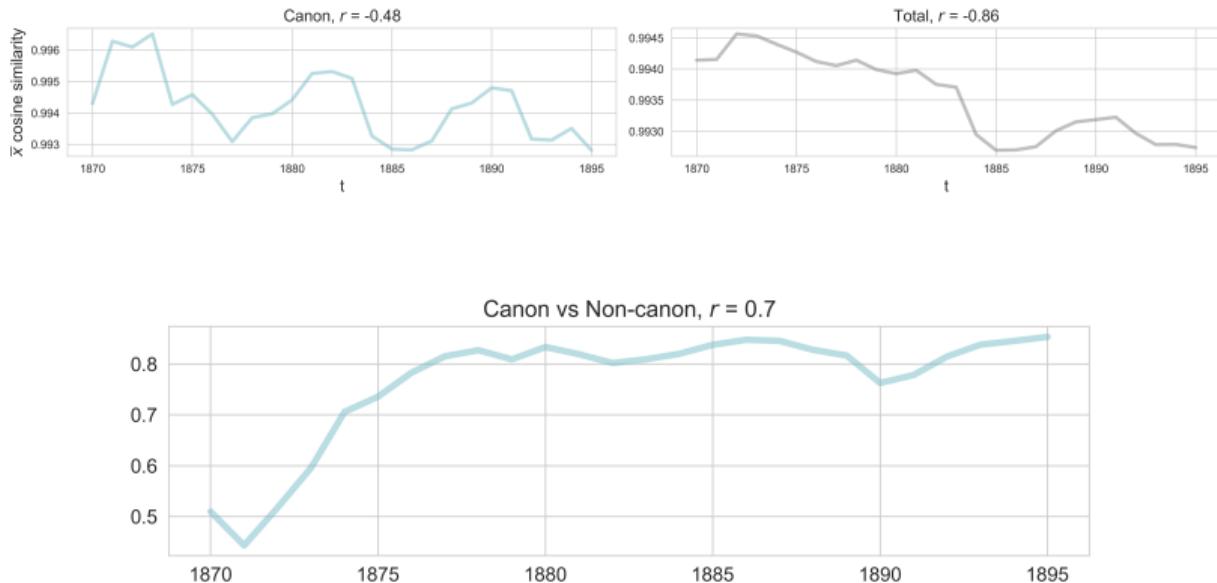


# SIMILARITY: DIACHRONIC CHANGE

---

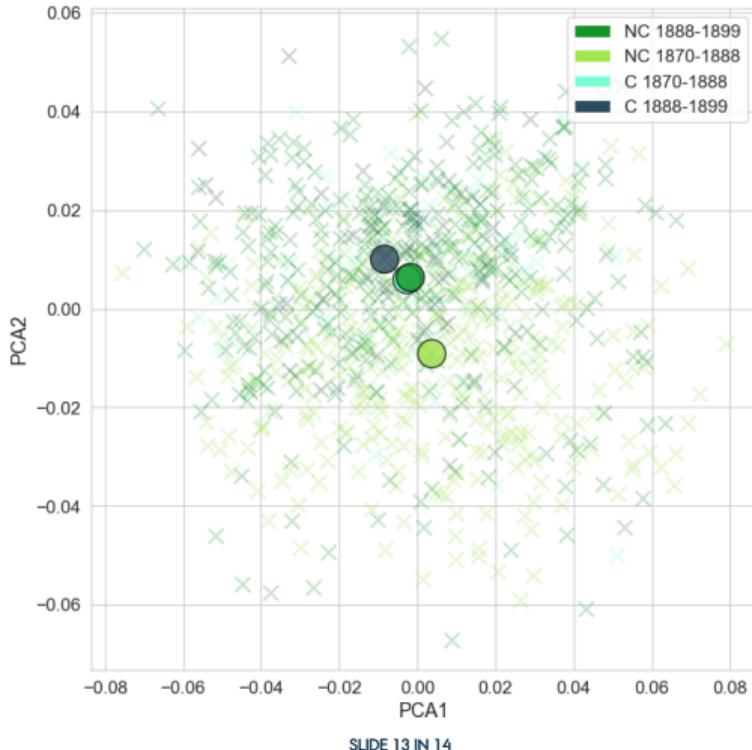
Intra-group  
dissimilarity

Inter-group  
similarity



# FUTURE WORK: DIRECTIONALITY

---



**THANK YOU**  
try: questions?

[pascale.moreira@cc.au.dk](mailto:pascale.moreira@cc.au.dk)

[chc.au.dk](http://chc.au.dk)

**SLIDES**

[pfeldkamp.github.io/pascalefeldkamp/pdfs/canon.pdf](http://pfeldkamp.github.io/pascalefeldkamp/pdfs/canon.pdf)



Tagged corpus



Code

This research is part of the *The Golden Array of Danish Cultural Heritage* and *Golden Imprints of Danish Cultural Heritage* projects supported by the Carlsberg Foundation and Aarhus University Research Fund.



AARHUS  
UNIVERSITY  
CENTER FOR HUMANITIES COMPUTING