

# The Dostoyevskian Trope

State Incongruence in Danish Textual Cultural Heritage (DaTeCH)

*Culture Analytics Reunion @Lake Arrowhead|UCLA*

Kristoffer L Nielbo

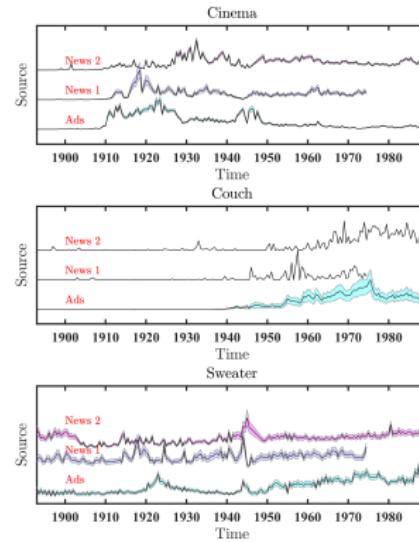
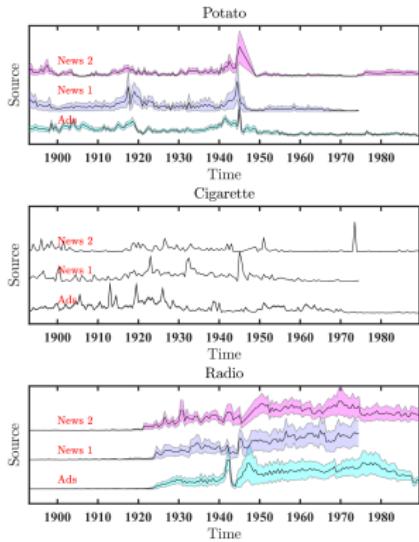
[knielbo@sdu.dk](mailto:knielbo@sdu.dk)

[knielbo.github.io](https://knielbo.github.io)

**datacube**|Department of History|University of Southern Denmark

## **PROGRAM**

<b>0.01</b>	<b>INTRODUCTION</b>	Kehre, <b>The Dostoyevskian Trope</b> , Geertz V
<b>0.10</b>	<b>METHODS</b>	IT, AFA, LDA
<b>0.15</b>	<b>RESULTS</b>	suggestions
<b>0.20</b>	<b>DISCUSSION</b>	...



"They say I'm old-fashioned, and live in the past, but sometimes I think progress progresses too fast" (Dr. Seuss)

## RESEARCH PROBLEMS in TEXTUAL CULTURAL HERITAGE

---

### DIE KEHRE

- “writers of fiction (and non-fiction) undergo personal paradigm shifts”

### DOSTOYEVSKIAN TROPE

- “writers’ creative state is inversely related to their emotional state”
- “writers’ creative state *depends* on their emotional state” (state incongruent writers)

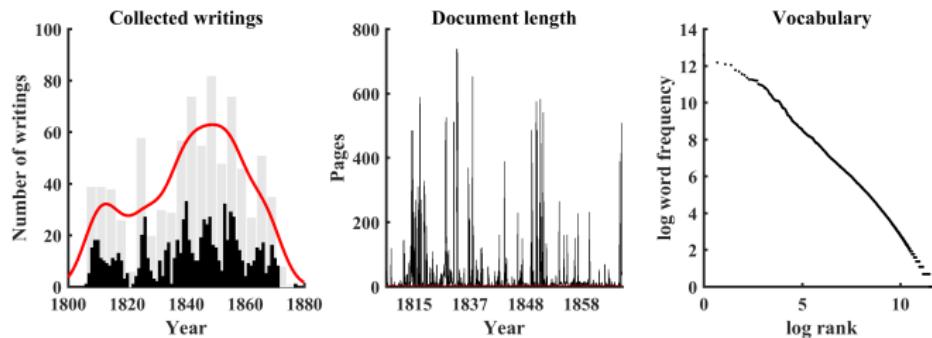
### GEERTZ V-FUNCTION

- particular case where the *kehre* is reflected in a topicality shift for state congruent writers, which might characterize successful innovators

### DaTeCH sandbox

---

- N.F.S. Grundtvig, active years 1804-1871 (born: 1783, dead: 1872)
- H.C. Andersen, active years 1829-1874 (born: 1805, dead: 1875)
- S.A. Kierkegaard, active years 1830-1855 (born: 1813, dead: 1855)



## DATA|danish textual cultural heritage (DaTeCH)

- $N = 1329$  documents ( $n_{NSFG} = 921$ ,  $n_{HCA} = 194^*$ ,  $n_{SAK} = 214$ )
- language normalization (orthographic variation and casefolding)
- length normalization ( $slen = 10^2 - 10^3$  words)

## **historical text normalization**

---

### **orthographic variation**

- automated analysis depends critically on existing tools and data (ex. sentiment dictionaries)
- NLP and TM resources “suffer from presentism”
- project often try to adapt the tool (ex. modify dictionary to historical data set)
- this solution scales badly due to lack of standardization

### **solution**

- statistical spelling corrector that treats historical variants as errors of contemporary Danish
- rule-based improvement

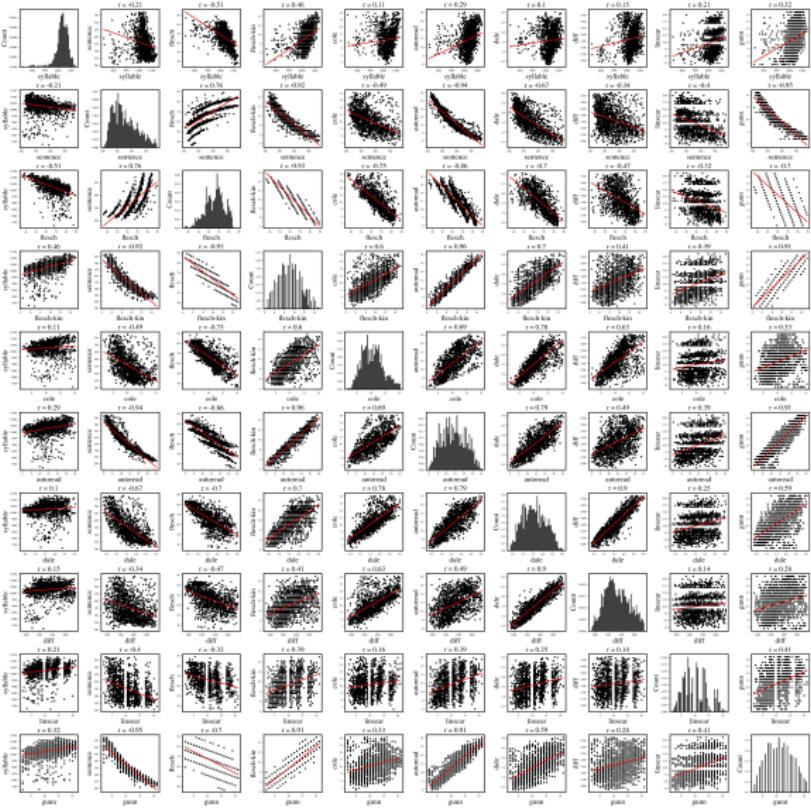
## LEXICAL VARIABILITY ~ CREATIVITY #1

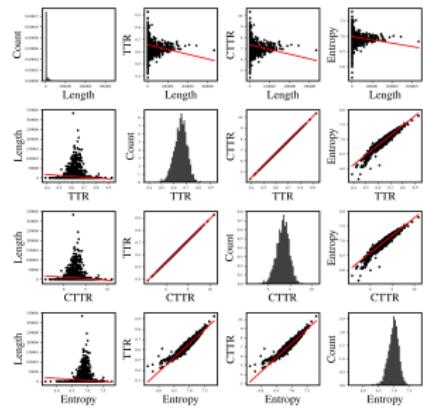
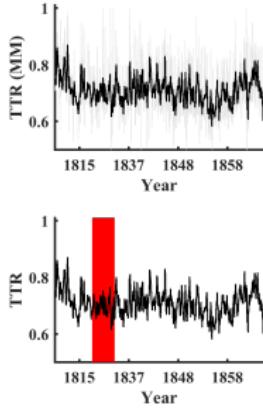
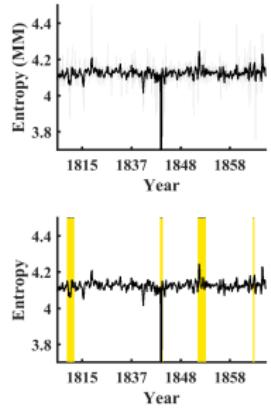
- multiple measures for lexical density (*TTR*, *CTTR*) and readability (*Flesch-Kincaid*, *SMOG*, *LIX*) that captures aspects of “text complexity” in terms of syllables, word and sentence length ~ character- and word-level characters ngrams
- issues pertaining to normalization of length, algorithms scale badly + domain dependency

$$H = - \sum_{i=1}^n p_i \times \log_2(p_i) \quad (1)$$

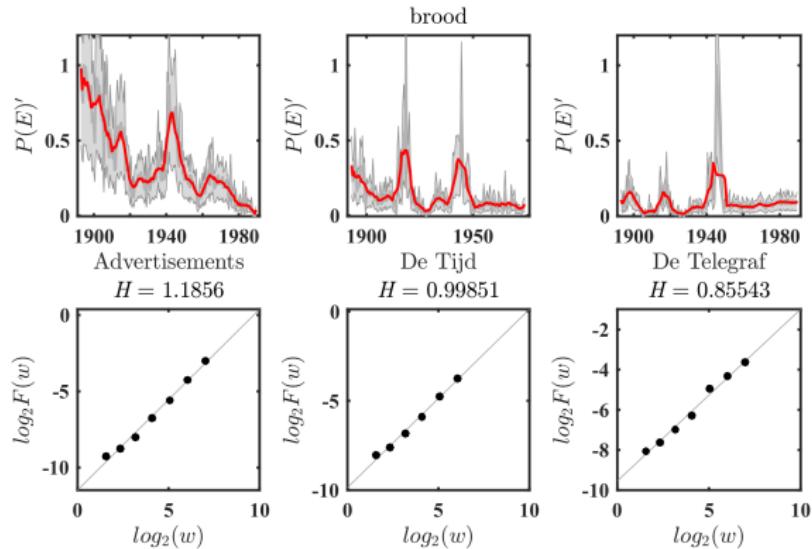
$$p_i = \frac{Fr(w_i)}{\sum_i^n Fr(w_i)} \quad (2)$$

- IT offers a range of information measures that capture lexical variability, that is, we can think of  $H[X]$  as the variability of some term-vectors  $X \sim \log$  of the effective number of values it can take
- $H[X, Y]$  then is the variability is associated with the pair term-vectors and  $H[Y | X]$  the variability that remains when  $X$  is known





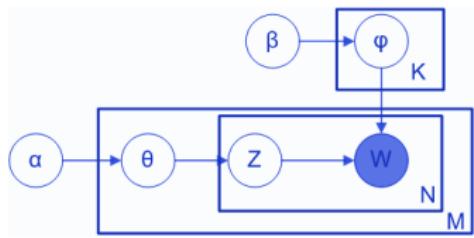
## LONG-RANGE DEPENDENCIES ~ CREATIVITY #2



### LONG RANGE DEPENDENCIES

- creativity is “flow” ~ persistent correlations in lexical variability
- Adaptive Fractal Analysis to estimate Hurst exponent: *antipersistent correlations*:  $0 < H < \frac{1}{2}$ , *memoryless*:  $H = \frac{1}{2}$ , *persistent correlations*:  $\frac{1}{2} < H < 1$

## LDA DIVERGENCE $\sim$ semantic innovation

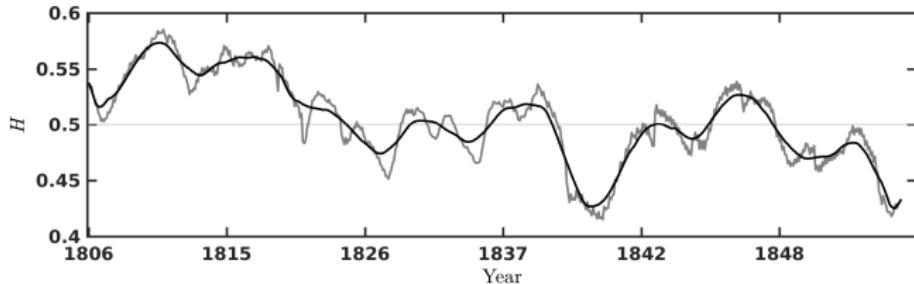


$\alpha$	Dirichlet prior for per-doc topic dist - proportions parameter
$\beta$	Dirichlet prior for per-topic word dist - topic parameter
$\theta_i$	word dist for topic - per-document topic proportions
$\phi_k$	word dist for topic $k$ - topics
$Z_{ij}$	topic for $j^{th}$ word in doc $i$ - per-word topic assignment
$W_{ij}$	the observed word

- model semantic innovation as “variation on a theme”  $\Rightarrow$  use a simple Bayesian model to capture lexical semantics
- model each document as a distribution on lexical topics (e.g.,  $P_1 = [0 .09 .78 .11 .2]$ ), where each ‘topic’ is a distribution on words, and compare document similarity

$$D_{KL}(P \parallel Q) = \sum_{i=1}^n P(i) \times \log_2 \frac{P(i)}{Q(i)} \quad (3)$$

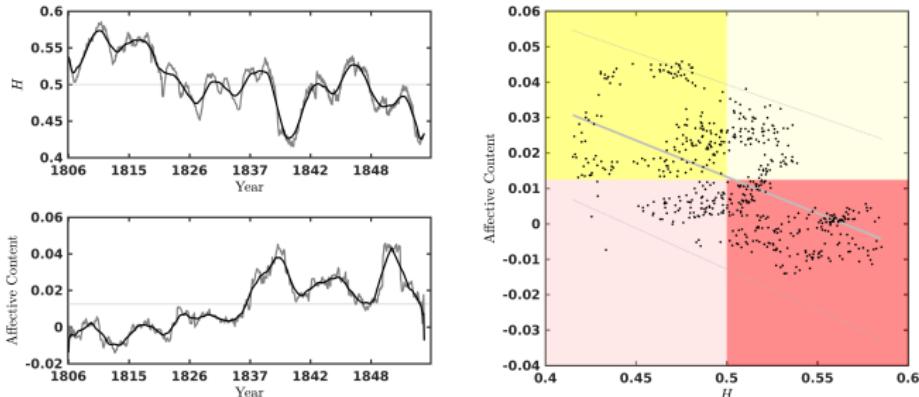
- bracket concrete semantics and only compare relative entropy between documents on topics (“variation on a theme”)



---

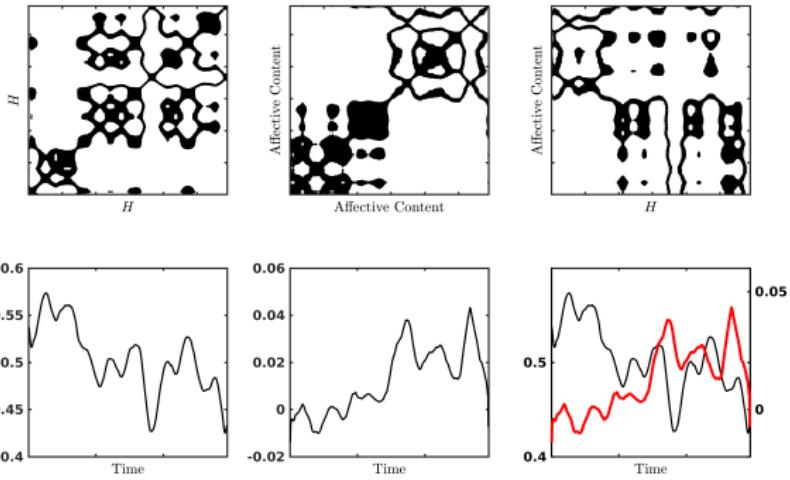
### CREATIVITY|Grundtvig

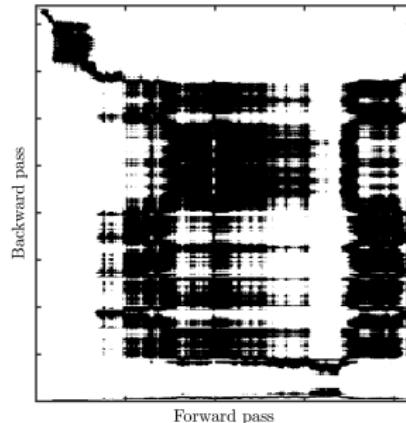
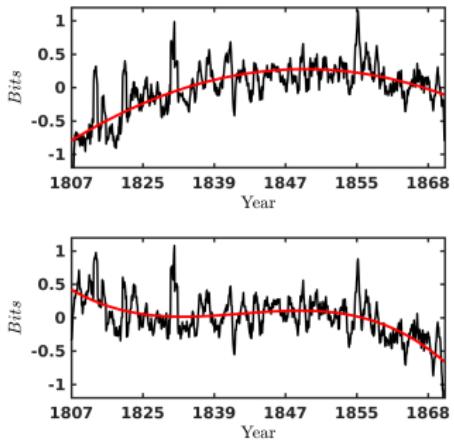
- early phase: persistent behavior ("flow in lexical variability")
- late phase: mixture of persistent and anti-persistent behavior
- creative kehre in early forties



## EMOTION|Grundtvig

- early phase: negative affective tone
- late phase: positive affective tone
- inverse relation → state incongruent writer
- emotional state Granger-causes creative state → dostoyevskian trope





## INNOVATION|Grundtvig

- early phase: fast increase in semantic innovation
- late phase: (extended) flat to decreasing development ~ uninventive

## COMPARISON|artist typology

### Grundtvig

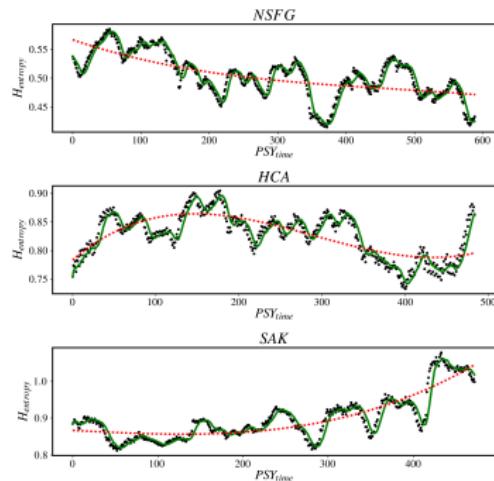
- “Kehre” – age-dependent change point (30-40 yrs)

### Andersen

- “eternal child”
- optimal creativity throughout life

### Kierkegaard

- extended “27 Club”
- production terminated before a Kehre



## COMPARISON|artist typology

### Grundtvig

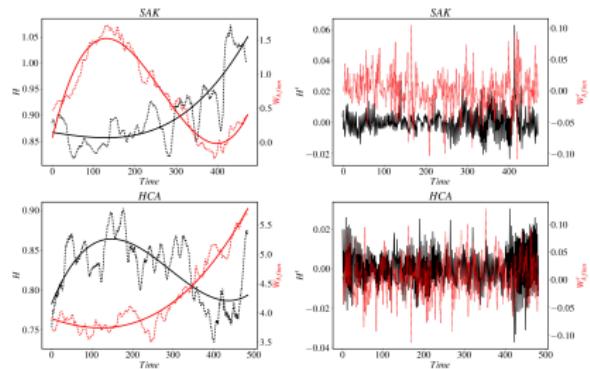
- state incongruence
- dostoyevskian trope

### Andersen

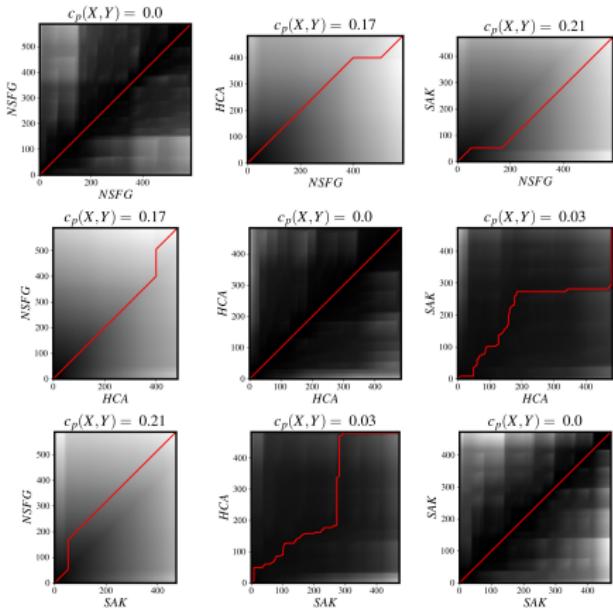
- state incongruence
- but *not* dostoyevskian trope

### Kierkegaard

- state incongruence
- dostoyevskian trope



## COMPARISON|sequence alignment



## CONCLUSION

---

### KEHRE

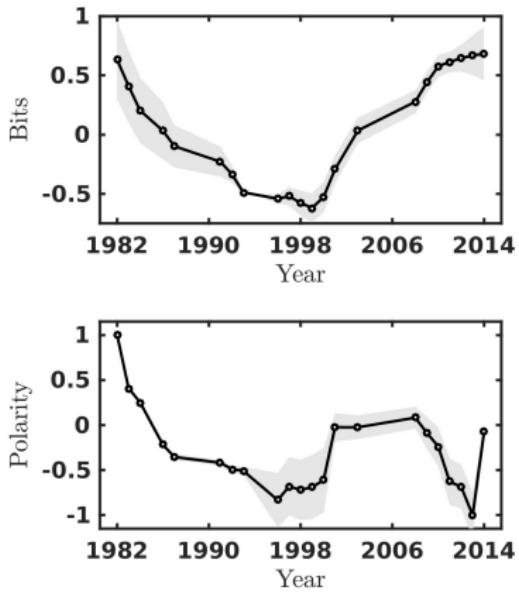
- “No mathematician should ever allow himself to forget that mathematics, more than any other art or science, is a young man’s game.” (G.H. Hardy, *A Mathematician’s Apology*, 1940)
- 30-40 yrs
- change in scaling dynamics for creativity that separates early and late writer: NFSG, SAK\*

### DOSTOYEVSKIAN TROPE

- state incongruent writer (creativity::affect): NFSG, HCA, SAK
- writes on affective state (affect→creativity): NFSG, SAK

### GEERTZ V-FUNCTION

...



## **THANK YOU**

[knielbo@sdu.dk](mailto:knielbo@sdu.dk)  
[knielbo.github.io](https://knielbo.github.io)

**& thanks to**  
bin liu & jianbo gao  
katrine f. baunvig

**data**

[adl.dk](http://adl.dk), [grundtvigsværker.dk](http://grundtvigsværker.dk), [sks.dk](http://sks.dk)

**funding**

danish agency for science and innovation & (partially) mellon foundation