Folklore Fellows' NETWORK

No. 54 | December 2020

Folklore Fellows'
NETWORK
No. 1 | 2020

FF Network is a newsletter, published twice a year, related to FF Communications. It provides information on new FFC volumes and on articles related to cultural studies by internationally recognised authors.

Publisher
Finnish Academy of Science and Letters, Helsinki

Editor
Dr. Frog
mr.frog@helsinki.fi

Editorial secretary
Petja Kauppi
secretary@folklorefellows.fi

Folklore Fellows on the internet
http://www.folklorefellows.fi
ISSN-L 0789-0249
ISSN 0789-0249 (Print)
ISSN 1798-3029 (Online)

Subscriptions
http://www.folklorefellows.fi

Contents

Cover image by Arno Niesner from Pixabay

The Corona Cocoon

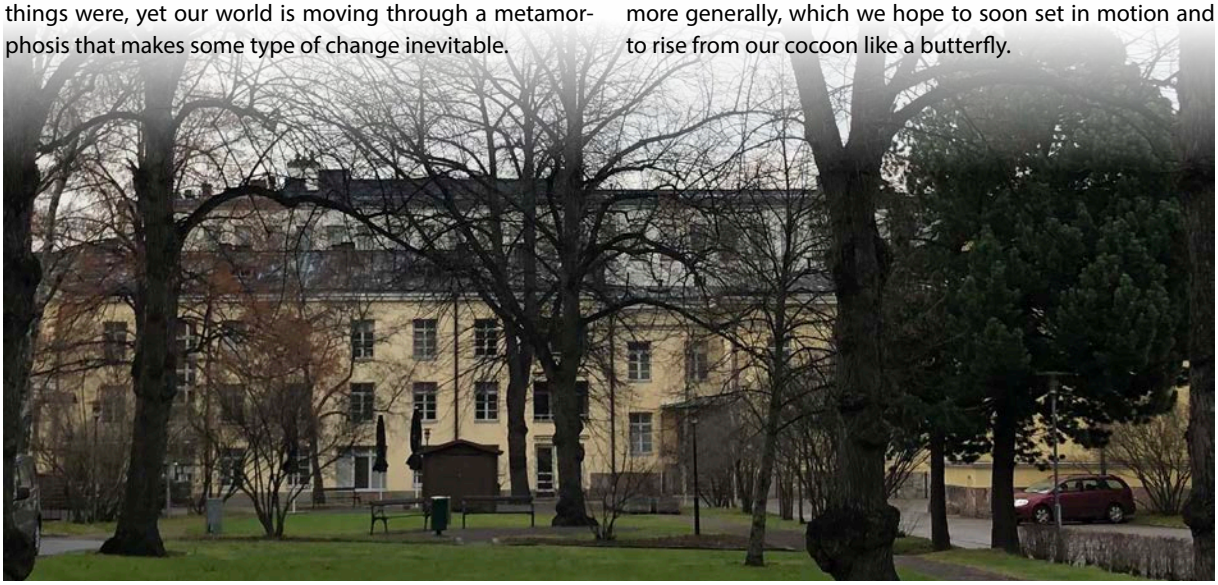
Frog

Yesterday, after hours of hopping on and off of public transportation amid various errands, I stopped by my office, thankful that my key still works while the university premises are otherwise closed owing to the Covid-19 pandemic. As I entered the abandoned courtyard, I took off my mask and virtually swooned with a breath of rich, damp, green fresh air. It is mindboggling that our world could change so abruptly this spring. International mobility is still largely suspended, safe distancing has replaced personal space and self-quarantine has become part of daily life. Institutions closed their doors and those of us who are lucky have continued work at home, while others have been left unemployed. As the physical world closed down around us, virtual worlds opened with unprecedented speed in order to compensate. The prevention of so much in our established ways of life has been a push into new technologies, bringing about an explosion of emergent practices that are gradually taking shape and perhaps also taking hold. The world of human society has transformed, placed under self-imposed constraints like a caterpillar entering a cocoon.

From the perspective of folklore, this is a fascinating time. New types of performance, customs and meaning-making are developing all around us and being negotiated in countless networks simultaneously. What will happen when we break from the Corona cocoon remains an open question: there is a dream of a return to the way things were, yet our world is moving through a metamorphosis that makes some type of change inevitable.

In tandem with the world entering a period of transformation, FF Communications and FF Network have been gestating in a cocoon of their own. Somewhat more than a year ago, the Finnish Academy of Science and Letters abruptly decided to change their publication profile and discontinue relationships with all existing series. After countless meetings and discussions, the Folklore Fellows have organized a new collaboration with the Kalevala Society as the publisher of FF Communications and FF Network beginning from January 2021. The Kalevala Society is an esteemed learned society with a century of close ties with the Folklore Fellows. Whereas so much of the future is unsure in other areas of life, we predict with confidence that FFC will continue without interruption to its processes, publications or distribution. The Kalevala Society will keep our publications within the Finnish Federation of Learned Societies (TSV) rather than moving to a for-profit publisher. This will allow FFC to remain affordable for purchase and allow us to continue plans for an open-access counterpart to print publication on TSV's new platform which assures long-term sustainability.

Change is inevitable after a long period in a cocoon, but change is not bad and may propel us into the future. For FFC, this not only includes looking forward to an open-access platform; discussions have started for new electronic resources for the Folklore Fellows and folklore research more generally, which we hope to soon set in motion and to rise from our cocoon like a butterfly.



Covid Conspiracies

A Computational Approach to Rumor and Conspiracy in a Time of Pandemic

Timothy R. Tangherlini & Vwani Roychowdhury

University of California, Berkeley and UCLA

The Covid-19 pandemic has led to an explosion of storytelling—much of it framed as believable first-person accounts—across many media. A large number of these stories reflect aspects of both rumor (which can be seen as a hyperactive transmission state of legend) and of conspiracy theory (Rosnow and Fine 1976; Tangherlini 1990; Fine et al. 2005). If we conceptualize folklore at least in part as the informal circulation of cultural expressive forms on and across social networks, thereby incorporating both the performance aspects of folklore and its dependence on social interactions embedded both in time and space, it should be of little surprise that the current situation has engendered a great deal of storytelling (Tangherlini 2018).

It is certainly well established that rumors love an information void (Shibutani 1966; Starbird et al. 2014). Given the current situation where high confidence information is hard to come by, and trusted information sources have been called into question by ideologically motivated groups, the conditions are perfect for the types of information cascades that rumors represent. Similarly, the information vacuum allows for the inventive alignment of stories, often from different domains of human interaction, into larger coherent narrative frameworks that rely on monological thinking and people's receptiveness to explanatory narratives (Goertzel 1994). We conceptualize these broad, interrelated storytelling cycles that often propose to uncover a nefarious background for an observable phenomenon such as the sudden appearance of the Covid-19 virus and its subsequent rampage across the globe as conspiracy theories (Shahsavari et al. 2020).

A great deal of interaction during this time of social distancing takes place online on social media platforms such as Facebook and Twitter, and on social media forums such as Reddit and 4chan. Unlike face-to-face interaction, social media has several features that make it particularly suitable to the propagation of rumor and the creation of conspiracy theories. First, the rapidity with which signals can propagate across these platforms is striking. Second, messages can be passed in part or in their entirety with simple clicks, while the original message can be edited quite easily. Third, there is a degree of anonymity on many of these platforms, confounding to an even greater degree the amount of trust a person may have in the source of the information. Fourth, given the assortativity of links on many social media sites (predicated on the notion of homophily in social networks), there is a tendency for these conversations to become self-reinforcing, creating what some have labeled "echo chambers" (Del Vicario et al. 2016). Fifth, there is an amplification of the signal that is uncommon in normal face-to-face interaction. This latter aspect of amplification becomes all the more critical in the context of automated agents, such as "bots", that can tirelessly send messages out into the social network again and again (Ferrara 2020).

Computational Modeling of Rumor and Conspiracy Theory

In our group's work in computational folkloristics, we aim to understand the dynamics of rumor propagation and conspiracy theory formation related to the Covid-19 pandemic in real-time. We hope to devise methods that will allow us to track how various stories gain traction in different forums, circulate, are modified and, in some cases, are dropped. We are also interested in understanding the interaction between the stories circulating in social media and news reporting about those stories. The traditional news media play an interesting role in this information eco-system, since the stories circulating in various social media forums become fodder for the reporting in two essential ways. First, the stories circulating on social media influence real world behaviors, thus leading to newsworthy events. These range from people ingesting untested remedies to people "resisting" public health measures. Second, the stories have entered the policy arena, with the president of the United States repeating rumors and conspiracy theories in press briefings, and directing government officials to explore legislative measures to address aspects of these stories. Both of these types of news stories may then feed back into the social media forums.

Our work is based on several key observations on storytelling in general, and online conversational behaviors in particular. Because computational measures require some

formalization of a problem so that they can be applied consistently, we have developed a model of storytelling inspired by Algirdas Greimas's actantial model (Greimas 1966), and the model of personal experience narrative proposed by William Labov and Joshua Waletzky (1967). To limit the scope of a story corpus, we rely on the intuition of George Boole who, in his classic definition of a discourse domain, noted that "In every discourse, whether of the mind conversing with its own thoughts, or of the individual in his intercourse with others, there is an assumed or expressed limit within which the subjects of its operation are confined" (Boole 1854). This latter aspect of our work allows us to hypothesize that, for any given domain, there should be an underlying generative storytelling framework that allows people to, once they have internalized that framework through the process(es) of enculturation, produce stories or story segments that align with the expected storytelling in that group (Tangherlini 2018). This observation aligns well with Carol Clover's now classic formulation of the idea of an "immanent narrative" underlying traditions such as epic singing, and also aligns well with fieldwork observations that people only infrequently tell complete stories (Clover 1986; Laudun 2001). This latter observation is particularly apt for social media, where jumping into a forum can feel like jumping into a conversation among strangers in a noisy bar.

The generative model for the forum posts that either recount a story in whole or in part has three main levels, offering a modification of the standard two-level model that has animated a great deal of narrative theory over the years. The goal of the model is to wed stable features of narrative to domain-specific aspects of storytelling related to a particular event or type of event, and also to the performance-specific aspects that form the basis of our ethnographic observations.[1] We labels these levels, in turn, the macro-, meso- and microlevels.

On the macro- or tradition level, the model posits that there are structural aspects of narrative that are genre-specific, albeit not genre-exclusive. For legend—and its closely related counterpart the rumor—this macrolevel consists of stable structural features recognized by Labov and Waletzky (1967), and modified by Nicolaisen (1987) for the folklore genre legend, and includes the orientation, the complicating action and the resolution. As a modification of the complicating action, we propose a two-part structure, comprising threat/disruption and strategy (Tangherlini 2018). Currently, our computational methods are not able to match actants to structural roles consistently.

Instead, our computational methods focus on the two other levels of the model: the mesolevel, which provides an understanding of the bounds of the discourse to which observed stories or story parts contribute; and the microlevel, which comprises the collected data in the form of blog posts and threads. The mesolevel provides the domain-specific anchoring of the story—the admissible actants and their range of possible actions given that domain. The microlevel, in turn, is the level that is observable, and includes all of the evaluative statements, framing elements and other aspects that make the performance of a story vibrant, even if that performance is simply a very brief conversation that alludes to a shared knowledge between interlocutors (Figure 1).

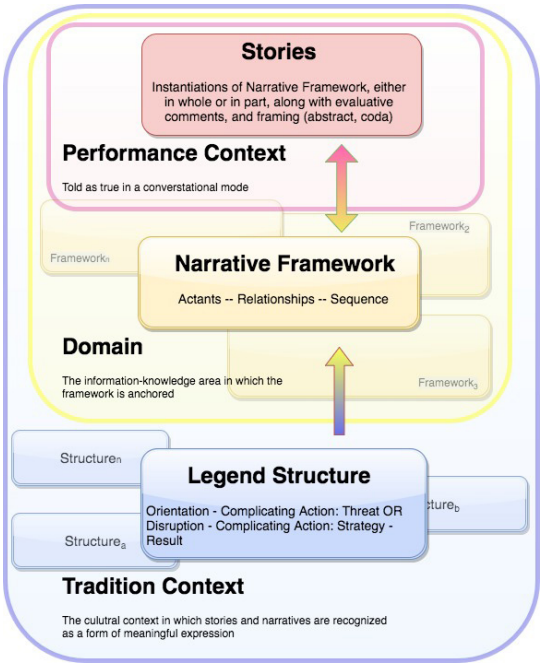


Figure 1: A three-level model of the generative processes undergirding legend and rumor (Tangherlini 2018).

Estimating Narrative Frameworks

Because of these interrelated features characterizing storytelling and story domains, we conceptualize the underlying narrative frameworks as a graph (which can be weighted and dynamically updated), where the nodes comprise the actants, and the edges comprise the interactant relationships. Although a complex hyperedge model may be more appropriate to capture complex relationships, we base this simpler model on the observation that most hyperedges can be decomposed into a series of pairwise edges. Aggregating thousands of storytelling events allows one to map the range of the admissible actants and relationships, as well as discover the emerging actant roles and interactant relationships. The more an actant is mentioned in the

1 The model is exhaustively presented in Tangherlini (2018).

storytelling, the more central they become to the network. Similarly, the more frequently an interactant relationship is mentioned, the more heavily that edge is weighted in the network. Evaluating this over time provides a dynamic view of the growth of these narrative frameworks.

The goal of our work, then, is to estimate the generative narrative framework graph for any corpus. This narrative framework identifies the main actants and presents the interactant relationships in a context-dependent manner. To illustrate the context dependent nature of these relationships, consider the Pizzagate conspiracy theory. One need only recognize that Hillary Clinton can play multiple roles in various domains of discourse to understand this context-dependent understanding of Hillary Clinton: in Democratic politics, she is a former presidential candidate; in the context of the Clinton foundation, she is a key and founding member; in the context of her family, she is the wronged spouse of Bill Clinton; and in the context of Satanic cannibalistic cults, she is a ring leader. Consequently, she, like many other actants, plays multiple roles in multiple domains, all dependent on context.

Because of the straightforward nature of this approach, we have developed a computational pipeline that, given a corpus of sufficient size, allows one to rapidly generate the narrative frameworks undergirding that corpus. As such, the pipeline provides a method for creating summary narrative graphs, long a desideratum in fields as diverse as discourse analysis, political science, sociology and folklore (Bearman et al. 2000; Lehnert 1980). The pipeline is agnostic to the data source, whether it be a corpus of social media posts, news articles, or even public policy planning documents (Mohr et al. 2013). If there is no underlying framework, the results appear as a disconnected graph. For a well-established narrative framework, by way of contrast, the results appear as a single giant connected component.

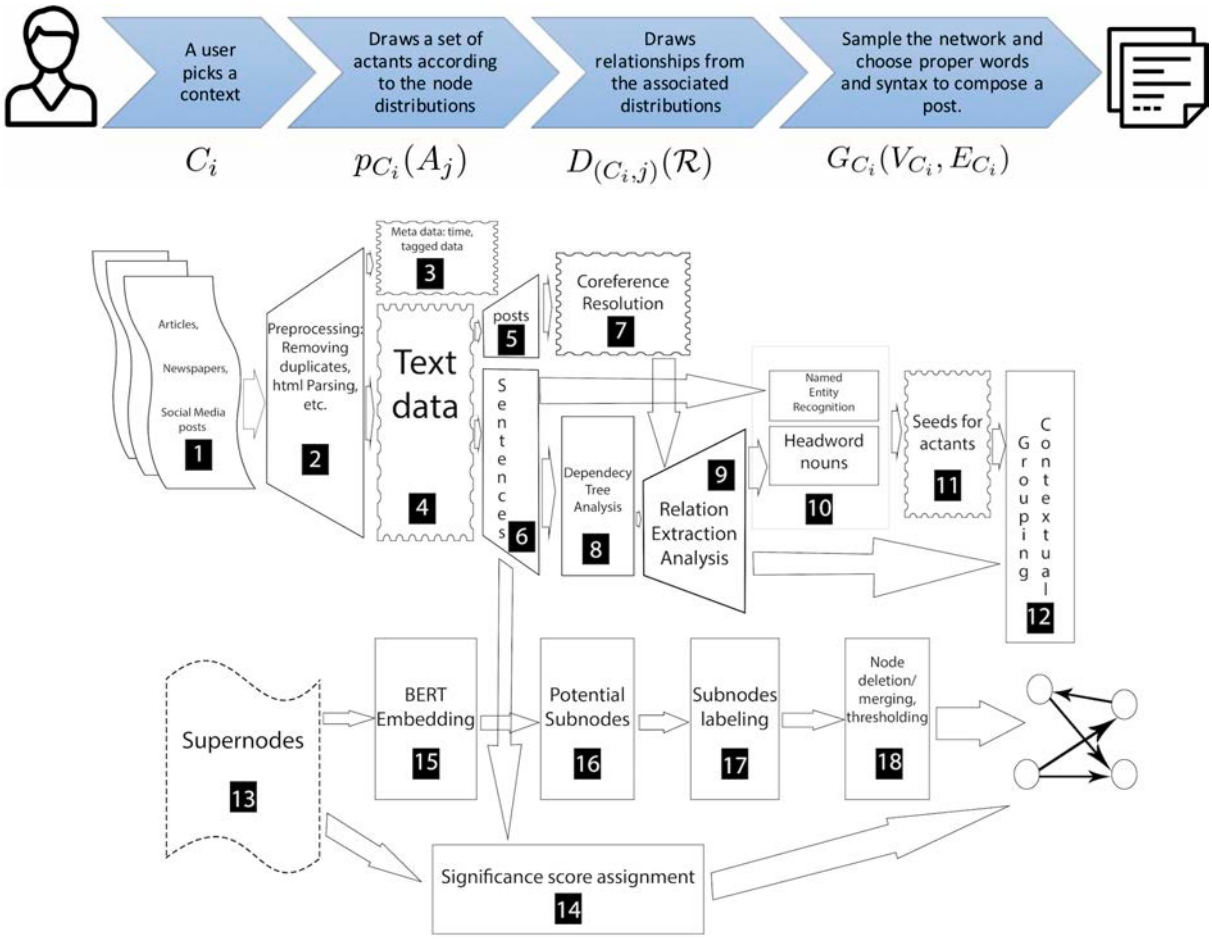


Figure 2: (a) an overview of how a person generates a post in this model (b) an overview of the pipeline (Tangherlini et al. 2020).

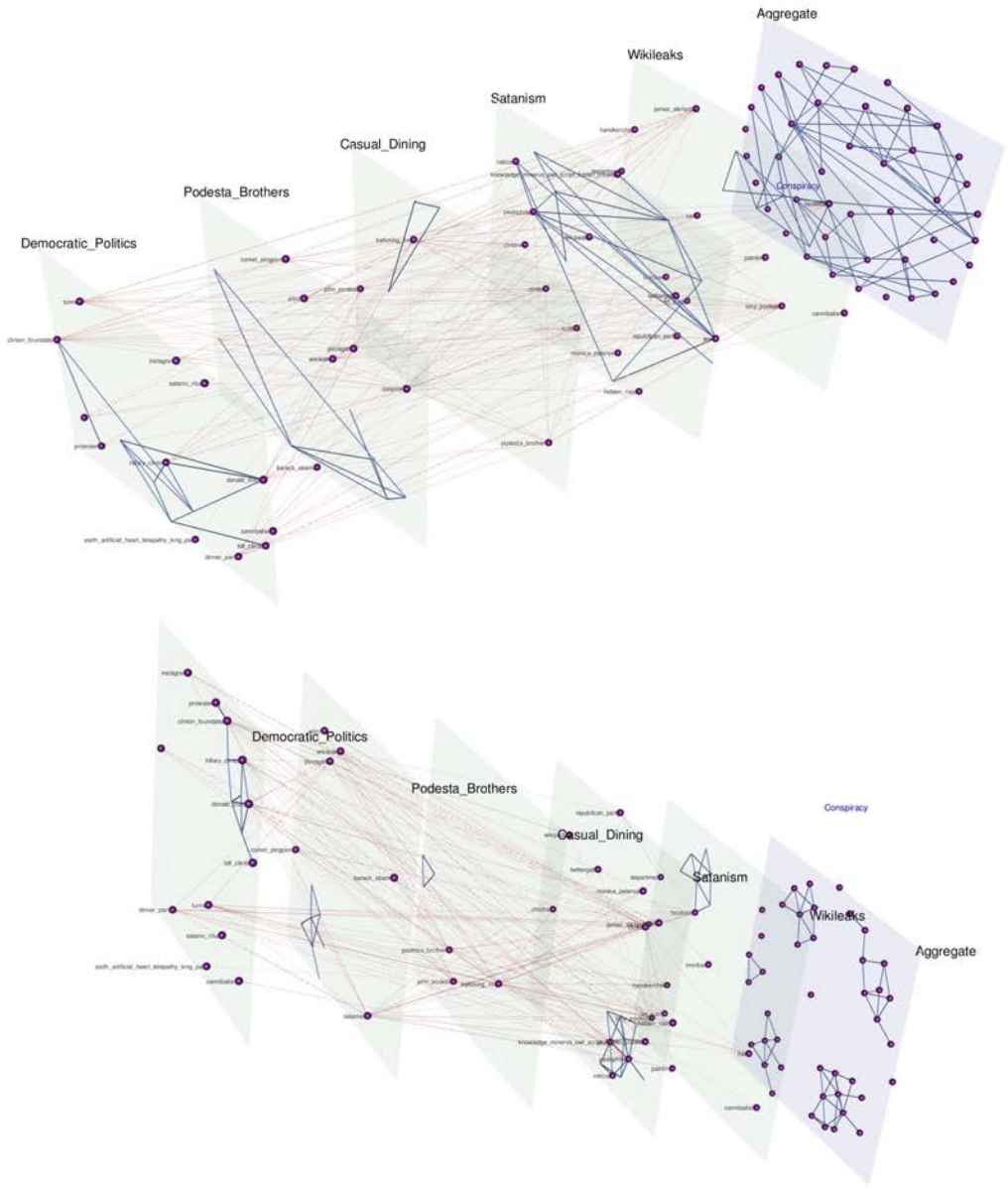


Figure 3: (a) shows the effect of Wikileaks emails in aligning otherwise unlinked domains; if the Wikileaks nodes and edges are removed from the graph, the aggregate network becomes a group of unconnected components; and (b) shows the automatically discovered main actants in an unlabeled corpus of reddit discussions related to Pizzagate. Not only do our methods match the graph presented in the NY Times (purple edges), but we also discover the importance of Bill Clinton and the Clinton Foundation to the conspiracy theory as presented by the main "theorists" themselves. (Tangherlini et al. 2020)

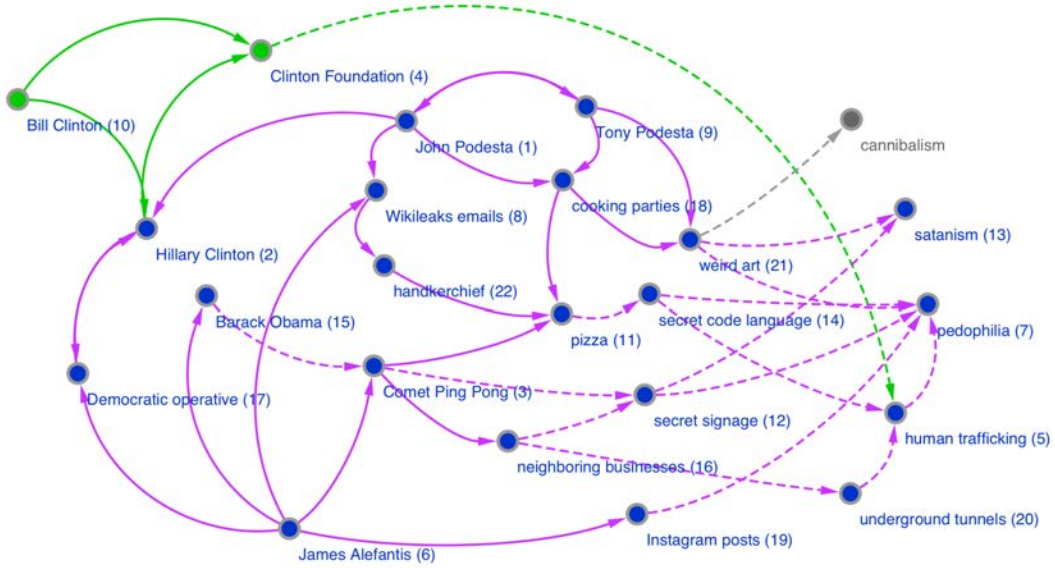


Figure 3: (b)

Applications of the Method

The results of applying this approach to a series of tens of thousands of posts on social media concerning the Pizzagate conspiracy theory—a theory that alleged a relationship between Clinton, various Democratic party operatives, Satanic pedophilic child trafficking and a northern Washington DC pizza restaurant, were fairly striking. For instance, our top-level graph captured the overall contours of the conspiracy theory, while also highlighting the reliance of the conspiracy theorists on the trove of leaked Wikileaks emails hacked from the Democratic National Committee as a means for linking the otherwise unrelated domains of Democratic politics, casual dining, DC area businesses and Satanic child sex trafficking (see Figure 3). The study also reveals how conspiracy theories often are created through the alignment of multiple, otherwise unaligned domains of interaction. This alignment is only possible because the theorists either collectively or individually have access (or gain access) to special, often hidden or secret knowledge, which they can interpret given their own (implicit) skill at understanding this knowledge. In the case of Pizzagate, it was the trove of Wikileaks emails that provided this key, while in the case of Q-Anon, for instance, it is the "bread crumbs" left by the eponymous Q on various messaging boards.

The limits of visualization software often make it difficult to present the rich semantic labeling that characterizes the relationships between actants. In our system, we are devising methods to provide this rich labeling in smaller or "zoomed-in" views of the graph (Figure 4).

Narrative Frameworks and the Emergence of Conspiracy Theories in Social Media

In the context of Covid-19, we have focused on two corpuses: one derived from Reddit and 4chan, and one derived from a series of news reports focused on conspiracy theories. Unlike the Pizzagate conspiracy theory, where Wikileaks is used by the conspiracy theorists as a Rosetta stone and thus allows for a single dominant narrative framework to move to a position of prominence, the pandemic discussions have yet to settle on a main narrative framework.

Instead, in our pipeline discovery, we find a series of emerging narrative frameworks, each of which proposes a different explanatory theory. Although several of the conspiracy theories are based on existing and long-standing conspiracy theories, others, such as the suggestion that the virus is activated by microwaves emanating from the new 5G network, are novel. A particularly prominent one, and one that has encouraged people to take real-world action, proposes that the virus is a hoax, no worse than the common cold, and that the hospitals are not in any dire situation; this has led to the "film your hospital" movement, which has been compared by analysts to the truther campaigns that suggested that President Obama's birth certificate was fake (Sommer 2020).

The interaction with the news is equally intriguing. While the news often "chases" after the formation of conspiracy theories, as was clearly the case in the well-documented Pizzagate conspiracy theory, the news and social media appear to be in a cycle of mutual reinforcement for the Covid-19 pandemic, where conspiracy theories are

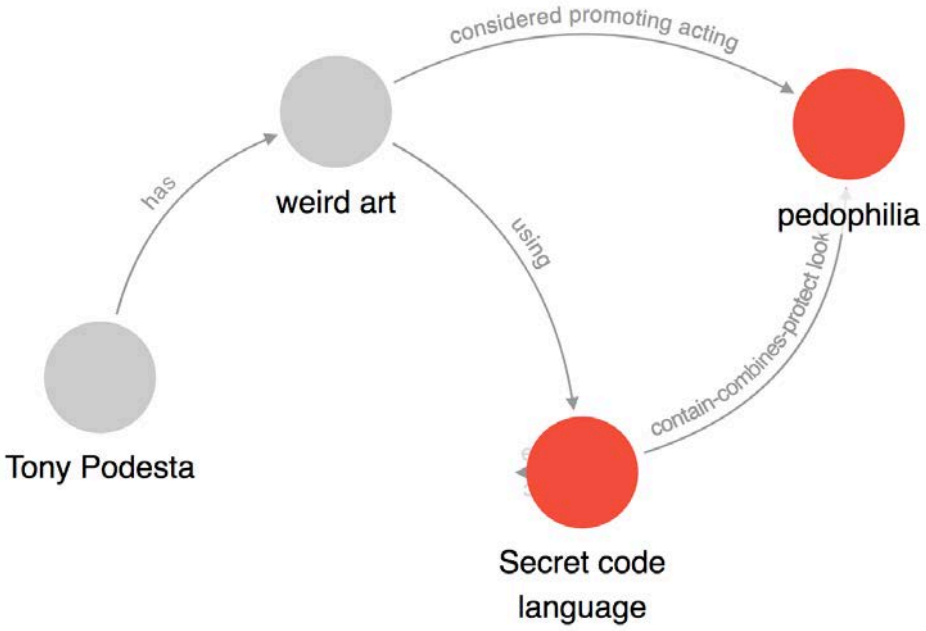


Figure 4: A zoomed in view of the relationship between Tony Podesta, weird art and pedophilia in the Pizzagate graph. The relationship edges are automatically generated by the pipeline and consist of the highest ranked relationship labels for a particular edge.

picked up by the news and then feed back into the social forums, where they take advantage of mentions in the news as validation of the narrative itself.

In the news reporting that we considered, we discovered that the general mentions of conspiracy theories, or the specific conspiracy theories themselves, lagged considerably behind the emergence of these in social media. In January, for instance, conspiracy theory reporting focused almost entirely on Q-anon and other non-Covid-19-related conspiracy theories (Figure 6a). Conspiracy theories began to be reported on more closely with Covid-19 in mid-March (Figure 6b) and, by mid-April, reporting on conspiracy theories was very closely linked to Covid-19 reporting (Figure 6c) (Shahsavari et al. 2020).

We expect to witness several phenomena as the Covid-19 pandemic continues to develop. If theories about monologic belief systems and their role in conspiratorial thinking are correct, then several of the conspiracy theories should align to form a single, totalizing narrative framework (Goertzel 1994). It appears that this may already be happening in the context of the 5G narrative, the bioweapons narrative and the globalist cabal narrative. A single conspiracy theory may then emerge as the dominant narrative about the pandemic, while smaller frameworks are either abandoned or break off to form alternative explanatory theories. Another phenomenon that may emerge is the lack of coordination across these narratives, with a constant churn in the discussion space. New narrative framework nucleations will appear in these discussion forums, gain some momentum, and then be abandoned. Certainly, these developments will be worth watching. Our methods can support this type of coarse level of surveillance, while providing early clues to guide more directed analysis.

Conclusion

Although observation and modeling can be important as we develop a more sophisticated understanding of how storytelling develops and functions during a global crisis, it may be equally important to apply our knowledge of these processes to help efforts to minimize the risk posed by people taking real-world action motivated by these explanatory narratives. Our methods, for instance, could be used to track the emergence of potentially dangerous information, such as the efficacy of household cleaning products as medicines to combat the virus, that form part of widely-accepted narratives on social media. Interestingly, these narratives about using household goods to combat Covid-19 *predate* Trump's misinformed suggestion that cleaning products are the subject of scientific testing for people suffering from the

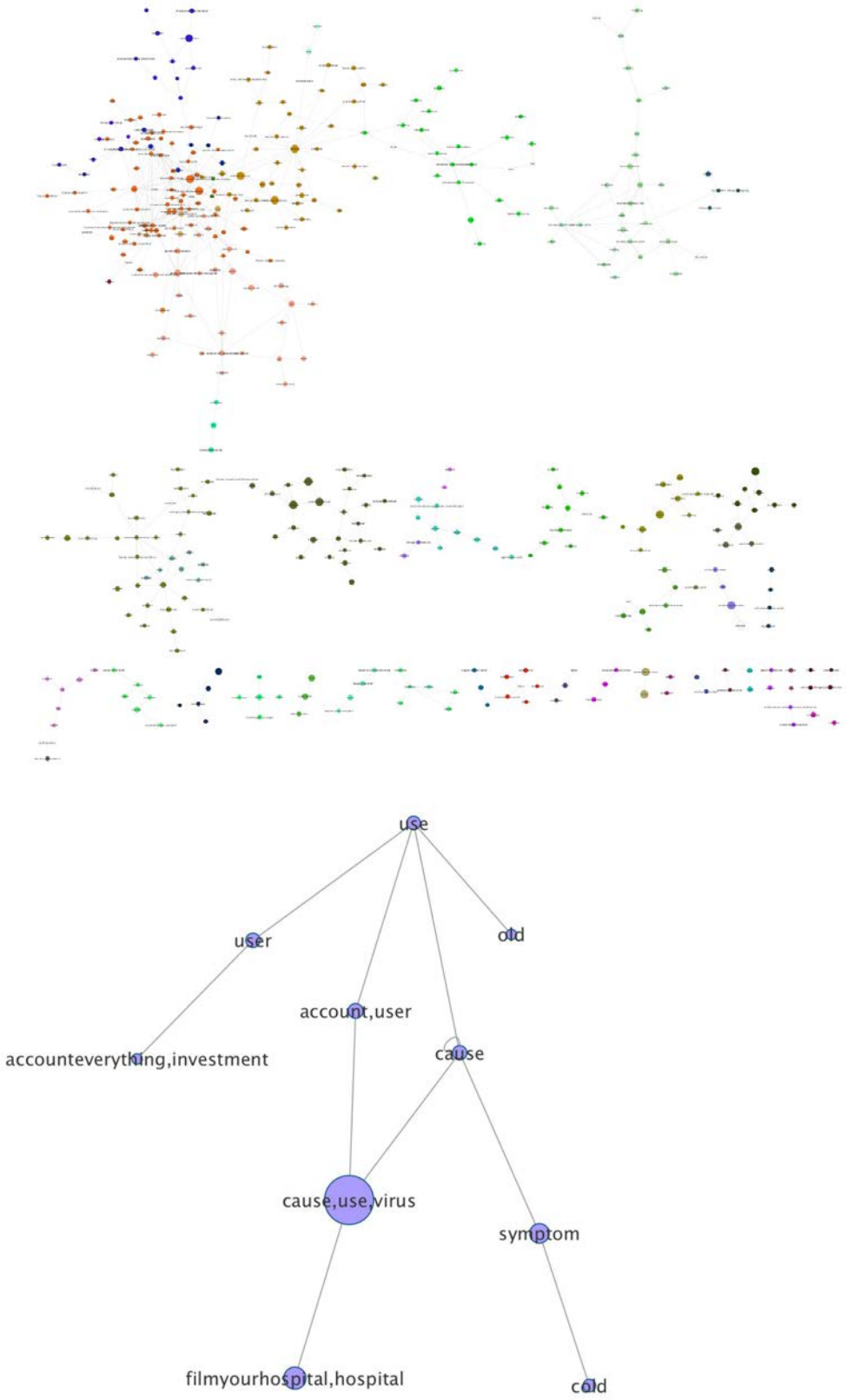
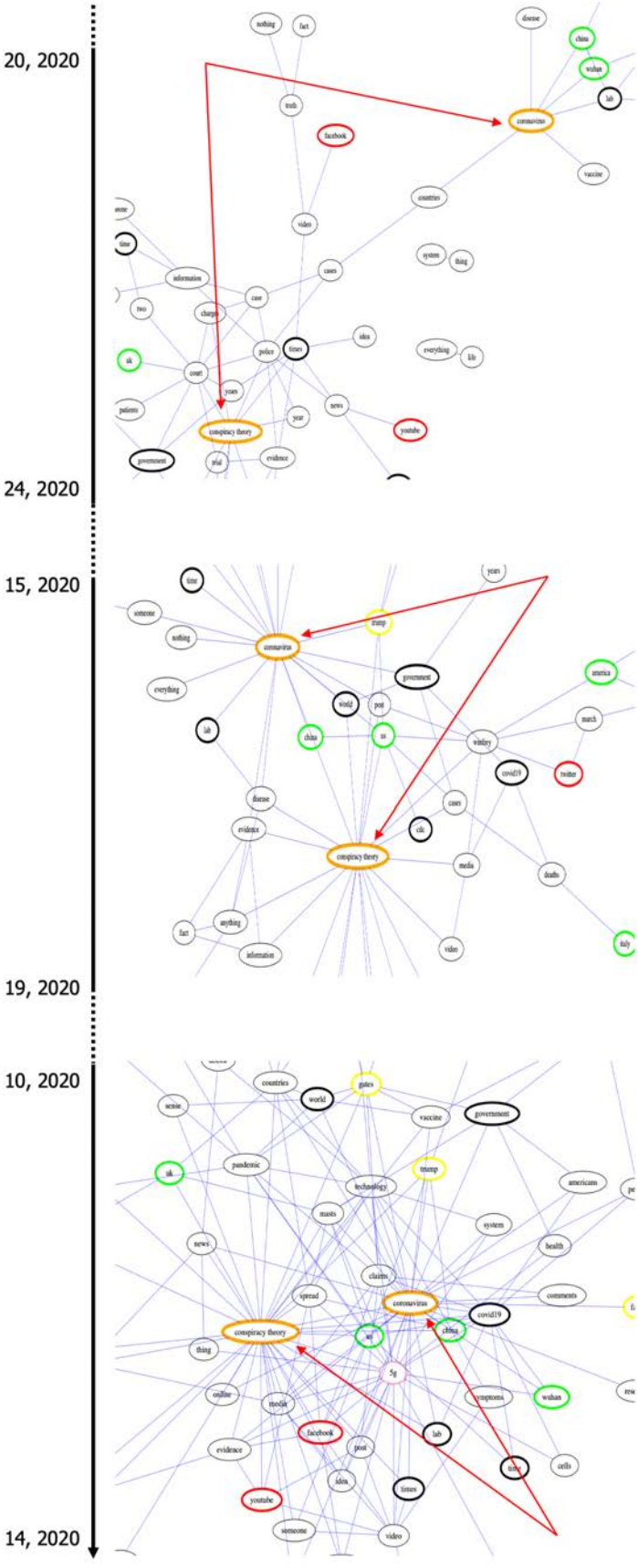


Figure 5: (a, above) An overview of the 52 different narrative framework "communities" in the overall social media discussion forum space. (b, below) A close up of a narrative framework, linking nodes from a series of communities, proposing that the Covid-19 virus is a hoax, and urging people to "film your hospital."



virus. Similarly, given the generative nature of our model, the approach may be helpful for public health officials who need to rapidly produce messaging that can disrupt the spread of information that has the potential for broad negative social impact, such as stories insisting that the virus is essentially harmless (which have led people to defy stay-at-home orders).

There can be little doubt that the current pandemic is an opportunity for folklorists to apply our hard-earned knowledge of the complex mechanisms that underlie storytelling in and across human societies. Among our many goals as scholars should be to document the ubiquitous storytelling that has rushed to fill the information vacuum, as well as to assist in developing methods that allow us to track these developments across many different media in a consistent and reproducible manner, allowing us to use this knowledge for the broader public good.

Figure 6: A time series overview of the reporting on conspiracy theories in news harvested from Google's GDELT project. The timeline illustrates the increased reporting on conspiracy theories starting in late January and the progressively closer relationship between conspiracy theory reporting and Covid-19 reporting.

Works Cited

Allport, Gordon W & Leo Postman. 1947. *The Psychology of Rumor*. New York: H Holt and Co.

Bearman, Peter S & Katherine Stovel. 2000. "Becoming a Nazi: A Model for Narrative Networks." *Poetics* 27(2-3): 69-90.

Boole George. 1854. *An Investigation of the Laws of Thought: On which are Founded the Mathematical Theories of Logic and Probabilities*. London: Walton and Maberly.

Clover, Carol J. 1986. "The long prose form." *Arkiv for nordisk filologi* 101: 10-39.

Del Vicario, Michela, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli & Walter Quattrociocchi. 2016. "Echo Chambers: Emotional Contagion and Group Polarization on Facebook." *Scientific Reports* 6: 37825.

Ferrara, Emilio. 2020. "# COVID-19 on Twitter: Bots, Conspiracies, and Social Media Activism." arXiv preprint arXiv:2004.09531

Fine, Gary Alan, Véronique Campion-Vincent & Chip Heath. 2005. *Rumor Mills: The Social Impact of Rumor and Legend.* New Brunswick: Aldine Transations.

Goertzel, Ted. 1994. "Belief in Conspiracy Theories." *Political Psychology* 15: 731-742.

Greimas, Algirdas Julien. 1966. "Éléments pour une théorie de l'interprétation du récit mythique." *Communications* 8(1): 28-59.

Labov, W. Waletsky & Joshua Waletzky. 1967. "Narrative Analysis: Oral Versions of Personal Experience." In, *Essays on the Verbal and Visual Arts*. Edited by June Helm. Seattle: University of Washington Press. Pp. 12-44.

Laudun John. 2001. "Talk about the Past in a Midwestern Town: It Was There at that Time." *Midwestern Folklore* 27(2):41-54.

Lehnert, Wendy G. 1980. "Narrative Text Summarization." In *AAAI-80 Procedings*. Palo Alto: Association for the Advancement of Artificial Intelligence. Pp 337-339.

Mohr, John W, Robin Wagner-Pacifici, Ronald L. Breiger & Petko Bogdanov. 2013. "Graphing the Grammar of Motives in National Security Strategies: Cultural Interpretation, Automated Text Analysis and the Drama of Global Politics." *Poetics* 41(6): 670-700.

Nicolaisen, Wilhelm FH. 1987. "The Linguistic Structure of Legends." *Perspectives on Contemporary Legend* 2(1): 61-67.

Rosnow, Ralph L & Gary A Fine. 1976. *Rumor and Gossip: The Social Psychology of Hearsay*. New York: Elsevier.

Shahsavari, Shadi, Pavan Holur, Timothy R. Tangherlini & Vwani Roychowdhury. 2020. "Conspiracy in the Time of Corona: Automatic detection of Covid-19 Conspiracy Theories in Social Media and the News." arXiv preprint arXiv:2004.13783

Shibutani, Tamotsu. 1966. *Improvised News: A Sociological Study of Rumor*. Indianapolis: Bobbs-Merrill.

Sommer, Will. 2020. "Naturally, We Now Have a Cottage Industry of Coronavirus Truther Assholes." *The Daily Beast*, March 30, 2020.

Starbird, Kate, Jim Maddock, Mania Orand, Peg Achterman & Robert M Mason. 2014. "Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing." In, *IConference 2014 Proceedings*. Pp 654-662. Grandville, MI: iSchools.

Tangherlini, Timothy R. 2018. "Toward a Generative Model of Legend: Pizzas, Bridges, Vaccines, and Witches." *Humanities* 7(1): 1. doi.org/10.3390/h7010001

Tangherlini, Timothy R. 1990. "'It Happened Not Too Far from Here...': A Survey of Legend Theory and Characterization." *Western Folklore* 49(4): 371-390.

Tangherlini, Timothy R., Shadi Shahsavari, Behnam Shahbazi, Ehsan Ebrahimzadeh, Vwani Roychowdhury. 2020. "An Automated Pipeline for the Discovery of Conspiracy and Conspiracy Theory Narrative Frameworks: Bridgegate, Pizzagate and Storytelling on the Web. *Under review*.

Historical Oral Poems and Digital Humanities

Starting with a Finnish Corpus

Kati Kallio
Finnish Literature Society and University of Helsinki

Eetu Mäkelä
University of Helsinki Centre for Digital Humanities

Maciej Janicki
University of Helsinki Centre for Digital Humanities

In this essay, we describe early experiments in a computational folkloristics project FILTER[1] aimed at studying formulaic intertextuality, thematic networks and poetic variation across regional cultures of Finnic oral poetry. Due to the vast amount of linguistic and poetic variation and historical biases in the corpora (see e.g. Anttonen 2005; Harvilahti 2013; Tarkka et al. 2018; Ilyefalvi 2018; Mäkelä et al. 2020b), existing automated approaches (see e.g. Moretti 2013) are unusable. Instead, advances must be made through intelligently interleaving computational and manual analysis (Säily et al. 2018; Hämäläinen et al. 2018; Isoaho et al. 2020).

In this project, the idea is to gradually develop tools in tight collaboration between folklorists and computer scientists (Mäkelä et al. 2019; 2020a). The folklorists describe what they tend to do and what they dream of being able to do with the source material, while computer scientists think of what may be possible and how this might be achieved. We first discuss the ideas, proceed to some test computations and then interpret these – and the possible problems – in relation to our humanistic and computational background knowledge of the data itself. If the results seem promising, some prototype interface may be developed, and the folklorists begin experimenting with it, evaluating what does or does not work, and describing what they do so that the computational scientists are able to understand the humanistic needs and the interpretive problems in the data. Folklorists continue dreaming what they would like to do, potentially leading again to new computational solutions and new evaluations in the cycle. In such experiments, even those that are only briefly tried often reveal new aspects of the data and help us to understand it better.

While we aim to build tools and processes that serve our specific project, we are also making them as broadly applicable as possible for researchers working with the same corpus or with similar questions with other materials, particularly for other small languages and oral-derived corpora. On the side of folkloristics, the project builds on the long research history of Finnic oral poems, on advances in computational folkloristics (see e.g. Abello et al. 2012; Arvidson et al. 2018; Harvilahti 2019; Hakamies et al. 2019; Sarv 2019; Tangherlini 2013; 2016) and on discussions with colleagues, especially Frog, Lauri Harvilahti, Janika Oras, Jukka Saarinen, Venla Sykäri and Senni Timonen.

In this essay, we describe our early experiments thus far. At this stage, the main computational question has been how to help the humanist researcher to find relevant sub-corpora or sets of texts, how to tackle complex textual variation, and what tools might be used to find similar, yet varying instantiations of verses and motifs. The central questions have been: (*a*) how to define folkloristically relevant research questions that are narrow enough for the development of new tools and yet help to produce and test tools with potential for wider use; and (*b*) how to analyse and explain the quite complex and versatile processes of reading, contextualising and analysis that folklorists tend to do with historical poetic texts, so that the computational scholars can help to make these processes easier.

Finnic Oral Poetry and the SKVR Corpus

Historical Finnic oral poetry – *runo*-songs, *regilaul*, or Kalevalaic poetry – makes a versatile corpus in multiple dialects and archaic forms of Estonian, Finnish, Karelian, Ingrian (Izhorian) and Votic languages. All in all, there are over 240,000 digitized texts of Finnic tetrametric oral poetry in the Finnish Literature Society and Estonian Literary Museum, and more archival texts and sound recordings in other Finnish, Estonian and Russian archives. (Harvilahti

1 Academy of Finland no. 333138, 308381, 322071 and 288119

2013; Sarv & Oras 2020; Kallio et al. 2017). In this preliminary work, we've focused on the Finnish SKVR corpus of 89,247 items in Karelian, Ingrian and Finnish languages, but we are currently working to add the Estonian corpus (see Sarv & Oras 2020), the unpublished (but digitized) Finnish corpus and some 19th-century literary works in Kalevala-meter.

The poems in SKVR were recorded from 1564 to 1939 and were originally edited and published in the 34 volumes of *Suomen Kansan Vanhat Runot* (SKVR) 'The Ancient Poems of Finnish People' (1908–1948 and 1997). The corpus is biased, for example, towards epic, narrative and poetically coherent texts (see e.g. Anttonen 2005; Tarkka 2013; Kalkun 2015; Tarkka et al. 2018; Timonen 2004), but it contains a wide variety of poetics and genres from epics and lyrical songs to incantations, ritual songs and lullabies (e.g. Harvilahti 2013; Kallio et al. 2017; Tarkka 2013).

Although not created for contemporary research questions, the corpus is unique in the scope of its documentation of local, historical Finnic oral traditions. Nevertheless, the sheer size of the data, the complex historical

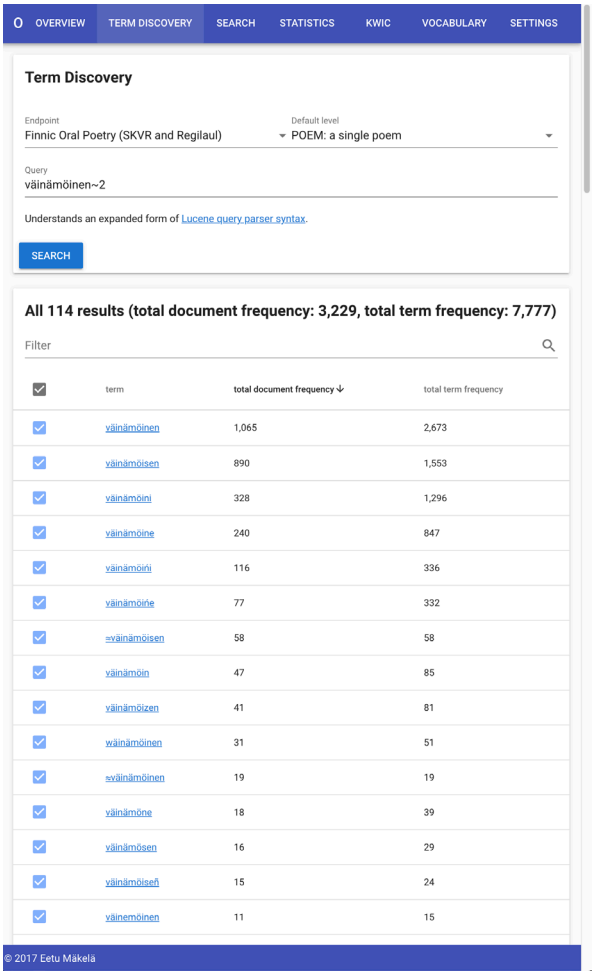


Figure 1. First results on Octavo for one set of variations for *vanha Väinämöinen* 'Old Väinämöinen', with metadata on collector, theme ID of the type index and the place of recording, and one sentence of text around each occurrence.

Figure 2. Term discovery search on Octavo for one set of variations of *Väinämöinen*, the results showing how many times each variation appears in the SKVR corpus.

and contextual knowledge needed in interpreting it, and the ample linguistic and poetic variation of texts make aims for macroscopic views (see Tangherlini 2013; 2016) difficult. The texts make use of diverse dialectal, morphological, poetic and archaic wordings, written down with various orthographies. Some folklore collectors used standard literary language, while others applied detailed phonetic transcription. Furthermore, motifs and storylines were used in versatile ways related to local understandings of poetics, genres and performance situations. (See e.g. Harvilahti 1992; Frog 2010; Timonen 2004; Saarlo 2005; Tarkka 2013; Kallio & Mäkelä 2019.) The multilevel variation and uneven quality of the data poses challenges for any computational experiments.

In the SKVR corpus, the metadata is structured, which offers possibilities for various analyses and visualisations according to the recorder of the text and the place and time of documentation. In addition, the corpus contains a typological index. Yet, the metadata also presents some problems. Although research interests today tend to concern people and society, these are not represented well in the metadata. Some dates and places of documentation are incorrect or unknown, or only vaguely identified with a region or century. The performers of the songs often remain unidentified. For the most part, the nineteenth century collectors did not think that information about informants was relevant, and many singers also preferred to remain anonymous. In the typological index, the main etic genres – like narrative poems, lyric poems, incantations, wedding songs or children's songs – have been analysed according to slightly different principles. For some genres, the index mostly reproduces those used in the printed SKVR, which in many cases were developed by the editors of the particular volumes and never unified; for others, especially lyric songs, the types are the product of recent, detailed analytical work. (See https://skvr.fi/skvr-runohakemisto.) In addition, a significant amount of essential information about the data is only found in the manuscripts, footnotes of earlier research, and prefaces of SKVR's printed volumes.

How to Browse the Complex Corpus?

A basic need for almost any user of a corpus of texts is to be able to find individual texts – whether a particular text, comprehensive corpus or some representative examples – on the basis of some criteria, such as a certain word, formula, line, motif or poetic type, or metadata such as year, place, collector or archival signum. Small differences in the functionalities of user interfaces can thus significantly impact on what kinds of research actions are feasible. The functionalities determine the flexibility of the interface, how easy it is to move between the list of results and individual texts, how the hits are indicated and whether it possible to sort the results. In the current online SKVR database (www.skvr.fi), there are several problems for advanced use: the hits within texts are not indicated, the user cannot arrange the results by the metadata, and the possibilities for free text searches are limited (see https://skvr.fi/ohje).

In our preliminary work, the SKVR poems were loaded into the Octavo system. The Octavo system is a service Eetu Mäkelä has developed to support humanities and social science research based on combinations of large, varied and 'noisy' text corpora along with attendant metadata. The system has been developed in collaboration with multiple humanities and social science research projects. On that background, the aim has been to transcend individual datasets and questions to provide functionalities of broader relevance, while at the same time ensuring that the functionalities are able to help answer actual research questions in individual projects.

The core of the Octavo system is its rich functionalities for delineating a subset of interest out of originally large and varied datasets. These include multiple mechanisms for dealing with different types of variation in the textual content, as well as the capability to query both metadata and content at the same time. After delineating a subset of interest, the system then offers further functionalities for both close reading (as seen in Figure 1) as well as subjecting results to statistical analysis, both in terms of metadata as well as vocabulary. Further, the system has been particularly designed to support iterative workflows, where the researcher can easily experiment with and amend their query constraints in response to the results they get and the analyses they make. In addition, some result views (Figure 2) are explicitly designed to help discover new variant forms for the query terms. Due to this, a researcher can start with the most obvious and certain query forms, but through iterative improvement ensure that they are also capturing the totality of the textual phenomenon of interest, while at the same time filtering out what does not belong to it.

For the most common cases across the various humanities and social science projects, the system provides ready web-user interfaces. However, feeding these are more expressive open programmatic interfaces. Due to this, the system is able to provide its most important workflows easily for all to use, but at the same time it does not limit more tech-savvy users from amending and modifying the workflows to better suit their exact needs.

Thus far out of Octavo's functionalities, the present project has mostly used the interfaces aimed at overcoming textual variation, as well as close reading of the query results. A typical search process proceeds as a chain of different types of searches. The researcher may check the variation of some individual words (Väinä*; Väinämöinen~2) and formulas ("va* van*"~1), limit the obtained results using word forms or metadata (-vanga*; -themeID:605002230), arrange the results on the basis of metadata, take a look at only the searched verses or formulas or at longer sequences

of poems, look at the whole texts either in Octavo or the SKVR database, make similar searches on parallel verses to look for unnoticed variations of the first verse, or use the type index to find similar texts without the textual feature that has been searched for or to see how these relate to the earlier analyses. (See Kallio & Mäkelä 2019).

These kinds of search processes reveal that e.g. the name of the old sage Väinämöinen may occur in over 200 forms, including *Väinö*, *Väinämö*, *Väilämöinen*, *Viänämöinen*, *Vainämoinen*, *Wäinämöisen*, *Väinämöizen*, *Väinämyösen*, and *Väinämöinji* – of which *Väinämöinen* is the most popular with 1,017 occurrences – and with numerous inflections such as *Väinämöistä*, *Väinämöisten*, *Väinämöistennin*, *Väinämöinä* etc., sometimes added with various diacritics. In formulas and poetic lines, this kind of variation accumulates. *Väinämöinen* most often appears in the formula *vaka vanha Väinämöinen* 'steady old Väinämöinen'. Yet, he can be wise instead of steady, or the formula may get shorter to incorporate verbs or other words, such as, for example:

Tuop oli vanha Väinämöinen
that was old Väinämöinen

Tuopa viisas Väinämöinen
that wise Väinämöinen

Olipa ennen vanha Väinö
there once was old Väinö

Sano vanha Väinämöinen
said old Väinämöinen

Päälle polven Väinämöisen
onto the knee of Väinämöinen

The formula often has a parallel line *tietäjä iänikuinen* 'the eternal sage', which again may have inflections and variations or be replaced with other parallel formulas. Yet, if compared with some short, wide-spread sequences of formulas (*standard sequences* or *multiforms*, see Harvilahti 1992; Frog 2016), such as the ones on making a journey, the set of formulas on *Väinämöinen* is quite simple, narrow and stable (Kallio & Mäkelä 2019).

When mapping and understanding of this kind of variation is done, and various exceptions and special cases have been interpreted, the researcher has a sub-corpus to proceed with, for example, when analysing various uses of a particular formula, motif or poetic type, or the relation of these to different local or genre-specific practices, literary influences or other features.

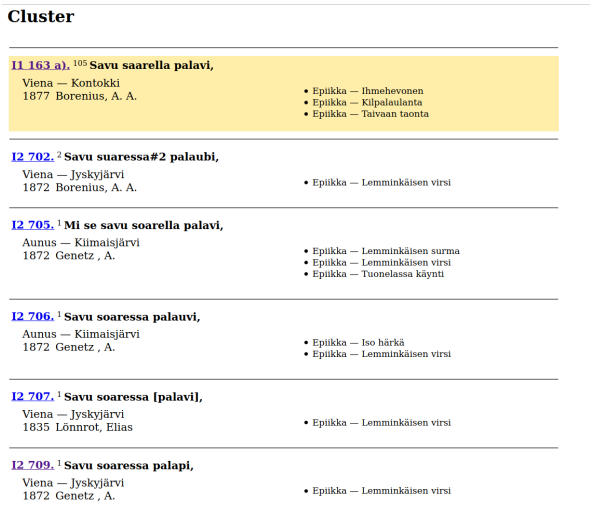


Figure 3. Part of the cluster of the verse 'Savu soarella palaabi' ("Fire is burning on the isle").

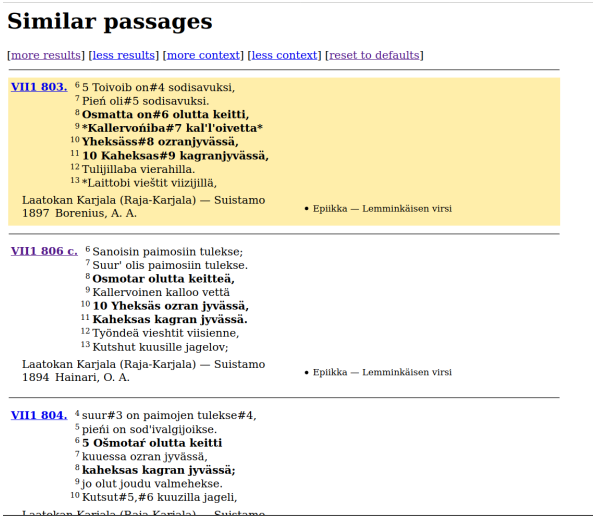


Figure 4. Search for passages similar to "Osmatta on#6 olutta keitti, *Kallervońiba#7 kal'l'oivetta*, Yheksäss#8 ozranjyvässä, 10 Kaheksas#9 kagranjyvässä" ('Osmatta brewed beer, Kalervoini (brewed) malt-water, in nine grain of barley, in nine grain of oat').

Yht' ei kut°tsun Lemmingäistä.°	15 Kut°tšu veri-sogeat,°
*Rujot (ne) reillä reissuaabi,	Ruiot re'ellä rembutteli,
Rammat rat°tšahin ajeli,°	Rammat rat°tšahin ajeli,°
	Sogeat venosin souti,
20 Sogiat venozin soudi.*	Yht' ei kut°tšu Lemmingästä.°
Lemmingäin on poiga_piilo	20_Lemmingäne on piilopoiga
Pillojah on piilemässä,#10	Pilloja on piilemässä,
Pahojah pagenemassa.#11	Pahoja pagenemassa.
"Hoib om moamo, kandajańi,	"Hoi on moammoni, kantajani,
25_Armas maijon andajańi,	Armas maion antajani,
Ihalan imettäjäni,	25_Ihala imettäjäni,
Et°tsib om miul pelvoi paid[a],°	Tuos miull' sot'isobani,
Ennembä neidona kuvottu,	Kannas paloini-paita!"
Kassabapeän#12 on kalkuteltu,	Emo varsin vastajeli:
30 Kannabas paloni paid[a]."	Noin on#2 virkki, näin pagiši:

Figure 5. The side-by-side view of two automatically aligned versions of *The Song of Lemminkäinen*.

From Similarity of Character Bigrams to Verses, Sections and Poems

Octavo, by design, allows the user fine control in driving their discovery and exploration. However, this requires an expert user who is able and willing to put in the often significant time required to craft queries in its language, to understand its affordances and limitations and to manually keep tabs on their exploration process. Consequently, the results are still substantially dependent on the competence of the user on the variation and complications of the corpus. Thus, we are actively searching for means to make the interaction easier. Particularly, we are looking at ways to use the corpus itself to iteratively drive the search.

To this end, Maciej Janicki has started developing a prototype tool for exploring the similarity within the corpus on verse, passage and poem level. The main computational idea is to measure the similarity between individual verses as the cosine similarity on character bigrams. Roughly speaking, this amounts to how many pairs of adjacent letters the two verses have in common. For example, the verses *Armazb maijon andajańi* and *Armas maion antajani*, despite having differences in every word, have many common bigrams: "Ar", "rm", "ma" twice, "ai", "an", "aj" etc. Importantly, besides allowing for orthographic, morphological and dialectal variation, this similarity metric is also insensitive to word order and only weakly sensitive to word compounding.

After discovering the most similar pairs of verses based on bigram analysis, the verses are clustered using the Chinese Whispers algorithm (Biemann 2006), which results in groups of verses similar to each other. The Chinese Whispers algorithm starts by assigning each verse to their own group. Then, it proceeds by selecting a verse, and going through every other verse it is pairwise similar to. From the clusters that these other verses belong to, it finds the one that contains most similar verses overall to the one under evaluation, and moves the verse to that group. This is done in random order for all verses, and further repeated until no group changes occur anymore. In a network representation of the corpus, with verses being nodes and similarities between verses edges, the Chinese Whispers algorithm computes groups of nodes that are especially densely connected with each other, as compared to the rest of the network. The resulting groups of similar verses can be used to explore how a given type of verse or sequence of verses appears in the corpus regardless of surface-level variation (Figures 3 and 4).

To align two poems, the minimum edit distance algorithm (Wagner & Fischer 1974) is used. The algorithm aligns the verses between the poems in a way that maximizes the poems' overall similarity (i.e. the sum of verse-wise similarities). The same algorithm can be applied to align the paired verses themselves at the character level. The result is a side-by-side view of two poems (Figure 5), in which both the differences on the verse level (equivalent vs. non-equivalent parts) and on the character level within equivalent verses are highlighted.

The main drawback of the current approach is its inability to capture and visualize changes in verse ordering (see *Yht' ei kuttsun Lemmingäistä* in Figure 5) or to explore similarities below the verse level. Also, the bigram-based similarity metric underestimates the similarity in cases of many small phonetic differences and could be improved by taking the phonetic similarity into account (e.g. substituting a vowel with a different vowel is a much smaller difference than with a consonant). We are going to address these points in further work.

Our future idea is to test the coverage of recognising similarity by comparing the results of the interface with more manual search results on Octavo, and on the existing type index and earlier manual studies on certain poetic types. Further, it is quite essential to add possibilities for manual adjustments – what verses are most relevant, what kinds of features should count as similar or should be highlighted in comparison – and think of effective ways to visualise and interpret the similarities of large groups of verses, sections or texts. For example, Stefan Jänicke and David Joseph Wrisley (2017) visualise versions of *Chanson de Roland* in a way that helps even someone not familiar with formulaic poetry to easily understand the scope and character of variation. In short, we are experimenting with how to take the strong points of each tool and combine them into something that is both powerful as well as easier to use.

Collaboration in Practice

Currently, research in computational social science and digital humanities rarely permeates back into their core disciplines. The problem is that current tools and approaches are often borrowed from fields where both data and research protocols are much more standardized. In the humanities, on the other hand, available datasets often have not been created for today's research, and, as a result, they are rife with complex biases. If not properly handled, these biases easily invalidate any computational research based on the corpora. Invariably, there are also gaps between what can be produced through automated means, and the nuanced human categories of interest. Thus, to produce results of interest to the subject domain, computational research by necessity would need to interleave computational inference with manual interpretation to produce the final data conclusions are based on.

Here, a challenge for a humanist is how to describe and document work processes well enough not only to give other humanists the possibility to reach similar conclusions, but to help the computational scientist to understand the process in order to make some parts of it easier. Due to the complexity of variation in the corpus, an efficient process must be equally complex and flexible, and enable the movement between quantitative views and manual interpretation of individual texts.

Works Cited

Abello, James, Peter Broadwell & Timothy R. Tangherlini 2012. "Computational Folkloristics". *Communications of the ACM* 55(7). Pp. 60–70. <https://doi.org/10.1145/2209249.2209267>

Anttonen, Pertti 2005. *Tradition through Modernity: Postmodernism and the Nation-State in Folklore Scholarship*. Helsinki: Finnish Literature Society. <https://doi.org/10.21435/sff.15>

Arvidsson, Alf, Lauri Harvilahti, Audun Kjus, Clíona O'Carroll, Susanne Österlund-Pötzsch, Fredrik Skott & Rita Treija (eds.) 2018. *Visions and Traditions: Knowledge Production and Tradition Archives*. Helsinki: Academia Scientiarum Fennica.

Biemann, Chris 2006. "Chinese Whispers: An Efficient Graph Clustering Algorithm and Its Application to Natural Language Processing Problems". *Proceedings of TextGraphs: The First Workshop on Graph Based Methods for Natural Language Processing* (June 2006). Pp. 73–80. <https://dl.acm.org/doi/10.5555/1654758.1654774>

Frog 2010. *Baldr and Lemminkäinen: Approaching the Evolution of Mythological Narrative through the Activating Power of Expression: A Case Study in Germanic and Finno-Karelian Cultural Contact and Exchange*. UCL Eprints. London: University College London. <http://eprints.ucl.ac.uk/19428/>

Frog 2016. "Linguistic Multiforms in Kalevalaic Epic: Toward a Typology". In *The Ecology of Metre*. Ed. Ilya Sverdlov & Frog. Special issue, *RMN Newsletter* 11: 61–98.

Harvilahti, Lauri 1992. "The Production of Finnish Epic Poetry: Fixed Wholes or Creative Compositions?". *Oral Tradition* 7(1): 87–101. <http://journal.oraltradition.org/issues/7i/harvilahti>

Harvilahti, Lauri 2013. "The SKVR Database of Ancient Poems of the Finnish People in Kalevala Meter and the Semantic Kalevala". *Oral Tradition* 28(2): 223–232.

Harvilahti, Lauri 2019. "History of Computational Folkloristics in Finland and Some Current Perspectives". *Folkloristics in the Digital Age*. Ed. Pekka Hakamies & Anne Heimo. Helsinki: Academia Scientiarum Fennica. Pp. 158–175.

Hakamies, Pekka, & Anne Heimo (eds.) 2019. *Folkloristics in the Digital Age*. Helsinki: Academia Scientiarum Fennica.

Hämäläinen, Mika, Tanja Säily, Jack Rueter, Jörg Tiedemann & Eetu Mäkelä 2018. "Normalizing early English Letters to Present-Day English Spelling". *Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Ed. Alex B. Degaetano-Ortlieb et al. Stroudsburg, PA: Association for Computational Linguistics. Pp. 87–96.

Ilyefalvi, Emese 2018. "The Theoretical, Methodological and Technical Issues of Digital Folklore Databases and Computational Folkloristics". *Acta Ethnographica Hungarica* 63(1): 209–258.

Isoaho, Karoliina, Gritsenko, Daria & Mäkelä, Eetu. 2020. "Topic Modeling and Text Analysis for Qualitative Policy Research". *Policy Studies Journal*. <https://doi.org/10.1111/psj.12343>

Jänicke, Stefan, & David Joseph Wrisley 2017. "Visualizing Mouvance: Toward a Visual Analysis of Variant Medieval Text Traditions". *Digital Scholarship in the Humanities* 32(suppl_2): ii106–ii123.

Kalkun, Andreas 2015. *Seto laul eesti folkloristika ajaloos: Lisandusi representatsiooniloole*. Tartu: Eesti Kirjandusmuuseum.

Kallio, Kati, Frog & Mari Sarv 2017. "What to Call the Poetic Form: Kalevala-Meter or Kalevalaic Verse, *regivärss*, Runosong, the Finnic Tetrameter, Finnic Alliterative Verse or Something Else?' *RMN Newsletter* 12–13: 94–117. <http://hdl.handle.net/10138/305420>

Kallio, Kati, & Eetu Mäkelä 2019. "Suullisen runon sähköisestä lukemisesta". *Elore* 26(2): 25–40. <https://doi.org/10.30666/elore.84570>

Moretti, Franco 2013. *Distant Reading*. Brooklyn: Verso.

Mäkelä, Eetu, Mikko Tolonen, Jani Marjanen, Antti Kanner, Ville Vaara & Leo Lahti 2019. "Interdisciplinary Collaboration in Studying Newspaper Materiality". *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference (DHN 2019), CEUR Workshop Proceedings* 2365: 55–66. <http://ceur-ws.org/Vol-2365/07-TwinTalks-DHN2019_paper_7.pdf>

Mäkelä, Eetu, Anu Koivunen, Antti Kanner, Maciej Janicki, Auli Harju, Julius Hokkanen & Olli Seuri 2020a. "An Approach for Agile Interdisciplinary Digital Humanities Research: A Case Study in Journalism". *Proceedings of Twin Talks at the Digital Humanities in the Nordic Countries 2020. CEUR Workshop Proceedings*. <http://ceur-ws.org/Vol-2717/paper01.pdf>

Mäkelä, Eetu, Krista Lagus, Leo Lahti, Tanja Säily, Mikko Tolonen, Mika Hämäläinen, Samuli Kaislaniemi & Terttu Nevalainen 2020b. "Wrangling with Non-Standard Data". *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020), CEUR Workshop Proceedings*. <http://ceur-ws.org/Vol-2612/paper6.pdf>

Saarlo, Liina 2005. *Eesti regilaulude stereotüüpiast: Teooria, meetod ja tähendus*. Tartu: Tartu Ülikooli Kirjastus.

Säily, Tanja, Eetu Mäkelä & Mika Hämäläinen 2018. "Explorations into the Social Contexts of Neologism Use in Early English Correspondence". *Pragmatics & Cognition* 25(1): 30–49.

Sarv, Mari 2019. "Poetic Metre as a Function of Language: Linguistic Grounds for Metrical Variation in Estonian Runosongs". *Studia Metrica et Poetica* 6(2): 102–148. <https://doi.org/10.12697/smp.2019.6.2.04>

Sarv, Mari, & Janika Oras 2020. "From Tradition to Data: The Case of Estonian Runosong". *Arv: Nordic Yearbook of Folklore* 76: 105-117.

Tangherlini, Timothy R. 2016. "Big Folklore: A Special Issue on Computational Folkloristics". *Journal of American Folklore* 129(511): 5–13. <https://www.jstor.org/stable/10.5406/jamerfolk.129.511.0005>

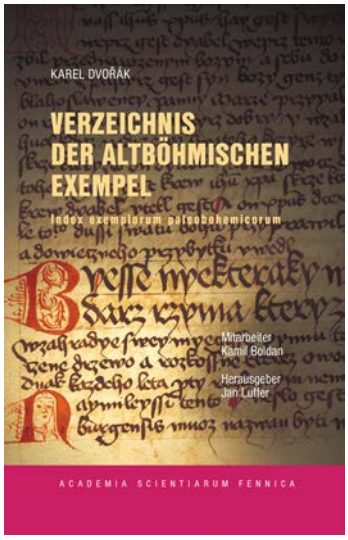
Tangherlini, Timothy R. 2013. "The Folklore Macroscope: Challenges for a Computational Folkloristics". *Western Folklore* 72(1): 7–27. <https://www.jstor.org/stable/24550905>

Tarkka, Lotte 2013. *Songs of the Border People: Genre, Reflexivity, and Performance in Karelian Oral Poetry*. Helsinki: Academia Scientiarum Fennica.

Tarkka, Lotte, Eila Stepanova & Heidi Haapoja-Mäkelä 2018. "The *Kalevala*'s Languages: Receptions, Myths, and Ideologies". *Journal of Finnish Studies* 21(1–2): 15–45. <http://hdl.handle.net/10138/301432>

Timonen, Senni 2004. *Minä, tila, tunne: Näkökulmia kalevalamittaiseen kansanlyriikkaan*. Helsinki: SKS.

Wagner, Robert A. and Michael J. Fischer 1974. 'The String-to-String Correction Problem". *Journal of the ACM* 21(I): 168–173. <https://doi.org/10.1145/321796.321811>



Verzeichnis der altböhmischen Exempel
Index exemplorum paleobohemicorum

Karel Dvořák

Mitarbeiter
Kamil Boldan

Herausgeber
Jan Luffer

Suomalainen Tiedeakatemia
Folklore Fellows' Communications 318
Tallinn 2019, 307 pp.
ISBN 978-951-41-1141-9
Available at the Tiedekirja bookstore, €38.00.

Bengt af Klintberg

Karel Dvorák (1913–1989) was a Czech folklorist who became professor of ethnology and folkloristics at the Charles University in Prague. Like many other folklorists, he first studied literature, and his special field as a scholar was the relationship between Czech medieval literature and folk culture.

A first edition of the book that has now been published in German translation was printed in Czech already in 1978. It soon became a useful tool for Czech folklorists and medievalists, but Dvorák realized that the material was also of interest to the international folklore community. His goal was to bring about an extended version in German for the Folklore Fellows' Communications, but he did not live to see the fulfillment of this project. During long periods the work was at a standstill, and it took more than forty years before it was brought to a successful close.

It is thanks to folklorist Jan Luffer and medievalist Kamil Boldan that an international edition of the Czech exempla index has finally been realized. The latter made contacts with Dvorák as a young scholar during his research into a manuscript in the Czech national library, *Historiae variae moralisatae*, written in the end of the fourteenth century. The 230 exempla texts there were not included in the first edition but were incorporated into the present edition by Dvorák and Boldan in collaboration, a work that was completed by Boldan after Dvorák's death.

After that the manuscript remained untouched for several years. What was missing was an up-to-date introduction and indexes. The latter were brought about by Dvorák's folklorist colleague Dagmar Klimova, and the introduction, a thorough presentation of the exempla as a genre and the Czech sources, was written by the medievalist Anezka Vidmanová. The extensive final editorial work was done by Jan Luffer and took ten years. The book now appearing in the international folklore scene is thus the result of impressive scholarly efforts.

Most folklorists apprehend the exempla literature as a genre which flourished in the high or late Middle Ages, narratives with a moralizing and edifying message. Vidmanová's introduction puts the term into a wider context and outlines its pre-history: the exampla are not just "monk tales" but constitute a literature, both religious and secular, which has existed since classical times and whose characteristic is that it conveys useful knowledge. The medieval versions of Aesop's fables, for instance, make up a substantial part of the exempla literature. The renaissance of the twelfth century resulted in the spread of many classical narratives, both philosophical and pseudohistorical, translated from Latin into the local languages.

Vidmanova's introduction is followed by Dvorák's own introduction to the first edition from 1978. He establishes that the Czech exempla literature reached its peak during the second half of the fourteenth century. When the Hussite movement swept over Bohemia in the fifteenth century, the Church took a stern view of the many stories which had no connection to the Bible. In other parts of Europe, the classical heritage could live on longer.

The book which has been the model for Dvorák's *Verzeichnis* is Frederic C. Tubach's large *Index exemplorum: A Handbook of Medieval Religious Tales* of more than 500 pages and published in 1969 as number 204 in the Folklore Fellows' Communications series. Dvorák stresses his debt of gratitude to Tubach, but he admits that one often would

like Tubach's plot summaries to be more detailed. He also notes that additions could be made to Tubach's references to secondary literature.

Like Tubach, Dvorák presents his material in alphabetical order. The most significant word in the caption of the exemplum decides its place. Dvorák has maintained Tubach's type numbers and English captions, but the contents are summarized in German. The type numbers not found in the Czech material have been left out, but in return Dvorák has introduced a great many types not found in *Index exemplorum*. These have the same number as a caption with the same catchword, but an asterisk indicates that it is a new type. This can be illustrated through three exempla that have been adapted to Tubach's type 23, "Abraham and wife Sarah". This story from Genesis is missing from Dvorák's material. He has, however, added 23*, "Abraham as a host", 23**, "Abraham erecting the family tomb" and 23***, "Abraham's sons", none of which are in Tubach's index.

After summaries of the contents follow references, first to Czech sources, thereafter to other sources where the narrative is found. This section is followed by references to secondary literature and finally also to proverb collections containing texts related to the exemplum in question. The last-mentioned information makes Dvorák's index a useful aid for paremiologists.

How can a non-Czech folklorist use Dvorák's *Verzeichnis*?

My impression is that all folklorists, regardless of specialization, will find something of interest among the more than a thousand narratives, some well-known, others never heard of. Those who have understood exempla as being religious narratives in a Christian setting will probably have to revise their opinion. To be sure, most of the stories, such as the animal fables which make up a substantial part of the index, are moralizing and edifying narratives, but they are not set in a religious context. One finds, as a matter of fact, types in Dvorák's index which are neither religious nor narratives. This is especially true about some of the texts in *Historiae* that have no dramatic plot; they just mediate useful knowledge in general terms. For example, 4946*, "Travelling in winter", goes like this in translation: "He who undertakes a journey in winter must above all take heed not to become hungry on his way. With an empty stomach the natural bodily warmth disappears, which is what happens with the light in a lamp when the oil is gone." A wise remark, to be sure, but there is no hint in the summary that this should be given a Christian interpretation.

Those readers who are folktale scholars will note that no less than 130 types have a number in Aarne-Thompson-Uther's *The Types of International Folktales*. A comparison of the presentation in Dvorak's *Verzeichnis* with the one in ATU shows that the information given is far from the same in the two handbooks. In some cases, it is more extensive in ATU, in others the *Verzeichnis* contains more details. This shows that folktale scholars have good reason to consult the Czech index.

As a random sample, I have compared the presentation of type 3378, "Monk Felix", in Dvorák's (and Tubach's) index and the ATU type 471A, "The Monk and the Bird". This well-known Christian legend is about a monk who listens to heavenly, beautiful bird song. When he returns to the monastery, it appears that several hundred years have passed. A folklorist who would like to investigate this legend should not forget to consult Dvorák's index. It contains not only references to Czech variants in the collections "Arnesti Pragensis" and "Klaret" but also around ten references to literature which are not found in ATU.

My personal interest in contemporary narrative tradition made me note that some of the Czeck exempla have survived into our time. Here are three examples:

1482*, "Death, threefold, predetermined by fate". – The son of a merchant is told that he will drown, die from being dragged after a wagon and be hanged. This story (ATU type 934) has a modern offspring in the story about a man who tries to commit suicide by simultaneously drowning, shooting and hanging himself but fails.

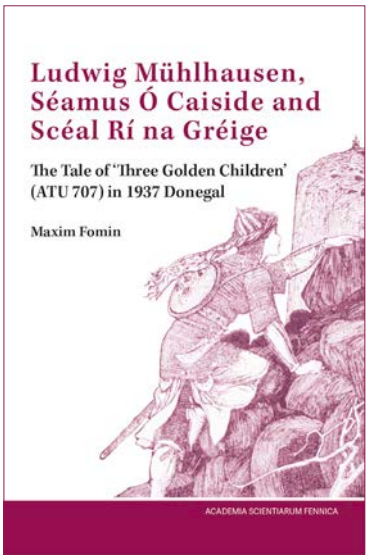
3251, "Medicaments, interchanged".** – A mistake is committed at the drugstore: a knight who looks forward to a love night gets laxative instead of aphrodisiac and a monk with constipation gets the aphrodisiac. This burlesque story is still alive today.

5279*, "Wife instead of harlot on rendezvous with husband". – A wife in disguise takes the place of a prostitute with whom her husband has arranged a rendezvous. Afterwards she reveals who she is and scolds him. In the contemporary versions the wife is a prostitute without her husband knowing it. He finds out this when he becomes her customer. The event has been reported in daily papers several times as a true story.

In the end of the book the reader finds three indexes. The first is a concordance with *The Types of International Folktales*. It shows that the most frequent folktale category in the exempla index are the animal tales, with 48 items. The second index is about the protagonists of the tales. Here one notices that persons from classical antiquity make up a larger portion than both Biblical persons and saints. One finds no less than 36 stories from the cycle of Alexander the Great, which could be compared to the only 19 stories about Christ. The third index is a detailed subject index running over 25 pages.

Indexes of the kind represented in this book have already from the very start been distinctive to the series Folklore Fellow's Communications. Karel Dvorák's *Verzeichnis der altböhmischen Exempel*, exemplarily edited by Jan Luffer, will without doubt stimulate research not only in the field of medieval exempla but in folktale research as a whole.

Folklore Fellows´Communications in 2020

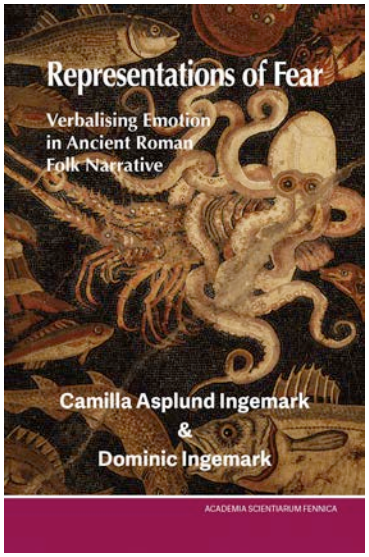


Ludwig Mülhausen, Séamus Ó Caiside and Scéal Rí na Gréige. The Tale of 'Three Golden Children' (ATU 707) in 1937 Donegal

Maxim Fomin

Suomalainen Tiedeakatemia
Folklore Fellows' Communications 319
Helsinki 2020
234 pp.
ISBN 978-951-41-1142-6
Available at the Tiedekirja bookstore, 34€

Read more at https://www.folklorefellows.fi/ffc-319/

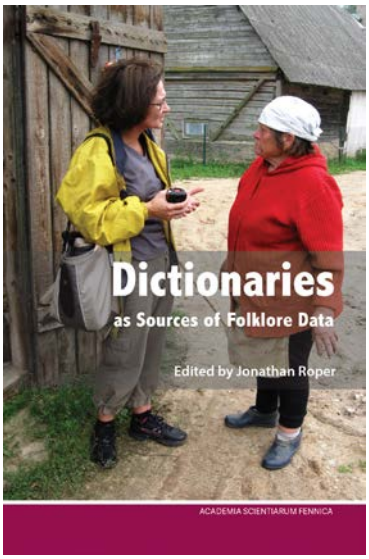


Representations of Fear. Verbalising Emotion in Ancient Roman Folk Narrative

Camilla Asplund Ingemark & Dominic Ingemark

Suomalainen Tiedeakatemia
Folklore Fellows' Communications 320
Helsinki 2020
362 p.
ISBN 978-951-41-1156-3
Available at the Tiedekirja bookstore, 38€

Read more at https://www.folklorefellows.fi/ffc-320/



Dictionaries as Sources of Folklore Data

Ed. Jonathan Roper

Suomalainen Tiedeakatemia
Folklore Fellows' Communications 321
Helsinki 2020
246 pp.
ISBN 978-951-41-1157-0
Available at the Tiedekirja bookstore, 28€

Read more at https://www.folklorefellows.fi/ffc-321/

Find the whole catalogue at https://www.folklorefellows.fi/folklore-fellows-communications/complete-catalogue/