A Curious Case of Entropic Decay: Persistent Complexity in Textual Cultural Heritage

Kristoffer L. Nielbo[1,2], Katrine F. Baunvig[1], Bin Liu[3] and Jianbo Gao[4,5,6,7]

[1]Department of History, University of Southern Denmark, Odense, Denmark
[2]SDU eScience Center, University of Southern Denmark, Odense, Denmark
[3]Business School, Guangxi University, Nanning, Guangxi 530005, China
[4]State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
[5]Center for Geodata and Analysis, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[6]Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[7]Institute of Complexity Science and Big Data Technology, Guangxi University, Nanning, China

Abstract

To understand an author's developmental trajectory, the static traits and properties of author reconstruction and profiling are not sufficient. Instead, it is necessary to focus on high level indicators of the complex set of variables that underlie the author's transient mental states during his or her creative production. We propose a method that combines information theory with random fractal theory in order to study the mental dynamics of an author as indicated by text complexity. To illustrate its application, we analyze the developmental trajectory of the culturally influential and "graphomanic" 19th century Danish pastor N.F.S. Grundtvig. This approach can detect an age-related trend (entropic decay), a significant Kehre (turning point), and multiple event-related change points in his production. We argue that the approach is applicable beyond the specific case and can be extended to comparative analysis within and between authors, and, finally, to dynamic analysis of cultural information systems.

Keywords— author profiling, entropy, fractal analysis, Hurst exponent, text analysis

Introduction

Author reconstruction and profiling can broadly be construed as the use of textual features to identify more or less static traits (e.g., openness, extraversion) and properties (e.g., gender, age) of an author (Stamatatos 2009). The static nature of such traits and properties can however blur the fact that textual features reflect an inherently dynamic process of an author's transient mental states during his or her creative production and that these states vary as a function of a complex set of variables. To understand the developmental trajectory of a prolific author, it is therefore necessary to track an indicator of these complex mental dynamics. We propose a method that combines information theory and random fractal theory in order to partially reconstruct the mental dynamics that underlie an author's writings.

Multiple research fields are showing a renewed interest in case studies of the mental dynamics that underlie textual sources. Historians and literary scholars explore the impact of writers' biographies on writing style and content (Renders, Haan, and Harmsma 2016; Norris 2016). Philosophers and cognitive scientists model the knowledge trajectories of intellectual ancestors (Murdock, Allen, and DeDeo 2015), and computer linguists and psychologists detect author-specific patterns of mental traits and impairments in the writings of culturally significant persons (Berisha, Wang, LaCross, and Liss 2015; Le, Lancashire, Hirst, and Jokel 2011). In part, this "biographical trend" is geared by the rapid development of techniques for automated text analysis combined with massive digitization of textual cultural heritage. Digital collections of prolific authors allow for an algorithmic reconstruction of past minds (Nielbo, Nichols, and Slingerland 2017), or, more exactly, the mental states reflected in the authors' text production.

As evidenced by the plurality of metrics and methods, reconstruction of mental states from texts is not a trivial task. Mapping linguistic units, such as words or sentences, onto mental states is complicated by the fact that a text can have many layers and is not typically a direct representation of the author's mental condition. Lexical matching based sentiment dictionaries, for instance, might be adequate for detecting tone and preferences of social media and consumer reviews, but how does the average sentiment score of a paragraph narrated by an unreliable fictitious narrator map onto an author's mental state? Even in the case of autobiographies, the author is neither bound by the truth nor does she have unfiltered access to her mind's information processing. A second set of problems is related to the displacement between production and publication. While it is possible to sort texts temporally based on their publication dates, it is complicated to do so in terms of their production date. Some texts might have taken years to produce, while others are produced in a matter of months or even days, because ideas and representations develop at multiple time scales (Donald 2007; Kahneman 2011). At shorter time scales, randomness in the sequential order of texts is therefore expected and for states to characterize an author, it is necessary to detect dynamics at longer time scales.

To adequately capture a prolific author's mental states, it is therefore essential to: a) identify a metric that summarizes mentally relevant linguistic features, which the writer did not exert direct control over during writing; and b) analyze variation in this metric with a method that is sensitive to dynamics at multiple time scales with a particular focus on long time scales. To do this we resort to Shannon's source entropy or just entropy as a measure of text complexity, and combine it with an index of persistence from random fractal theory in order to capture the dynamics of mental states reflected in a specific case study of the collected writings of Danish pastor N.F.S. Grundtvig. The approach (i.e. the combination of entropy and random fractal theory) is applicable beyond the specific case and therefore of general relevance to text-based studies of dynamics.

While entropy and information theory are widely used for studying a) variability in discrete time series such as characters and words (Shannon 1948; Thoiron 1986; Zhang 2016) and b) dynamics of creativity and innovation (Barron, Huang, Spang, and DeDeo 2017; Bilder and Knudsen 2014; Gabora 2016; McGavin 1997; Murdock, Allen, and DeDeo 2015)(see appendix A, equation 1-2), fractal analysis is less common in the social sciences and humanities. Methods for fractal analysis are used in many areas of science to study self-similarity in dynamic systems (Liebovitch and Shehadeh 2003). Dynamic systems are systems that change over time or, more exactly, systems whose behavior can be described as a function of time. Dynamic systems exist in history and culture, at every level of abstraction, from the individual reading process or plot of a novel to the complex interplay of agents and objects that we refer to as societies and historical periods. Some dynamic systems exhibit self-similarity such that patterns of fluctuations at shorter time scales are scaled copies of fluctuations at longer time scales. In figure 1 self-similarity is pronounced in the persistent random walk (h), where rounded mountain-like structures can be observed across multiple time-scales. An example of self-similarity can be found in reading, where self-similar properties are found at multiple time-scales, because reading fluency and word comprehension are affected by the word, its immediate context (i.e. shorter time scales) as well as the larger text context of paragraphs and chapters (i.e., longer time scales) (O'Brien, Wallot, Haussmann, and Kloos 2014). At a higher level of abstraction, classes of cultural representations show differential self-similarity. Collective representations related to natural events (e.g., earthquakes, lightning, fires) have self-similar properties because they re-occur at regular intervals of varying lengths (Gao, Hu, Mao, and Perc 2012). Light earthquakes, for instance, occur every year, but great earthquakes only occur a couple of times each decade. In contrast, representations of social events do not exhibit self-similarity because their occurrence depends on a complex interplay of human agents, which seem to result in a more chaotic profile.

Many culturally relevant complex systems display self-similarity in psychology (Chater and Brown 1999), economy (Marchant 2008), sociology (Gao, Fang, and Liu 2017), health (Eke, Herman, Kocsis, and Kozak 2002), language (Gao, Hu, Mao, and Perc 2012) and music (Voss and Clarke 1975). In all these domains we find an important class of fractal objects called $1/f^{\alpha}$, which has attracted considerable attention due to findings of so-called "pink noise", where $\alpha = 1$, in a range of natural and man-made processes. The presence of $1/f^{\alpha}$ noise seems to be linked to self-organized criticality (SOC) in dynamic systems, which is fundamental for complex self-organizing processes (Bak, Tang, and Wiesenfeld 1987). Psychologists have argued that SOC represents a unifying principle in psychological systems that couple with a self-similar physical world (Chater and Brown 1999). In the present study we are targeting a subclass of $1/f^{\alpha}$ noise called a $1/f^{2H+1}$ process (see appendix A, equation 3). For $1/f^{2H+1}$ processes, we estimate the Hurst exponent which is an index of self-similarity and, by extension, the complex organization of the process. The Hurst exponent takes values of $0 < H < 1$ and measures the persistence of a time series, that is, the degree to which a pattern of fluctuation persists at multiple time-scales. The following heuristic can be used to identify different classes of or states in time series (Gao, Hu, and Tung 2011): For $0 < H < 0.5$, the time series is an anti-persistent process (i.e., increments are followed by decreases and decreases by increments); for $H = 0.5$, the time series only has short-range correlations also called short memory; and when $0.5 < H < 1$, the time series is a persistent process (i.e. increments are followed by increases and decreases by further decreases) that is characterized by long-memory[1] (Fig. 1). Returning to the previous examples of self-similarity, we can say that the reading and representations of natural events are persistent processes because it has been shown that for reading speed and word frequencies over time respectively $H > 0.5$.

The rationale behind our proposed combination of entropy and fractal analysis is as follows: First, the data consists of the collected writings of one prolific author, N.F.S. Grundtvig (Grundtvig henceforth) (see appendix A, Data). Entropy is estimated for each work and used as a direct measure of its lexical variability (i.e. text complexity) and an indirect measure of the complexity of Grundtvig's mental states during its production (see appendix A, Entropy). This measure is positively associated with the creativity and innovation of the text such that higher levels entropy measure higher complexity (see Related Work), which is indicative of changes in Grundtvig's mental states. Importantly, by itself the measure cannot distinguish between the different status or content of the mental states - whether Grundtvig was unbalanced, delirious or manic when writing. Second, Fractal analysis is applied to the time-ordered work entropies in windows of size n in order to estimate a time-varying Hurst exponent (see appendix A, Fractal Analysis). The time-varying Hurst exponent is then interpreted as a proxy for the persistence or continuity of the complexity of mental states during Grundtvig's productive career. We are particularly interested in windows where $0.5 < H < 1$ as expressions of creativity and innovation, because that indicates a self-organizing process underlying a sequence of productions or what might more popularly be called "creative flow" between the works.

Case Study

Nikolai Frederik Severin Grundtvig (1783-1872) was a "graphomanic" pastor, pedagogue, politician and polyglot; his language, individual texts and thoughts have had considerable impact on Danish national identity (Hall, Korsgaard, and Pedersen 2015). He was hyperactive as a poet, as a scholar, and as a polemic columnist. Today he is considered one of the most important characters of 19^{th} century Danish intellectual activity – more important even than his contemporaries, Søren A. Kierkegaard (1813-1855) and H.C. Andersen (1805-1875). Influenced by the Romantic Movement (Scharling 1947; Lundgreen-Nielsen 1980), Grundtvig is perceived to be a central agent in the construction of a modern Danish cultural identity focused on encompassing every social stratum in the socially revolutionary 19^{th} century (Nygaard 2012). It is widely acknowledged that Grundtvig worked for and, at least to some extent, succeeded in constructing a sense of community within the Danish people - the "Folk". The goal was to create a collective consciousness incorporating the social heterogeneous "Folk" as well as cultural heroes from the days of yore: an imagined community (Anderson 1983; Damsgaard 2014; Korsgaard 2004).

[1]Persistent and long-memory processes are used interchangeably when referring to time series for which $0.5 < H < 1$.

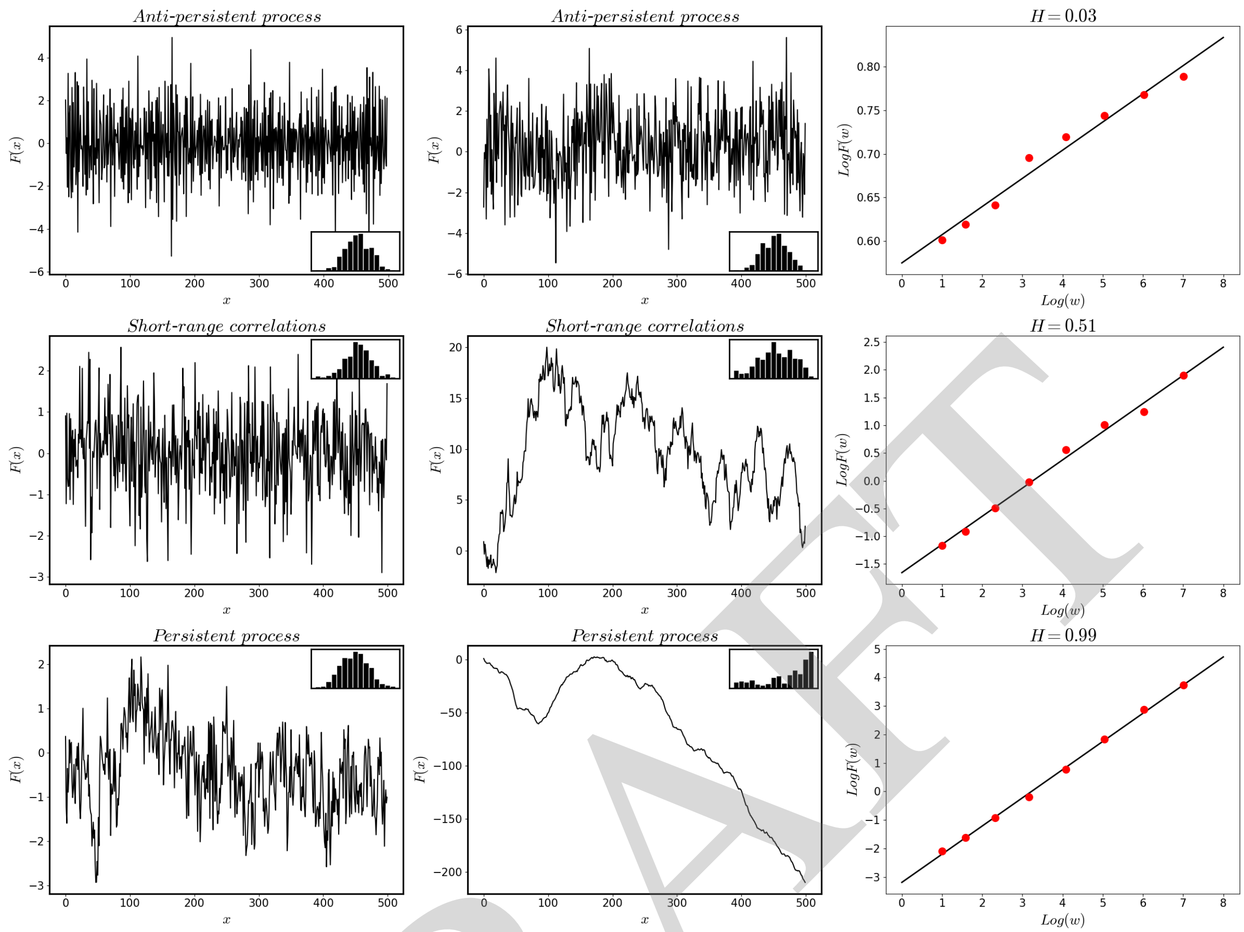

Fig. 1: Processes that exhibit anti-persistent (a-c), short memory (d-f), and persistent (g-i) behavior. Left to right shows noise-like processes (a,d,g), random walks obtained from the noise-like processes (b,e,h), and, finally, estimation of the Hurst exponent for matching process (c,f,i). Many cultural processes are noise-like (a,d,g), but in order to estimate the Hurst exponent it is necessary to transform them to random walks through first-order integration (see appendix A, equation 4). In most cases, we are therefore not studying the process directly, but its incremental structure (b,e,h). The behavior of the process (anti-persistent, short-term, persistent) can often be observed at the initial plot (a,d,g). The anti-persistent process (a) oscillates rapidly around its average at $F(X) = 0$ and short memory process (b) contains short wave-like spikes. The persistent process (c) on the other hand contains copies of itself across multiple time-scales. The last feature is an expression of self-similarity, which becomes more apparent for the persistent random walk process (h), where rounded mountain-like structures are embedded in the process at short, medium and long time scales. The degree of self-similarity in a process can be estimated by Adaptive Fractal Analysis (c,f,i),(see appendix A, equation 5-6), where a line is fitted to the residuals $F(w)$ of the detrended process over multiple time-scales or windows w. The Hurst parameter H is estimated as the slope of the linear fit and describes how fast the overall average amplitude $F(w)$ grows with increasing window size w. For the anti-persistent process, the slope < 0.5, for the process with only short-range correlations (i.e., short memory) it is approximately 0.5, and for the persistent process the slope is > 0.5. Notice that a linear fit is not a very accurate description of the anti-persistent process (c), which at short time scales ($w < 16$) is steeper, while at longer time-scale ($w \geqslant 16$) is more flat. This is an indication of a multi-fractal process. In this example, the process still remains anti-persistent because the slope is < 0.5 at both time scales.

Ambitions of identity formation were incorporated in his work on the Danish school system - mainly the secondary or tertiary levels of education were incorporated in his visions for the so-called folk high schools (Grell 1998) aiming to cultivate the "civil esprit" in the Danish youth (Balle 2014; Møller 2017). But identity formation was also dominant within his work on the institutional church culture. Questions of, on the one hand, church historical continuity and authenticity and, on the other, modern liberal reflections on religious freedom are prevalent in his scholarly-theological writings as well as his poetic, hymnic production (Baunvig 2017). Furthermore, it is held that Grundtvig helped form Danish democratic culture (Nevers 2011) through his work in the Constituent National Assembly (1848) and in parliament (1849-58, 1866)(Korsgaard 2012): He has gained the position of a founding father and is referenced within the manifestos of all the main political parties represented in parliament with the exception of one. Grundtvig is an integral part of Danish cultural self-understanding and his work is part of the Danish Ministry of Education's list over important historical roots of Danish democracy.[2]

Ever since its foundation within the late 1940'ies, research in Grundtvig as an academic discipline, which is driven by theologians, literary scholars and historians, has been directed by an eagerness to detect one or more biographical or literary turning points within the collected writings (Pedersen 2003). Within the ever-growing bulk of Grundtvig studies the years 1810, 1825, 1832, 1835, 1838, 1839, 1848 are competing for scholarly attention. Here discussions on psychological development and shifts in intellectual orientations are often condensed into teleological proclamations that one specific year represents Grundtvig's arrival at his 'authentic state' as it were. The work of theologian Kaj Thaning and his emphasis on 1832 is perhaps the most obvious representative for this line of work (Thaning 1963; Nielsen 2002; Pedersen 1994).

Grundtvig's collected writings consist of 921 works with a median work length of four pages. The data set covers a variety of genres (e.g., hymns, poems, speeches, dissertations and letters) and spans eight decades (1804-1871). Inversely to the year fetishism, the diversity of the text material has time and again been the target of scholarly reflection (Auken 2014). This interest can to some extent be seen as a direct reaction to what is considered to be an over-emphasis on biographical questions such as authorship development and (sequential) coherence. The sum of individual texts can hardly be considered to constitute an oeuvre as such being as they are highly bound to their specific production context, is the argument - echoing a new criticism logic (Auken 2005; Lundgreen-Nielsen 1980). We accept this criticism of the naïve interest in the Grundtvig biography. However, we wish to explore the mental dynamics of Grundtvig as an author (Baunvig and Nielbo 2017). Here a focus on text complexity and not on specific semantic statements within the texts is convenient.

[2]Danish version available at: http://static.uvm.dk/Publikationer/2008/demokratikanon/kap38.html

Related Work

There is a range of statistical measures of lexical variability originating in linguistics, which are based on the number of unique words (or types). The most widely known is probably the Type-Token Ratio: $TTR = N_{lex} / \sum_i^n Fr(w_i) \times 100$, where N_{lex} is the size of the lexicon and Fr is the frequency of word w_i. TTR and its derived measures reliably function as indices for linguistic development (Snowdon D.A., Kemper S.J., Mortimer J.A., Greiner L.H., Wekstein D.R., and Markesbery W.R. 1996) and carry information about mental decay, specifically, onset of dementia and Alzheimer's Disease (Berisha, Wang, LaCross, and Liss 2015; Le, Lancashire, Hirst, and Jokel 2011; Garrard, Maloney, Hodges, and Patterson 2005). Even in certain non-pathological cases measures of lexical variability can track age-related lexical decay (Snowdon D.A., Kemper S.J., Mortimer J.A., Greiner L.H., Wekstein D.R., and Markesbery W.R. 1996). Crime novelist Agatha Christie, for instance, showed a marked increase in lexical and syntactic redundancy with age (Lancashire and Hirst 2009). Importantly, TTR is also linked to creativity in sentence writing (Zhu, Xu, and Khot 2009) and musical lyrics (Hu and Yu 2011) with higher lexical diversification indicating greater creativity.

In similarity with TTR, entropy captures patterns of lexical variability, because it summarizes most of the information in the frequency distribution of a text. Entropy is more general than TTR because it applies to non-linguistic domains and originates outside linguistics in a formal and axiomatic approach to information (Shannon 1948).[3] For time series objects, such as a text, entropy captures its variability and is indicative of its complexity such that larger values indicate greater complexity (Bandt and Pompe 2002). Time series analysis of source and relative entropy for texts can be used to identify change points and trends in lexical complexity (Zhang 2016), lexical innovation and preservation (Barron, Huang, Spang, and DeDeo 2017; Evrard 1985), and finally model the impact of biographical events in an author's intellectual life (Murdock, Allen, and DeDeo 2015). For studies of cognitive development and cultural evolution, entropy has been suggested as a unifying principle and fundamental driver linked to restructuring, innovation and creativity (Bilder and Knudsen 2014; Gabora 2016; Rigau, Feixas, and Sbert 2007).

Although analysis of persistent $1/f^{2H+1}$ processes has great potential for cultural and historical research where multiple unknown parameters are typically the norm, it has to our knowledge rarely been applied. One study found that the Hurst exponent of article content is predictive of the global terrorism index, because countries impacted by armed conflict have a high degree of persistence in reports of deaths, injuries and incidents (i.e., $H > 0.5$ for negative events) (Gao, Fang, and Liu 2017). The Hurst exponent has also been shown to be sensitive to content (Gao, Hu, Mao, and Perc 2012), media type and the dynamics of consumer products in large scale text analysis (Wevers, Nielbo, and Gao *submitted*).

[3]The concept of entropy actually originates in classical thermodynamics, but it is Shannon's information theoretical construct that is the focus of this article.

Results

Entropy and Global Dynamics

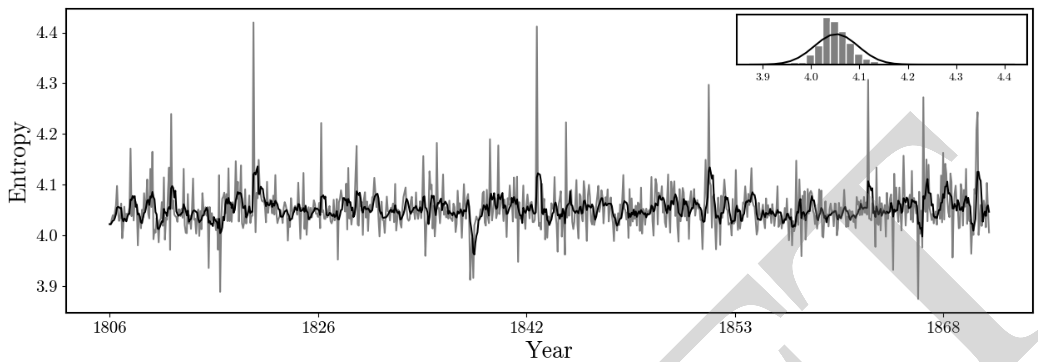

Fig. 2: Averaged entropy computed over slices of 1000 words for each of Grundtvig's time-ordered works. The distribution of entropy in upper right corner approximate a Gaussian normal as indicated by the fit (black line) with a positive skew.

The raw entropic process of Grundtvig's active years is shown in figure 2. The average amplitude of variation measured by the root mean square is $RMS = 4.05$, $SD = 0.05$ with some extreme values, which are responsible for the positive skew of the distribution. No overall linear trend can be discerned from the entropy of Grundtvig's writings. This was confirmed by using the ordinary least squares methods to fit entropy on year: $F < 1$. At first glance the data do, in other words, not indicate decay of lexical variability or text complexity. To identify possible persistent correlations, AFA was applied to the full time series in order to estimate its Hurst parameter. Results primarily indicated short-term correlations with a small tendency to anti-persistent behavior: $H = 0.47$. Over Grundtvig's entire life-time we do, in other words, not observe persistent trends in the Grundtvig's text complexity. Searching for persistence in the entire signal might however be misleading, because it assumes one coherent developmental trajectory that spans eight decades of psychological and social events. By investigating variation in the Hurst exponent over a moving time-window instead, we can capture long-range trends and remain sensitive to local dynamics.

Time-varying Hurst Exponent

For the time-varying Hurst exponent $RMS = 0.51$, $SD = 0.04$, which indicates virtually memoryless central tendency. By observing the signal's behavior as the difference, ΔH from the memoryless state of $H = 0.5$, it becomes apparent that the positive difference (i.e., a trend towards persistence) groups in the early part of Grundtvig's career, while the anti-persistent dynamics become more pronounced with age (see figure 3). To minimize effects of temporal displacement and emphasize stable trends, an adaptive filter was applied to the signal, which combines local linear trends with a global smooth trend (black bars in figure 3). The observable age-dependent trend was corroborated by correlating ΔH with age, which confirms in a statistically reliable inverse relationship: $r = -.67, p < .00001$. The time-varying Hurst exponent does in other words show evidence of entropic decay over the life-time of Grundtvig.

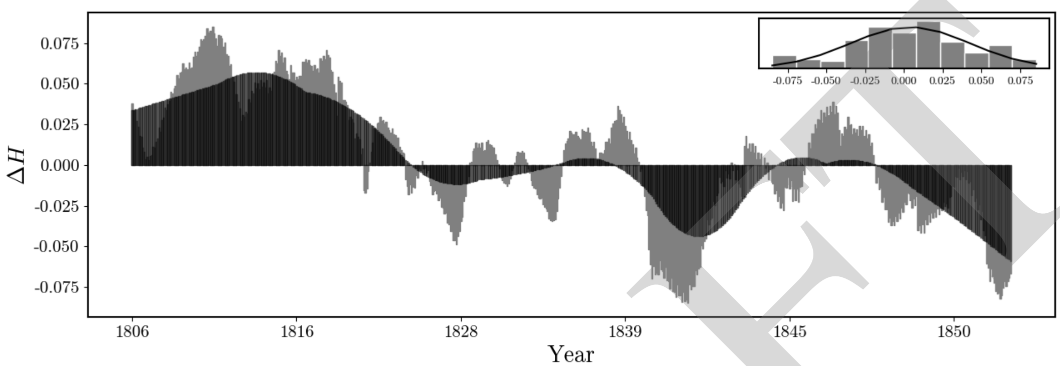

Fig. 3: The difference (ΔH) from short memory only ($H = 0.5$) for the time-varying Hurst exponent of entropy. Gray bars represent the original difference, while black bars are the results of adaptive filtering in order to emphasize stable trends.

To further explore this process of entropic decay, we emphasize two related levels of resolution: a coarse- and a fine-grained level (see figure 4 and table 1). The resolution levels are based on a simple mean-shift model for change point analysis that identifies changes in the mean of ΔH that are statistically reliable at two α-thresholds (Kulkarni, Al-Rfou, Perozzi, and Skiena 2015; Tyler 2000). At the coarse level, the persistence in text complexity of Grundtvig's writings can be split in two parts based on a change point in 1826. The early phase before 1826 is characterized by persistent dynamics such that a jump in complexity 'inspires' another jump up in the incremental process (and vice versa). The later phase, on the other hand, is characterized by a trend towards increasing anti-persistent dynamics such that Grundtvig can be said to rely increasingly on the same average level of complexity with age, that is, a jump up 'promotes' a jump down in the incremental process (and vice versa) resulting in the mean-reverting behavior of the signal. The fine-grained level detects multiple phases that oscillate between short memory and anti-persistent dynamics embedded in the second coarse phase. Immediately after the early persistent phase, Grundtvig switches into a phase dominated by short memory only, which lasts until 1839. Between 1839 and 1845 Grundtvig's writings show predominantly anti-persistent dynamics, which is succeeded by a three year period of short memory. Finally, Grundtvig's late writings switches back to mean-reverting behavior.

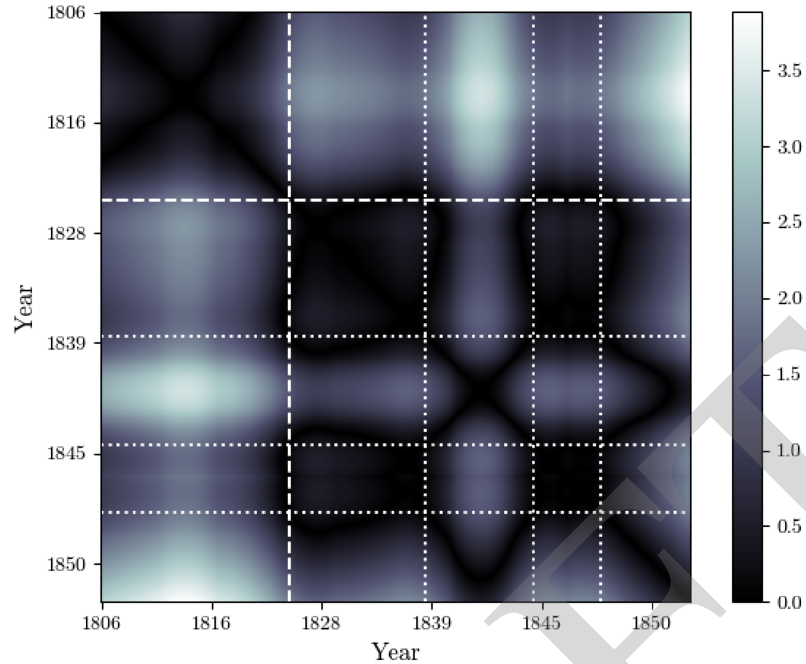

Fig. 4: Distance matrix for ΔH data with coarse (white dashed lines) and fine (white dotted lines) resolution level overlaid. Dark rectangular areas along the main diagonal indicates phases, while dark/light squares off the main diagonal indicate similar/dissimilar phases, respectively. It is apparent that the dynamics of the early 1806-1826 phase are dissimilar ($\Delta H > 0$) from all later phases ($\Delta H \leq 0$), but the fine grained changes in the later phase can also be observed by the checkered pattern after 1826. Normalized Euclidean distance was used as distance metric.

Table 1: Dominant Dynamic in the Phases of N.F.S. Grundtvig's Writings

Time period	Age of onset	Coarse	Fine	Behavior
1806-1826	23	$H > 0.5$	$H > 0.5$	*persistent*
1826-1839	43	$H \leq 0.5$	$H \approx 0.5$	*short memory*
1839-1845	56	$H \leq 0.5$	$H < 0.5$	*anti-persistent*
1845-1848	62	$H \leq 0.5$	$H \approx 0.5$	*short memory*
1849-1872	65	$H \leq 0.5$	$H < 0.5$	*anti-persistent*

To summarize, we can see that Grundtvig's collected writings do not indicate decay in the raw entropic signal. If, however, we model the time-varying Hurst exponent, we do find a negative trend that corresponds to decay in the persistence of the text complexity. Interestingly, the method allows us to explore the signal at multiple resolutions and detect several hierarchically related points of change in Grundtvig's writings.

Concluding Remarks

By combining measures from information theory and random fractal theory, the present study has modeled an aspect of a prolific writer's developmental trajectory as it is reflected in persistent trends of text complexity. The results pertaining to the case study of Grundtvig require further discussion as does possible extensions of the approach to multiple authors and other cultural information systems.

Case study

Based on entropy, there is no indication of lexical decay in Grundtvig's collected writings. This is not surprising because lexical measures that are similar to entropy, such as TTR, are primarily suited for identifying decay related to pathological conditions, and there are no records indicating that Grundtvig suffered from dementia. There are some indications that Grundtvig suffered from bipolar disorder (Helweg 1932), traces of which can be identified by analysis utilizing lexical variability (Baunvig and Nielbo 2017). The two-phase model based on the coarse-grained level of analysis essentially identifies decay in persistent complexity as a function of age with a change point round Grundtvig's 43$^{\text{th}}$ year (see table 1). It is tempting to interpret this change point as the creative parallel for art and science to Hardy's claim that "mathematics ... is a young man's game" (Hardy 1967). Beyond aging, the specific change point possibly reflects events in Grundtvig's life.

In the early 1820'ies Grundtvig became a deacon assisting ministerial work first in the rural village Præstø (1821) later in Copenhagen (1822). Grundtvig scholars point to 1821-1824 as an important period for Grundtvig's professional maturation (Pontoppidan Thyssen 1983). For the first time in his professional career Grundtvig regularly had the opportunity to test his theological reflection on a specific congregation through his sermons and hymnic production. For Grundtvig, scholars have repeated, this lead to a new definition of the protestant church emphasizing the role of congregational community. And as such 1821-1824 is generally considered a prelude to the climax of 1825-1826 (Koch 1959; Thodberg 1983; Pedersen 2003). This is considered one of the most important biographical turning points:

In 1825 Grundtvig publishes the highly polemic pamphlet *Kirkens Gienmæle* ("Response of the Church") criticizing professor H.N. Clausen and his recent theological dissertation. Grundtvig's critique is an application of his ideas on the importance of lay "energy" as opposed to overly academic theological reflection. His text led to a defamation verdict and censorship. As a consequence Grundtvig resigned as deacon in 1826. From 1826 and onwards, Grundtvig's writings become less theoretical and the polemic elements are somewhat reduced. Instead he strengthens a pragmatic focus with the congregation as the center of the church. Gradually, theoretical papers are substituted by psalms and pedagogical-devotional literature reflecting a fundamental transition from theological thinker to a more praxis-oriented and pragmatic pastor centered on 1825-26 (Baunvig 2013).

Based only on the persistence of aggregated complexity, the model manages to correctly identify this "Road to Damascus" moment in Grundtvig's life. A finer and more detailed account of the two-phase model's second phase can be obtained from the five phase model, which differentiates between four phases post 1826. During its second 1826-1839 phase, Grundtvig becomes a freelancer and conducts several study trips to the United Kingdoms. From the model, we know that his writings' complexity lack persistence during this phase and therefore only displays dependencies at short time scales (i.e., only works that are close by show dependence). We speculate that the lack of persistence is indicative of an author that acquires new experiences that only temporarily change or stabilize in the author's creative space. This line of thought corresponds with trends in the Grundtvig literature stressing the 1830'ies as a "Sattelzeit" (Thaning 1963; Christensen 2013).

In 1838 at the initiation of the third phase, Grundtvig receives his public breakthrough with the Mands Minde ("Living Memory") lectures held at Borch's College in Copenhagen; this was a decisive cultural event in mid-19th century Denmark (Barfod 1839; Kaalund 1839) - ultimately leading to the development of the Danish folk high schools (Koch 1954; Lundgreen-Nielsen 1998). Immediately following this, in 1839, Grundtvig receives his second pastoral employment, this time in Vartov Copenhagen. The position was given to him by king Frederik 6th of Denmark. During this third phase, Grundtvig's writings display anti-persistent behavior reverting to an average model of complexity. His mental states do in other words appear fixed or rigid, which is likely to reflect an early onset of old age combined with repetitions of slogans that the young men at Borch's College enjoyed (Baunvig 2014).

The fourth phase is a short intermediary period, where the writings briefly returns to a short memory process that lack persistence and continuity. During this stage Grundtvig is not well, struggling both with his physical and mental health. Early research explores Grundtvig's lack of "aandelig Ligevægt" (mental stability) (Rönning 1907; Helweg 1932) during these years.

Finally, this short memory phase is replaced by the final phase where Grundtvig returns to the anti-persistent dynamic. Beyond old age, it is likely that the mean-reverting behavior reflects the dispersion of Grundtvig's discursive activity into an increasing number of genres (Auken 2014) due to Grundtvig's dual (official) obligations as pastor and politician from 1949 and onwards.

Further Applications

The approach we have presented is applicable beyond the specific case study. We envision three applications, where the approach can contribute to the study of dynamics in the social sciences and humanities: author transitions, author comparisons, and dynamics of cultural information systems.

Results from the case study illustrate the approach's potential for *Kehre* detection, that is, the detection of significant transitions or change points in an author's production and the underlying mental correlate. The two-phase model of Grundtvig resonates with theoretical arguments for distinct younger and older profiles in the collected writings of, for instance, philosopher M. Heidegger, theologian M. Luther or author M. Kundera. Similarly, the construct "late style" describes a distinct late phase in an artist's production (Hutchinson 2016; McMullan and Smiles 2016; Said 2006). Our approach can be used to validate such claims of an author's mental *Kehre*. A current constraint is the number of samples necessary to get a reliable estimate of the Hurst exponent. Currently, we are using sample size of 256 works for an accurate measure of H, which limits the technique to highly prolific authors. It is possible to get accurate estimates with as little as 120 data points (Gao, Hu, Mao, and Perc 2012). Another way to circumvent this limitation is by sampling multiple times from the individual works. One could for instance treat slices of n tokens as works keeping in mind that this process will a) result in fewer samples for estimating entropy and b) introduce a hierarchy of slices where some slices come from the same original work and other from different original works. Currently we are only comparing entropy estimates that originate in different original works. Another concern originates in the bias that incorrect or insufficient dating might introduce. If for instance Grundtvig wrote all his early works in a time window that was significantly shorter than the period spanned by the publication dates and his later works were uniformly distributed across the publication dates, we would be over-emphasizing the persistence for the early Grundtvig. Whether this represents an issue, which in our case it does not, is an empirical problem related to the specific data set.

Compressing an author's production to a one-dimensional series of temporally ordered samples, which captures a relevant feature of the data makes several comparisons within and between authors possible. As an example of a relevant within author comparison, take the construct of the "suffering artist", which predicts an inverse and asymmetric relation between creative and emotional states of an author, such that emotional states predict creative states. Make the assumption that sentiment analysis can track an affective dimension of an author, [4] our approach can then test if an author is an instance of the "suffering artist". This can be done by modeling the asymmetric relations between (long-range dependencies in) affective content and text complexity as an estimate of causal directionality. Similarly, we can compare similarities and dependencies between authors' developmental trajectories using either basic correlational methods or more advanced sequence alignment. How, for instance, does H.C. Andersen's developmental trajectory differ from Grundtvig and to what extend are their production impacted by the same external events? An interesting extension of such between-author comparisons is a typology of author dynamics that controls for demographic variables such as age, gender, ethnicity and educational level. A challenge, however, is identification of metrics that can estimate relevant properties from the available data. An advantage of Shannon's source entropy, as well as other metrics based on information theory, is that they track properties that are related to general information processing (Bilder and Knudsen 2014; Gabora 2016). A further advantage of using information theory is that it is not confined to linguistic data, which in principle makes our approach applicable to other kinds of cultural media such as images and audio.

Finally, we want to mention that although our approach is developed for studying the dynamics of a single author, there is no principled reason that it cannot be applied to higher level systems. Whenever we can reasonably assume that a collection of texts is an expression of the evolution of a cultural information system, the approach can the applied to study the dynamics of the system's complexity. While the study of entropic trends in collections of homogeneous texts is not a new topic

[4]Although this assumption can be problematic, it is not uncommon in author profiling (Argamon, Koppel, Fine, and Shimoni 2003; Schler, Koppel, Argamon, and Pennebaker 2006).

(Zhang 2016), the advantage of our approach is that it goes beyond simplistic central tendencies such as RMS and focuses on the fractal properties, specifically long range dependencies, which can capture fine-grained structure in the system's variable and dynamic behavior. It is important to emphasize that for applications beyond a single author, the interpretation of the results has to rely on proper domain knowledge.

Acknowledgements

Part of this research was performed while K.L. Nielbo and J. Gao were visiting the Institute for Pure and Applied Mathematics (IPAM), which is supported by the National Science Foundation. The data analyzed here were retrieved from http://www.grundtvigsværker.dk/.

Appendix A. Methods

Data

The writings of Grundtvig were collected from the database available at www.grundtvigsværker.dk. At the time of acquisition, the database contained 921 works with a median work length of four pages. This data set covers a variety of genres (e.g., hymns, poems, novels, and letters) and spans eight decades (1804-1871). For estimation of word-level entropy, only alphabetic characters were kept in the data set and each work was sliced into segments of 1000 words for the purpose of length normalization.[5] The entropy of a work was then computed as the average entropy of the work's segments. For time series analysis, the works were sorted temporally by combining metadata for publication year with the sequential order from the database. The Hurst exponent was estimated for windows of 256 works using a sliding window with maximal overlap.

Entropy

Information theory offers a general framework for measuring informational variation and dependencies in strings such as character and word sequences. In the discrete case with K distinct values, classical source entropy h of Shannon is a measure of variability. For this study h is measured at the word level with a lexicon of K distinct types accordingly:

$$h = -\sum_{i=1}^{K} p_i \times \log_2(p_i) \tag{1}$$

[5]In general, the results are robust for segments of 250, 500 and 1000 words.

with p being defined as:

$$p_i = Fr(w_i) / \sum_{i=1}^{K} Fr(w_i) \tag{2}$$

where Fr is the absolute frequency of word w. Word-level entropy measures the lexical variability of a tokenized string and results in $0 \leqslant h \leqslant \log_2 K$ with $h = \log_2 K$ if the words are uniformly distributed (i.e., a string where each word occurs with equal probability) and $h = 0$ if the string is perfectly predictable (e.g., a string with repetitions of only one word). Entropy is, as other measures of lexical variability, influenced by text length. On the assumption that a corpus consists of homogenous texts, differences in length may be a relevant informational feature (Zhang 2016), this assumption, however, is not valid for authorships covering multiple genres and styles. In such cases length normalization is a more viable solution, where entropy is either normalized by text length or estimated based on slices of equal length.

Fractal analysis

Fractal Dynamics

Many natural and man-made complex dynamic systems are fractal in the sense that they display self-similar and scale-invariant behavior. In the context of stochastic systems, self-similarity means that the system's fluctuation patterns at faster time-scales resemble fluctuation patterns at slower time scales, while scale-invariance means that the measurement of these patterns does not depend on the resolution of the time scale of the measurement (Riley, Bonnette, Kuznetsov, Wallot, and Gao 2012). In order to determine if a system is fractal, fractal analysis therefore examines the relationship between the measurement and it's time scale, specifically, whether this relationship is characterized by power-law scaling (see figure 1).

A $1/f$ fractal process has a power-law decaying spectral density and it can therefore not be adequately modelled by standard techniques for time series analysis, such as an ARIMA model or a Markov process, because they have distinctly different spectral densities. A $1/f^{2H+1}$ process where $0 < H < 1$ is a non-stationary random-walk process, the differentiation of which is a covariance stationary stochastic process with mean μ, variance σ^2, and autocorrelation function (Cox 1984):

$$r(w) = E(X_t X_{t+w})/E(X_t^2) \sim w^{2H-2}, \;\; as \;\; w \to \infty. \tag{3}$$

where w is the time lag. In this case X has a power spectral density $1/f^{2H-1}$. To adequately model a $1/f$ process, a fractional order process has to be used such as the fractional Brownian motion model (Mandelbrot 1982). For fractal analysis it is helpful to understand the difference between fractional Brownian motion (fBm) and fractional Gaussian noise (fGn). Both types of signals are characterized by long-memory, that is, they exhibit correlations over longer time scales. But where fGn is a stationary process (i.e., its mean or variance do not change over time), fBn

is non-stationary (i.e., its mean or variance show a time dependent trend) and has a power-law increasing variance (t^{2H}), and power-law decaying power spectral density ($1/f^{2H+1}$) (Mandelbrot and Ness 1968; Beran 1994; Mandelbrot 1997; Kuznetsov, Bonnette, Gao, and Riley 2013). The two types of signals are related because fBn process can be created from a fGn through integration and fGn from fBn through differentiation (Eke, Herman, Kocsis, and Kozak 2002). The Hurst H exponent quantifies persistence or memory in time series, where $0 < H < 0.5$ is an anti-persistent process, $H = 0.5$ is a short-memory process, and $0.5 < H < 1$ is a persistent process (Gao, Hu, and Tung 2011). The interpretation of H, however, depends on characterizing the signal as fGn or fBn. H describes the correlation structure for fGn, while it describes the correlation structure for increments for fBm (Riley, Bonnette, Kuznetsov, Wallot, and Gao 2012; Cannon, Percival, Caccia, Raymond, and Bassingthwaighte 1997). For the present study of $1/f^{2H+1}$ processes should be interpreted as the latter. A persistent process indicates continuity of text complexity (i.e., entropy levels will last for a long time). An anti-persistent indicates rigidity (i.e., entropy will rapidly decay to a mean state), and finally, short memory indicates a lack of continuity (entropy will only be correlated at short time scales).

Adaptive Fractal Analysis

Detrended fluctuation analysis (DFA) (Peng, Buldyrev, Havlin, Simons, Stanley, and Goldberger 1994) is a widely used method for estimating the Hurst parameter for a time series. DFA consists of five steps, 1) initially a random walk process is constructed from the time series:

$$u(n) = \sum_{k=1}^{n}(x_k - \overline{x}), \quad n = 1, 2, \cdots, N, \tag{4}$$

where $\overline{x}$ is the mean of the series $x(k), \ k = 1, 2, \cdots, N$; 2) dividing the constructed random walk process into non-overlapping segments; 3) determining the local trends of each segment as the best polynomial fit; 4) getting the variance of the differences between the random walk process and the local trends; and 5) determining the average variance over all the segments. DFA may involve discontinuities at the boundaries of adjacent segments. Such discontinuities can be detrimental when the data contain trends (Hu, Ivanov, Chen, Carpena, and Eugene Stanley 2001), non-stationarity (Kantelhardt, Zschiegner, Koscielny-Bunde, Havlin, Bunde, and Stanley 2002), or nonlinear oscillatory components (Chen, Hu, Carpena, Bernaola-Galvan, Stanley, and Ivanov 2005; Hu, Gao, and Wang 2009). Adaptive fractal analysis (AFA) is an alternative to DFA that solves these problems (Gao, Hu, and Tung 2011). The main advantage of AFA over DFA is that it identifies a global smooth trend, which is obtained by optimally combining local linear or polynomial fitting, and thus no longer suffers from DFA's problem of discontinuities of adjacent segments. As a result, AFA can automatically deal with arbitrary, strong nonlinear trends (Gao, Hu, and Tung 2011; Hu, Gao, and Wang 2009).

AFA is based on a nonlinear adaptive multi-scale decomposition algorithm (Gao, Hu, and Tung

2011). The first step involves partitioning an arbitrary time series under study into overlapping segments of length $w = 2n + 1$, where neighboring segments overlap by $n + 1$ points. In each segment, the time series is fitted with the best polynomial of order M, obtained by using the standard least-squares regression; the fitted polynomials in overlapped regions are then combined to yield a single global smooth trend. Denoting the fitted polynomials for the $i - th$ and $(i + 1) - th$ segments by $y^i(l_1)$ and $y^{(i+1)}(l_2)$, respectively, where $l_1, l_2 = 1, \cdots, 2n + 1$, we define the fitting for the overlapped region as

$$y^{(c)}(l) = w_1 y^{(i)}(l + n) + w_2 y^{(i+1)}(l), \quad l = 1, 2, \cdots, n + 1, \tag{5}$$

where $w_1 = (1 - \frac{l-1}{n})$ and $w_2 = \frac{l-1}{n}$ can be written as $(1 - d_j/n)$ for $j = 1, 2$, and where d_j denotes the distances between the point and the centers of $y^{(i)}$ and $y^{(i+1)}$, respectively. Note that the weights decrease linearly with the distance between the point and the center of the segment. Such a weighting is used to ensure symmetry and effectively eliminate any jumps or discontinuities around the boundaries of neighboring segments. As a result, the global trend is smooth at the non-boundary points, and has the right and left derivatives at the boundary (Riley, Bonnette, Kuznetsov, Wallot, and Gao 2012). The global trend thus determined can be used to maximally suppress the effect of complex nonlinear trends on the scaling analysis. The parameters of each local fit is determined by maximizing the goodness of fit in each segment. The different polynomials in overlapped part of each segment are combined using Equation 5 so that the global fit will be the best (smoothest) fit of the overall time series. Note that, even if $M = 1$ is selected, i.e., the local fits are linear, the global trend signal will still be nonlinear. With the above procedure, AFA can be readily described. For an arbitrary window size w, we determine, for the random walk process $u(i)$, a global trend $v(i), i = 1, 2, \cdots, N$, where N is the length of the walk. The residual of the fit, $u(i) - v(i)$, characterizes fluctuations around the global trend, and its variance yields the Hurst parameter H according to the following scaling equation:

$$F(w) = \Big[\frac{1}{N}\sum_{i=1}^{N}(u(i) - v(i))^2\Big]^{1/2} \sim w^H. \tag{6}$$

By computing the global fits, the residual, and the variance between original random walk process and the fitted trend for each window size w, we can plot $\log_2 F(w)$ as a function of $\log_2 w$. The presence of fractal scaling amounts to a linear relation in the plot, with the slope of the relation providing an estimate of H (Fig. 1).[6]

[6]Code for computing DFA and AFA is available at https://github.com/knielbo/saffine.

References

Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. New York, NY: Verso.

Argamon, Shlomo, Moshe Koppel, Jonathan Fine, and Anat Rachel Shimoni. 2003. "Gender, genre, and writing style in formal written texts." *Text - Interdisciplinary Journal for the Study of Discourse* 23:321–346.

Auken, Sune. 2005. *Sagas spejl: Mytologi, historie og kristendom hos N.F.S. Grundtvig*. Gyldendal.

Auken, Sune. 2014. *Grundtvig og genrerne. En introduktion*. Forlaget Spring.

Bak, Per, Chao Tang, and Kurt Wiesenfeld. 1987. "Self-organized criticality: An explanation of the 1/ f noise." *Physical Review Letters* 59:381–384.

Balle, Thorstein. 2014. "Myten om Grundtvigs indflydelse på den danske folkeskole." *Grundtvig-Studier* 65:65–98.

Bandt, Christoph and Bernd Pompe. 2002. "Permutation Entropy: A Natural Complexity Measure for Time Series." *Physical Review Letters* 88.

Barfod, Frederik. 1839. *Den Sidste Aften paa Grundtvigs Forelæsninger: En lille Nyaarsgave til hans Venner og Tilhørere*. C.A. Reitzel.

Barron, Alexander T J, Jenny Huang, Rebecca L Spang, and Simon DeDeo. 2017. "Individuals, Institutions, and Innovation in the Debates of the French Revolution." *arXiv:1710.06867* p. 8.

Baunvig, Katrine. 2013. *Forsamlingen først. N.S.F. Grundtvigs og Émile Durkheims syn på fællesskab*. Aarhus University.

Baunvig, Katrine. 2014. *Det folkelige foredrag*. Forlaget Spring.

Baunvig, Katrine. 2017. *Kirkehistorisk adaptation. Om Grundtvig, (Grundtvigs) Luther og kulturel evolution*. Forlaget Anis.

Baunvig, Katrine F. and Kristoffer L Nielbo. 2017. "Kan man validere et selvopgør?" *Nordiskt Nätverk för Editionsfilologer Skrifter* 2:45–67.

Beran, Jan. 1994. *Statistics for Long-Memory Processes*. New York: Chapman and Hall/CRC, 1 edition.

Berisha, Visar, Shuai Wang, Amy LaCross, and Julie Liss. 2015. "Tracking Discourse Complexity Preceding Alzheimer's Disease Diagnosis: A Case Study Comparing the Press Conferences of Presidents Ronald Reagan and George Herbert Walker Bush." *Journal of Alzheimer's Disease* 45:959–963.

Bilder, Robert M. and Kendra S. Knudsen. 2014. "Creative cognition and systems biology on the edge of chaos." *Frontiers in Psychology* 5.

Cannon, Michael J., Donald B. Percival, David C. Caccia, Gary M. Raymond, and James B. Bassingthwaighte. 1997. "Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series." *Physica A: Statistical Mechanics and its Applications* 241:606–626.

Chater, Nick and Gordon DA Brown. 1999. "Scale-invariance as a unifying psychological principle." *Cognition* 69:B17–B24.

Chen, Zhi, Kun Hu, Pedro Carpena, Pedro Bernaola-Galvan, H. Eugene Stanley, and Plamen Ch. Ivanov. 2005. "Effect of nonlinear filters on detrended fluctuation analysis." *Phys. Rev. E* 71:011104.

Christensen, Bent. 2013. "Kirke og meninghed i Grundtvigs teologi og kirkepolitik 1806-1861." *Grundtvig-Studier* .

Cox, Timothy M. 1984. "Long range dependence: A review." In *Iowa State University*. Press.

Damsgaard, Johan. 2014. *Fædrelandssangen*. Forlaget Spring.

Donald, Merlin. 2007. "The slow process: a hypothetical cognitive adaptation for distributed cognitive networks." *journal of Physiology* 101:214–222.

Eke, A., P. Herman, L. Kocsis, and L. R. Kozak. 2002. "Fractal characterization of complexity in temporal physiological signals." *Physiological Measurement* 23:R1.

Evrard, Etienne. 1985. "Richesse et mode d'enrichissement d'un vocabulaire." *Actes de la 11e Conférence intern* pp. 147–152.

Gabora, Liane. 2016. "A Possible Role for Entropy in Creative Cognition." p. E001. MDPI.

Gao, Jianbo, Peng Fang, and Feiyan Liu. 2017. "Empirical scaling law connecting persistence and severity of global terrorism." *Physica A: Statistical Mechanics and its Applications* 482:74–86.

Gao, Jianbo, Jing Hu, Xiang Mao, and Matjaž Perc. 2012. "Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries." *Journal of The Royal Society Interface* p. rsif20110846.

Gao, Jianbo, Jing Hu, and Wen-wen Tung. 2011. "Facilitating Joint Chaos and Fractal Analysis of Biosignals through Nonlinear Adaptive Filtering." *PLoS ONE* 6:e24331.

Garrard, Peter, Lisa M. Maloney, John R. Hodges, and Karalyn Patterson. 2005. "The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author." *Brain* 128:250–260.

Grell, Helge. 1998. *Vision og virkeliggørelse: En redegørelse for Grundtvigs tanker om folkelig oplysning og en folkelig højskole og for forsøget på at virkeliggøre dem*. Grundtvig Selsskabet.

Hall, John A., Ove Korsgaard, and Ove K. Pedersen. 2015. *Building the Nation: N.F.S. Grundtvig and Danish National Identity*. Montreal ; Kingston ; Ithaca: McGill-Queen's University Press.

Hardy, G. H. 1967. *A Mathematician's Apology*. Cambridge University Press.

Helweg, Hjalmar. 1932. *N. F. S. Grundtvigs Sindssygdom*. Copenhagen, second edition.

Hu, Jing, Jianbo Gao, and Xingsong Wang. 2009. "Multifractal analysis of sunspot time series: the effects of the 11-year cycle and Fourier truncation." *Journal of Statistical Mechanics: Theory and Experiment* 2009:P02066.

Hu, Kun, Plamen Ch. Ivanov, Zhi Chen, Pedro Carpena, and H. Eugene Stanley. 2001. "Effect of trends on detrended fluctuation analysis." *Physical Review E* 64.

Hu, Xiao and Bei Yu. 2011. "Exploring The Relationship Between Mood and Creativity in Rock Lyrics." In *ISMIR*, pp. 789–794.

Hutchinson, Ben. 2016. *Lateness and Modern European Literature*. New York, NY: Oxford University Press, 1 edition.

Kaalund, H.V. 1839. *"Kæmpeaand fra Kraftens Alder", Den Sidste Aften paa Grundtvigs Forelæsninger: En lille Nyaarsgave til hans Venner og Tilhørere*. C.A. Reitzel.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 1 edition.

Kantelhardt, Jan W., Stephan A. Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H. Eugene Stanley. 2002. "Multifractal detrended fluctuation analysis of nonstationary time series." *Physica A: Statistical Mechanics and its Applications* 316:87–114.

Koch, Hal. 1954. *Den Danske Kirkes Historie VI*. Gyldendal.

Koch, Hal. 1959. *N.F.S. Grundtvig*. Gyldendal.

Korsgaard, Ove. 2004. *Kampen om Folket. Et dannelsesperspektiv på dansk historie gennem 500 år*. Gyldendal.

Korsgaard, Ove. 2012. *Grundtvig som politiker*. Djøf Forlag.

Kulkarni, Vivek, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. "Statistically Significant Detection of Linguistic Change." pp. 625–635. ACM Press.

Kuznetsov, Nikita, Scott Bonnette, Jianbo Gao, and Michael A. Riley. 2013. "Adaptive Fractal Analysis Reveals Limits to Fractal Scaling in Center of Pressure Trajectories." *Annals of Biomedical Engineering* 41:1646–1660.

Lancashire, Ian and Graeme Hirst. 2009. "Vocabulary Changes in Agatha Christie's Mysteries as an Indication of Dementia: A Case Study." Toronto.

Le, Xuan, Ian Lancashire, Graeme Hirst, and Regina Jokel. 2011. "Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists." *Literary and Linguistic Computing* 26:435–461.

Liebovitch, Larry S. and Lina A. Shehadeh. 2003. "Introduction to fractals." *Tutrorials in contemporary nonlinear methods* 24:178.

Lundgreen-Nielsen, Flemming. 1980. *Det handlende ord. N.F.S. Grundtvigs digtning, litteraturkritik og poetik 1798-1819*. Gad.

Lundgreen-Nielsen, Flemming. 1998. *"Løven i buret", 1848 – det mærkelige år*. Museum Tusculanum.

Mandelbrot, Benoit. 1982. *The Fractal Geometry of Nature*. San Francisco: Times Books, updated ed. edition edition.

Mandelbrot, Benoit B. 1997. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Selecta Volume E*. New York: Springer, 1997 edition edition.

Mandelbrot, Benoit B. and John W. Van Ness. 1968. "Fractional Brownian Motions, Fractional Noises and Applications." *SIAM Review* 10:422–437.

Marchant, Thierry. 2008. "Scale invariance and similar invariance conditions for bankruptcy problems." *Social Choice and Welfare* 31:693–707.

McGavin, Lee H. 1997. "Creativity as Information: Measuring Aesthetic Attractions." *Nonlinear Dynamics, Psychology, and Life Sciences* 1:24.

McMullan, Gordon and Sam Smiles (eds.). 2016. *Late Style and its Discontents: Essays in art, literature, and music*. New York, NY: OUP Oxford, 1st edition edition edition.

Murdock, Jaimie, Colin Allen, and Simon DeDeo. 2015. "Exploration and Exploitation of Victorian Science in Darwin's Reading Notebooks." *arXiv preprint arXiv:1509.07175* .

Møller, Jes Fabricius. 2017. "Grundtvigs betydning for samfundet." *Grundtvig-Studier* 55:103–130.

Nevers, Jeppe. 2011. *Fra skældsord til slagord*. Odense Univsersitetsforlag.

Nielbo, Kristoffer L, Ryan Nichols, and Edward Slingerland. 2017. "Mining the Past – Data-Intensive Knowledge Discovery in the Study of Historical Textual Traditions." *Journal of Cognitive Historiography* 3:93–118.

Nielsen, Mikkel Crone. 2002. ". . . at tale med de døde. Om sækularisering og hermenutik I Kaj Thanings forfatterskab." *Grundtvig-Studier* 53:121–141.

Norris, Stephen M. 2016. "A Biographical Turn?" *Kritika: Explorations in Russian and Eurasian History* 17:163–179.

Nygaard, Bertel. 2012. *Revolution: Masser af modstand*. Aarhus Universitetsforlag.

O'Brien, Beth A., Sebastian Wallot, Anna Haussmann, and Heidi Kloos. 2014. "Using Complexity Metrics to Assess Silent Reading Fluency: A Cross-Sectional Study Comparing Oral and Silent Reading." *Scientific Studies of Reading* 18:235–254.

Pedersen, Niels Arne. 1994. “Nekrolog over Kaj Thaning.” *Grundtvig-Studier* 45:8–15.

Pedersen, Niels Arne. 2003. *Den teologiske grundtvigforskning siden anden verdenskrig*. Makadam Förlag.

Peng, C.-K., Sergey V. Buldyrev, Shlomo Havlin, Michael Simons, H. Eugene Stanley, and Ary L. Goldberger. 1994. “Mosaic organization of DNA nucleotides.” *Physical review e* 49:1685.

Pontoppidan Thyssen, Anders. 1983. *Grundtvigs tanker om kirke og folk indtil 1824*. Anis.

Renders, Hans, Binne de Haan, and Jonne Harmsma (eds.). 2016. *The Biographical Turn: Lives in history*. Abingdon, Oxon ; New York, NY: Routledge, 1 edition.

Rigau, Jaume, Miquel Feixas, and Mateu Sbert. 2007. “Conceptualizing Birkhoff’s Aesthetic Measure Using Shannon Entropy and Kolmogorov Complexity.” In *Computational Aesthetics*, pp. 105–112.

Riley, Michael A., Scott Bonnette, Nikita Kuznetsov, Sebastian Wallot, and Jianbo Gao. 2012. “A tutorial introduction to adaptive fractal analysis.” *Frontiers in Physiology* 3.

Rönning, Frederik. 1907. *N.F.S. Grundtvig. 1907. Et Bidrag til Skildring af Dansk Åndsliv i det 19. Århundrede, vol. 3*. Det Schønbergske Forlag.

Said, Edward. 2006. *On Late Style*. London: Bloomsbury Publishing PLC.

Scharling, C.I. 1947. *Grundtvig og Romantikken: Belyst ved Grundtvigs Forhold til Schjelling*. Gyldendal.

Schler, Jonathan, Moshe Koppel, Shlomo Argamon, and James W. Pennebaker. 2006. “Effects of age and gender on blogging.” In *AAAI spring symposium: Computational approaches to analyzing weblogs*, volume 6, pp. 199–205.

Shannon, Claude Elwood. 1948. “A mathematical theory of communication.” *ACM SIGMOBILE Mobile Computing and Communications Review* 5:3–55.

Snowdon D.A., Kemper S.J., Mortimer J.A., Greiner L.H., Wekstein D.R., and Markesbery W.R. 1996. “Linguistic ability in early life and cognitive function and alzheimer’s disease in late life: Findings from the nun study.” *JAMA* 275:528–532.

Stamatatos, Efstathios. 2009. “A survey of modern authorship attribution methods.” *Journal of the American Society for Information Science and Technology* 60:538–556.

Thaning, Kaj. 1963. *Menneske først – Grundtvigs opgør med sig selv*. Gyldendal.

Thodberg, Christian. 1983. *Grundtvig som salmedigter*. Anis.

Thoiron, Philippe. 1986. “Diversity Index and Entropy as Measures of Lexical Richness.” *Computers and the Humanities* 20:197–202.

Tyler, W. A. 2000. “Change-point analysis: A powerful new tool for detecting changes.”

Voss, Richard F. and John Clarke. 1975. “‘1/fnoise’ in music and speech.” *Nature* 258:317.

Wevers, Melvin, Kristoffer L. Nielbo, and Jianbo Gao. *submitted*. “Tracking the Consumption Junction: Temporal Dependencies in Dutch Newspaper Articles and Advertisements.” .

Zhang, Yanhui. 2016. “Entropic evolution of lexical richness of homogeneous texts over time: A dynamic complexity perspective.” *Journal of Language Modelling* 3:569–599.

Zhu, Xiaojin, Zhiting Xu, and Tushar Khot. 2009. “How creative is your writing? a linguistic creativity measure from computer science and cognitive psychology perspectives.” In *Proceedings of the workshop on computational approaches to linguistic creativity*, pp. 87–93. Association for Computational Linguistics.