

THE DYNAMICS OF LITERARY FIELDS

TRACING INFLUENCE OF DANISH CANONICAL NOVELS (1870-1900)



LITERARY DYNAMICS IN TIME

The **Golden Matrix** project

Ambition: dynamics in the literary field of the Danish Golden Age (1750-1930)

Limitations: the “great unread”, national focus, selectivity of features/themes

Current approach: comprehensive snapshots

Tracing the canon through 30 years – based on the paper:

“Canonical Status and Literary Influence: A Comparative Study of Danish Novels from the

Modern Breakthrough (1870–1900)”, NLP4DH workshop at EMNLP 2024.

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



DANISH CANON

- What is “canon”?

Danish Culture Canon

文 A 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



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)?

We expect literary innovation in this period

We expect a decline of popular genres (e.g., historical novel)

! The full population of novels – a comprehensive snapshot of the literary field



DATA: THE MODERN BREAKTHROUGH

838 novels in Danish (1870-1900)

filename string · lengths	full_firstname string · lengths	auth_first string · lengths	auth_last_modern string · lengths	pseudonym string · classes
1893_Hamsun_RedaktoezLynge	2	Knut	2	Hamsun
1890_Hamsun_Sult	Knut	Knut	Hamsun	null
1876_JacobsenJP_FruMarieGrubbe	Jens Peter	J.P.	Jacobsen	null
1880_JacobsenJP_NielsLyhne	Jens Peter	J.P.	Jacobsen	null
1896_JensenJV_Danskeze	Johannes V.	Johannes V.	Jensen	null
1898_JensenJV_EinarElkaer	Johannes V.	Johannes V.	Jensen	null
1888_JuelHansen_EnUngDamesHistorie	Erna	Erna	Juel-Hansen	null
1895_JuelHansen_KaerlighedensVeje	Erna	Erna	Juel-Hansen	null
1881_JuelHansen_Melleml20g17	Erna	Erna	Juel-Hansen	Azne Wendt

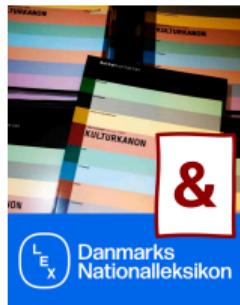


DATA: THE MODERN BREAKTHROUGH

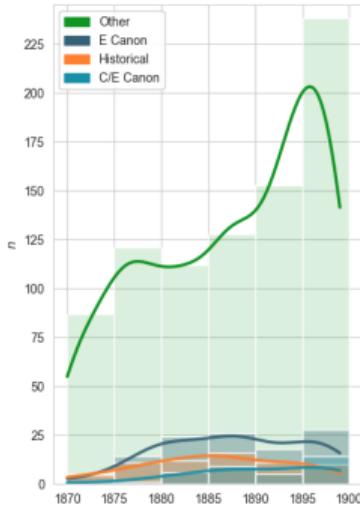
tags: historical & canon novels

	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

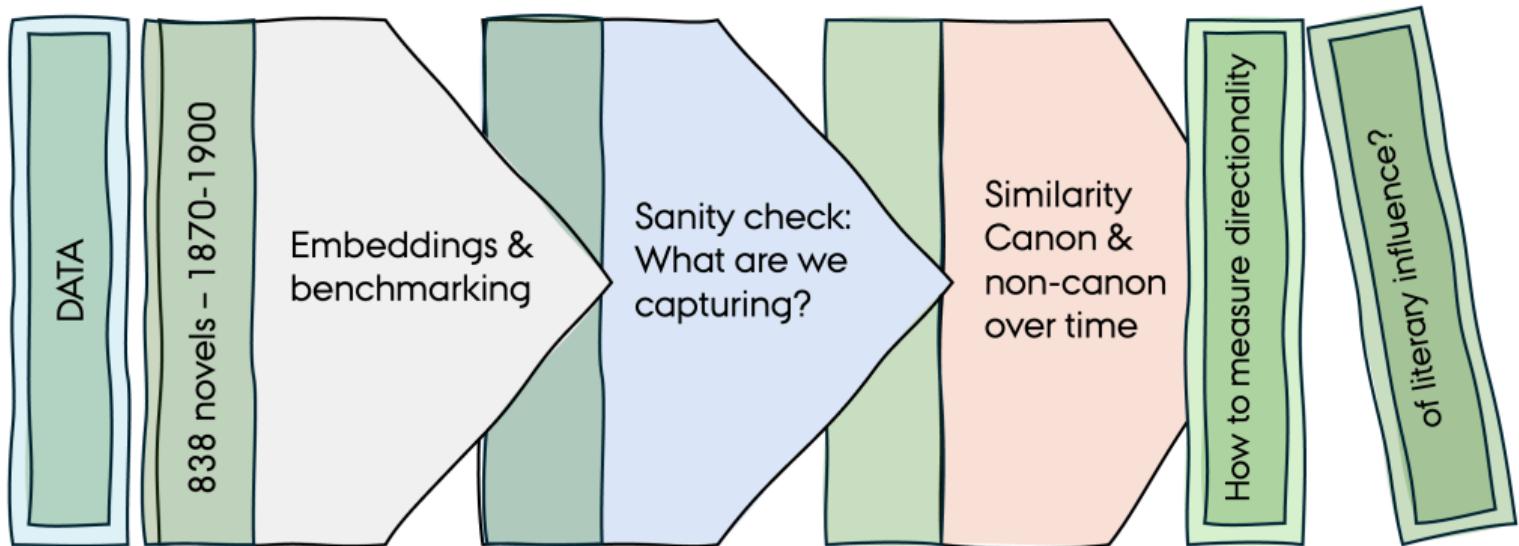
-> mime-memo.github.io/



Cultural & Educational
Canon / Danish Lexicon
Lemma



METHODOLOGICAL PIPELINE



EMBEDDING BENCHMARKING

Document embeddings: \overline{X} slices
4 models' performance on SEB

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	Bornholm Parallel	DKHate
1 multilingual-e5-large-instruct	65.7	4.8	✓	1024	182	64.6	55	67.1
2 text-embedding-ge4	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

& their performance on a **custom historical task**
(clustering chunks of novels according to author)

Result:

	\overline{x} SEB	Historical	Sol
Number of Datasets → Models ↓	24	1	25
m-e5-large-instruct	66.65	40.10	53.38
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.



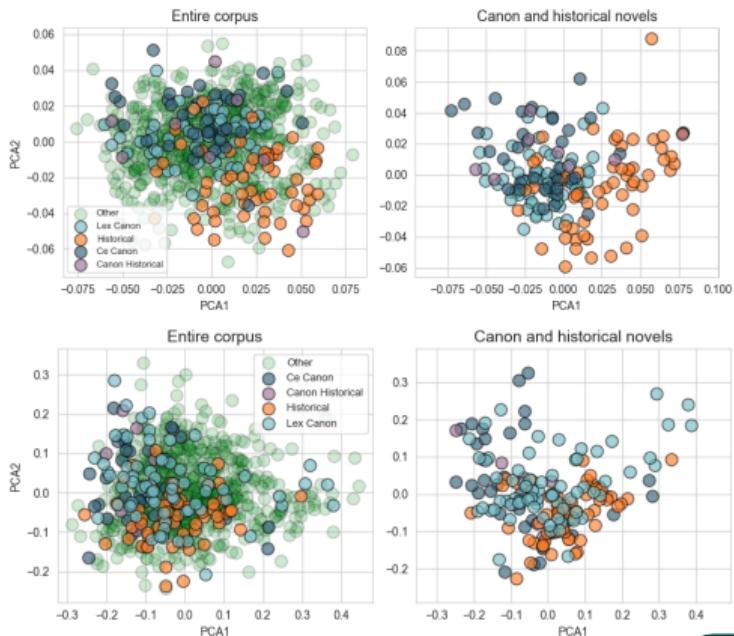
SANITY CHECK

Can we distinguish
between historical and
canon novels?

Embeddings (upper)

vs

TF-IDF (lower)

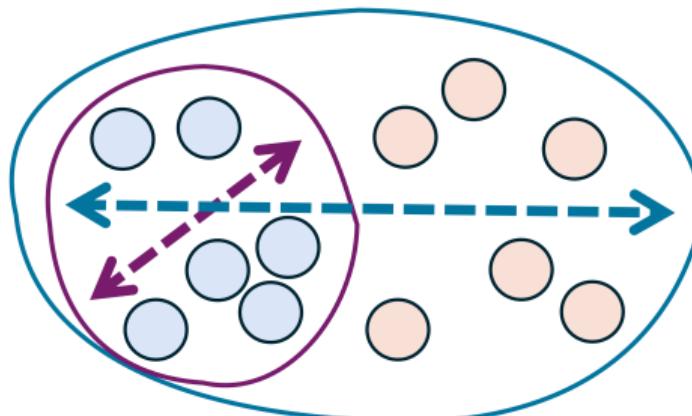


DIACHRONIC CHANGE

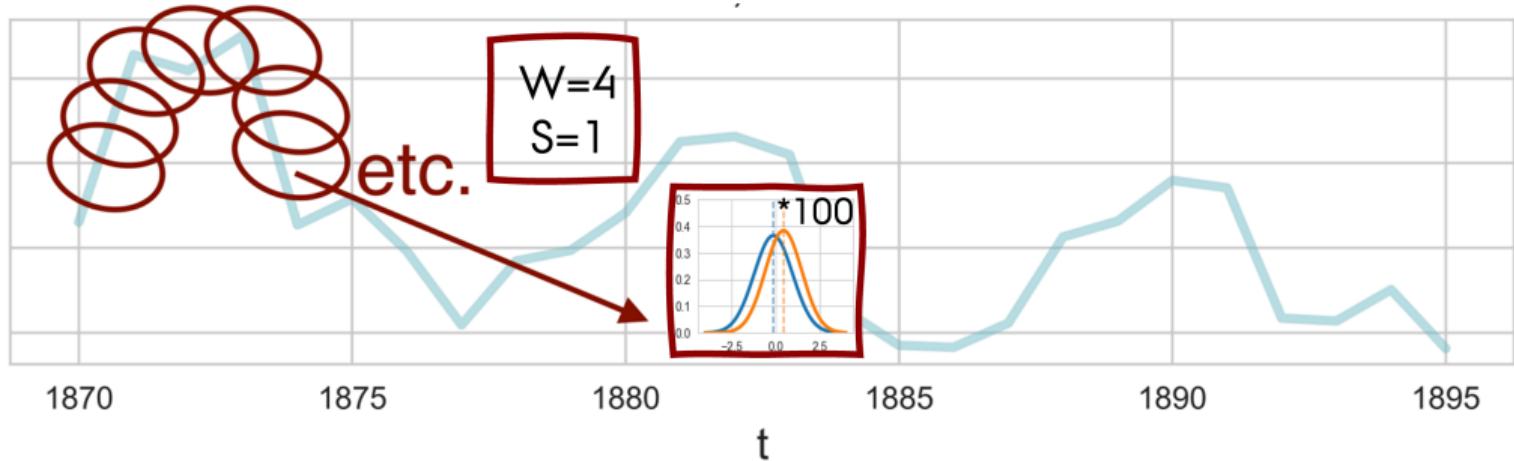
Intra-group similarity
(canon/canon)

&

Inter-group similarity
(canon/non-canon)



DIACHRONIC CHANGE

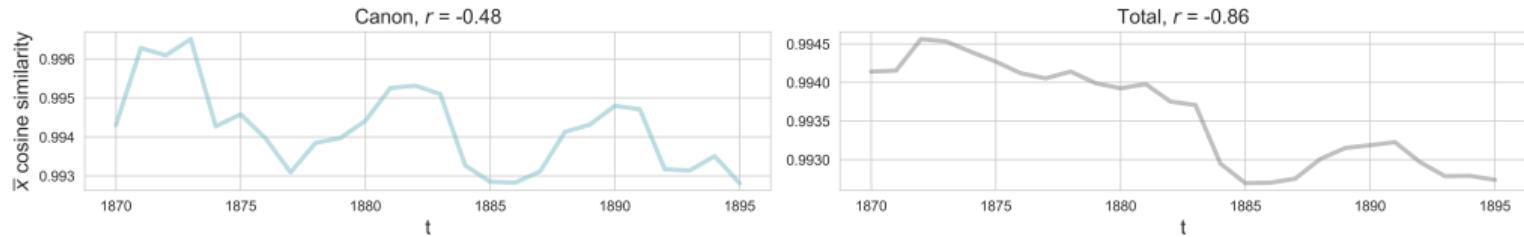


NB: similar results when not upsampling & when using bootstrapping – see paper



DIACHRONIC CHANGE

Intra-group similarity in the canon/full corpus across the period

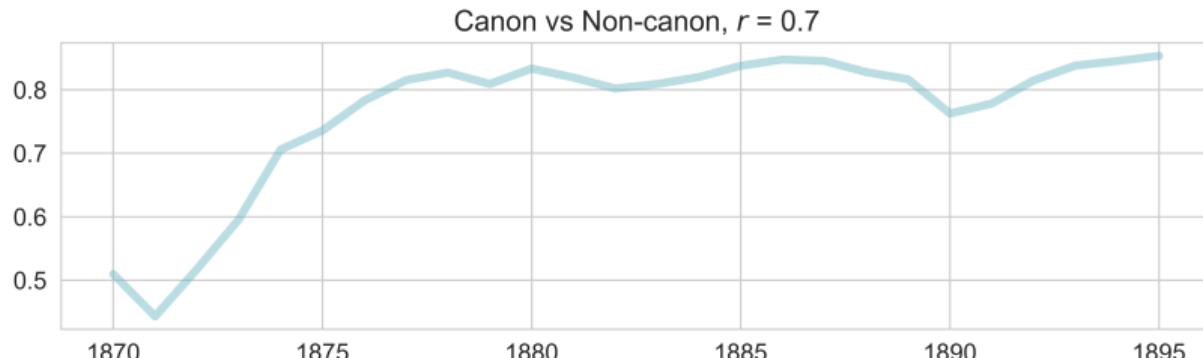


Spearman's ρ of the correlation between year and mean intragroup similarity. Both canon and full corpus tend to get **more internally diverse** across time, though canon books less so and less consistently.



DIACHRONIC CHANGE

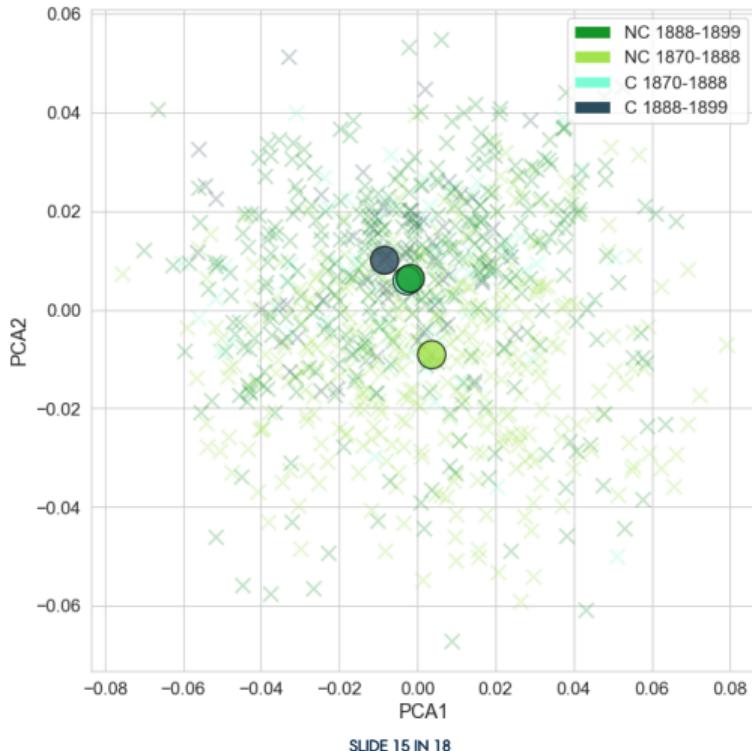
Inter-group (canon vs non-canon) similarity across the period



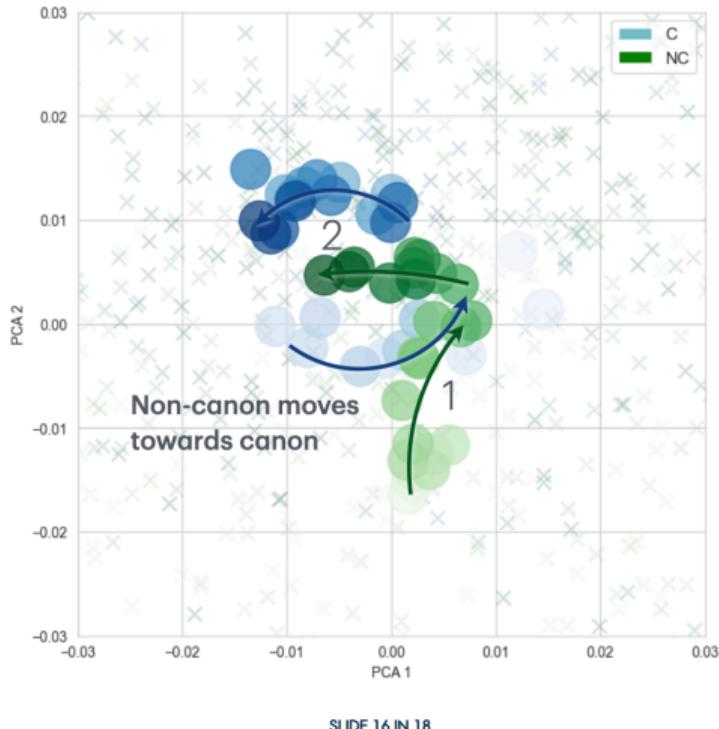
Note that a correlation with time persists after 1875 ($\rho = 0.42$).



FUTURE WORK: DIRECTION OF CHANGE



FUTURE WORK: DIRECTION OF CHANGE

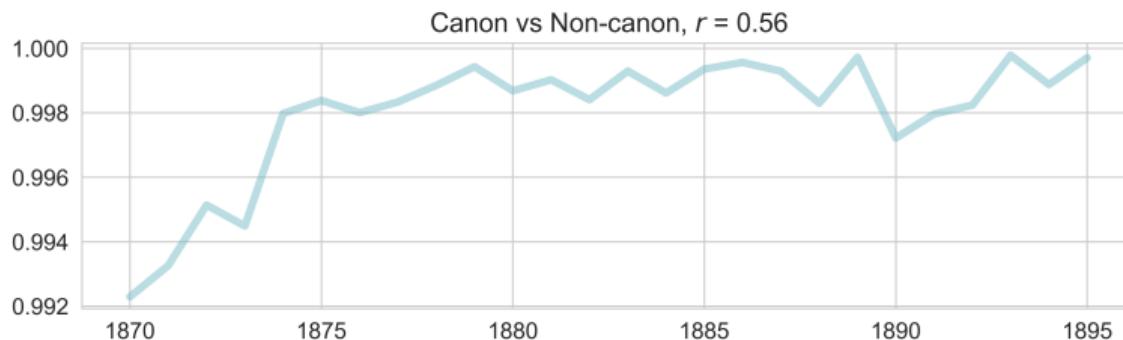


FUTURE WORK: MORE INTERPRETABILITY

While we can detect change here, how to characterize it? – the problem of opaque embeddings

Vector of stylistic measures:

'average word length' – 'msttr' – 'average sentence length' – 'bzip ratio' – 'bigram entropy' – 'word entropy' – 'coherence' – 'readability (flesch ease)'



THANK YOU
questions?

pascale.moreira@cc.au.dk

chc.au.dk



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