John Laudun / johnlaudun@gmail.com
American Folklore Society
2014 / Santa Fe, New Mexico

# *Counting Tales*

**Folklore's Contribution to the Computational Model of Narrative**

#

**ABSTRACT**

*A number of scientists and humanists have begun to consider possible quantitative approaches to the study of what seems like a most unquantifiable object of study, narrative. The goal has been both to speed up the process as well as to remove the ambiguity inherent in many human-generated results. This paper contains many of the same errors encountered by unwitting scientists, from clumsy compressions to awkward leaps of logics, but in focusing on the task of constructing a morphology of a small collection of tales, I hope to explore what computational abstractions might be possible, and useful, to folklorists.*

#

**PAPER**

Seven years ago I set out to understand the relationship between land and water in the imaginations of Louisiana residents. My exploration of what might be called the folk imaginary quickly focused on a single artifact, a recently invented but completely traditional artifact known locally as the crawfish boat. It was easy to fixate on the crawfish boat both because, first, it's just cool -- come on, look at that thing -- and, second, my attempts to find other clear examples of the human imagination dealing directly with the intersection of land and water had turned up so little. There are, no doubt, mentions of land and mentions of water, either in abstract or in particular, in the thousands across the many collections of Louisiana folklore, but there are relatively few texts that deal with the relationship between the two in a focused fashion.

The exception to this generalization are stories about buried treasure, which, in Louisiana, regularly place the treasure under a tree by a lake, or on shore of a mud island, or in a stump along a creek bed. As work on the crawfish boat drew to a close, I found myself culling stories about buried treasure from a variety of sources, in hopes of discovering why these tales, and seemingly no other, featured a dimension of the landscape that had been my obsession for so long. If they were seemingly singular in the context of Louisiana folk culture, would they reflect similar or dissimilar properties within American folk culture? And, just as importantly, I was also interested in how one might go about making such claims.

These two interests, one topical and one methodological, drove me to construct a small collection of oral texts, a mere 20 in all as well as to construct or seek out a number of reference collections, the most useful of which can be found in Gerard Hurley's mid-century survey of

"Buried Treasure Tales in America." Drawing upon treasure stories culled from print collections published largely in the 1930s and 1940s, Hurley concluded that American treasure tales have "three main characteristics": first, that they be "brief and factual"; second, that "their plots have a simple two-part structure"; and, third, that they "usually end with the treasure not being found" (197).

Setting aside the other dimensions of the plot that draw most of Hurley's attention in his survey--such as the nature of the treasure's protection or how legends make claims about knowledge of a treasure-- I would like to focus on Hurley's simple two-part narrative structure, which consists of, first, the treasure being hidden, and, then, the treasure being hunted for. Perhaps more importantly, he observed that "few tales give detailed attention to both parts" (199). This structure stood for a quarter century until Pat Mullen revisited it in 1978 in an attempt to work through the implications of Dundes' notion of folk ideas.

For Hurley, the fact that "American treasure tales usually end with the treasure not being found" revealed that such tales are wish narratives and not wish-fulfillment narratives and that the "essential interest is concerned with the idea of treasure and the thrill of the hunt" making "the treasure tale successful even though the search in it is not" (204). But for Mullen, the success or failure to find the treasure was a necessary third component of a treasure tale plot, though like Hurley he observed that most tales end in failure and that the accounts of failure are often where narrators invest most of the text's attention (209).

Mullen's principle interest, of course, is in how American treasure tales disrupt the folk idea of America as the land of unlimited good. Extending Dundes' argument, Mullen concluded that, given the many dimensions of at least morally questionable if not evil deeds contained within most treasure tales, the narratives are far from wish fulfillment stories and more akin to warning against seeking easy money. Whether "the initial burial of a treasure is clouded with acts of murder ... [or] the discovery of treasure leads to murder and treachery," the overriding theme of these stories is, as Mullen noted, "you don't get something for nothing" (218-219). The happy endings are few and far between.

In my own work, I came to much the same conclusions: all Louisiana treasure legends were of the same form, consisting of two components, which I called *alpha* (α) and *tau* (τ). I defined *alpha* as the securing or loss of the treasure in the impersonal and more distant past, and *tau* as the experience of seeking treasure, often located within a personalized, or localized, and more recent, past. Like Hurley and Mullen, not all the texts in my collection contained both parts, but I was confident enough in the similarities between those that did, and in the standalone legends that are either *tau* or *alpha* to say that in some fashion the other part's presence can, analytically, be inferred.

This is an important distinction. I am not suggesting in any way a devolution of an uber-legend that has all its constituent parts, but making what I think is the rather interesting observation that I have twenty texts, and of those twenty texts, fourteen consist solely of the experience of seeking treasure (τ); four are only about how a treasure came to be where it is (α);

and two contain both parts, though those parts are in inverse order to each other: one tale proceeds in a chronological order and the other in an expository order.[1]

An example of each component should be helpful here. Let's begin with the dominant narrative component, τau. (Handout.) The first thing to notice is the presence of a preface to the text -- "I went to meet an old man in Marrerro, and he told me a story." Such prefaces occur in fourteen of the twenty texts and typically establish the speaker as part of a diegetic chain. In some cases, the preface seeks to establish historical veracity, or, in true hybrid fashion, seeks to negotiate the divide that is commonly understood to exist between being the receiver of a story and being able to vouchsafe the historical truth of the story's contents. After the preface in this short text comes the *tau* component, and it offers up the first of two variants within the *tau*: that of the actant who digs for treasure.

The other *tau* variant can be seen in the second example where the actants do not seek the treasure, but rather comes across it while plowing, a version of digging, in pursuit of reward through work. Such a pursuit in fact leads to them actually finding the treasure.

If we compare the two *taus* as a series of narrative states (or functions), then we see that while A is exactly the same, and B is quite often similar on its face, it is C that is inverted. The only explanation for the difference is human intentionality. If you go to a location with the intent of finding treasure, you will not only not find it, but you will probably find yourself a spirit from which you will then have to escape. If, however, you go to a location with the intent of doing some kind of work, the kind of work that in some texts is tied to the regular accretion of wealth, like gathering pine knots, then you may very well find treasure. If you do not find treasure, then you will encounter a tomb or stone slab that in other texts is associated with treasure. Thus, the event states for the two forms of tau are as follows:

in $\tau_1$ the actant goes to a location, digs for treasure, and experiences a spirit.

in $\tau_2$ the actant goes to a location, performs an agricultural task (plows, gets cows, gathers moss, hunts), and finds a treasure.

Interestingly, the location the actants go to in both versions is often the same: woods dominate, and are either directly mentioned or suggested in six out of the fifteen tau components.

I should be clear that I arrived at this morphology the old-fashioned way, by hand. I was hoping to begin to glean some way to automate the process, which may sound like I am either monumentally lazy or just crazy. (Choose your *azy* wisely.) My efforts to do so were, in fact, born of efforts by computer scientists to automate morphological analyses of narrative. For those not familiar with recent work in the computational modeling of narrative, I will note that scientists like Mark Finlayson were drawn to folktales because, in their view, such tales are born of a "Darwinian-like natural selection process, in which portions of the narratives that are congruent with the culture are retained and amplified, and those that are incongruent are distorted or

discarded" (2009: 127).

Much of the work in computer science and computational linguistics has focused on characters as actants within distinct environments whose social networks can be graphed. Some of you may be familiar with the recent work by statistical physicists who applied "apply statistical mechanical tools to analyze the networks underlying three iconic mythological narratives with a view to identifying common and distinguishing quantitative features" (1). In contrast, Finlayson has been was fascinated by the idea that "all morphologies seem to share a common high-level structure ... and have significant overlap in rough identity of the functions, but vary considerably in specific function sequences and other details" (2009: 128).[2]

The goal of such studies is to develop computational methods of analysis for a given set of stories. The goal in doing so is both to speed up the process as well as to remove the ambiguity inherent in many human-generated results. That is, how much a particular morphology, for example, owes to the analyst, who may have been influenced by other morphologies; how much such a morphology reflects the narratives being examined; and how involved any verification of such a morphology must be. Most methods, for all their interesting moves--using such things as Bayesian models and analogical story merging--unfortunately depend upon a transcoding of narratives into a computer-readable format. Quite often this transcoding means that larger chunks of discourse get reduced to plot summaries and the analyst must, perforce, transform actual words used to analytical terms that are capable of comparing one narrative to another.

Finlayson's model is reasonable enough: many narratives proceed through a series of looped iterations that, while achieving a myriad of dramatic and/or cognitive effects, are often much the same in actual appearance. If we take such a claim at face value, it should be possible to reduce the representation of such texts to simpler forms, and perhaps in the process understand something about how the forms themselves get built "up" or "out" of simpler schema.

Unfortunately for Finlayson, and others, much of their work is built upon computational and corpus linguistics, which has largely attended to language as it occurs at the level of the sentence.

Computer scientists have bigger plans, however, and they want to leap from sentences to Shakespeare, from texts of tens of words to texts of tens of thousands, if not hundreds of thousands, of words. It's a really big leap, and I think folklorists could profit from the fact that we are in possession of a whole lot of texts not only in the middle in terms of size but a bit more hardened by regular use. (It's a tradition thing, you know?)

I have a number of ideas about how I would start building a computationally operable morphology, but much of it amounts to sketches on the backs of napkins, a great deal of napkins, and I'd be happy to discuss those possibilities with anyone brave enough to ask. What I would like to offer next is an example of such computational analysis and the particular insight to which it led me.

Some of you, okay probably none of you, read the essay in the *Journal of American Folklore* where my colleague Jonathan Goodwin and I attempted to understand the intellectual history of folklore studies through the use of topic models. Our particular interest was to see if "the turn

toward performance" as folklorists know it could be glimpsed through crunching how often words appear together. Of course, as a discourse that arrived with a particular terminology, the rise of performance studies could be discerned, but we were also fascinated by the adjacent discourses and the turn towards politics that seems, in fact, to be displacing performance studies over the past decade.

Put most simply, topic models are, from one point of view, a fairly simple way to approach the way texts mean something within a larger network or collection of texts: so the first thing a topic model does is to assume that meanings are relational. The second assumption is that meanings can be found in sets of words, or clusters of words, that tend to stick together over time, increasing their tendency to co-occur. As an introduction to topic modeling in a recent issue of the journal Poetics warns:

> Topic models capture co-occurrences regardless of these words' embeddedness within other complexities of language--such as syntax, narrative, or location within the text. Instead, each document is treated as if it were a so-called "bag of words." The goals of a topic model analysis are then to analyze these various word bags, to identify word co-occurrence patterns across the corpus of bags, and then to use these to produce a mapping of the distribution of words into the topics and of the topics into the bags. (Mohr and Bogdanov: 545)

We could have a longer discussion about topic models: if my experience at presenting this work to humanists elsewhere is any indication, there are a lot of heads, or tempers, on the verge of exploding. Bear with me for just a moment longer.

To keep it simple, let me demonstrate how words co-occur within the small collection of texts I am treating today. As you can see, out of the 16 times that the word *buried* occurs, it is accompanied by the word *money* exactly half the time: three times it comes before, and the remaining five times it comes after. Just as importantly, *money* is always within three words of *buried*, reflecting a kind of syntactic urgency to the relationship. (I have also highlighted treasure as a way of showing the way we can quite literally see synonyms at work within a corpus.)

If we were to take these word co-occurrences within the collection and imagine it as a network of words, wherein the words that most frequently occur with each other are linked and the links are directional such that we can see what words lead to other words, we can see that buried and money are indeed near the center of the Louisiana treasure tale constellation.

The locative/deictic *there* is also central, with many of the tales quite literally passing through it. (This makes for a very language-centered approach to thinking about texts that I find quite compelling.)

And here's where it gets interesting in a way that most folklorists find things interesting: looking at this graph I was struck by the convergence at the bottom of the graph of the word *old*

and then, from there, a divergence onto three seemingly equivocal objects: *slave*, *stage*, *station*. How, I wondered, is a slave like a stage? That makes no sense.

But such a question paired well with another question I had often asked myself when thinking about a legend that I myself had collected over a decade ago. In that story, the narrator stops by his family home and sees people digging in the backyard for treasure. He joins them in prayer and then takes on the job, along with his nephew, of taking water to the differs. As they shuttle back and forth, the two encounter a pirate in a tree, who asks for something to drink. They give him some. And the pirate asks again. They get scared and don't give him anything, and the Pirate threatens them. They run. And the story ends with a shovel flying through the air and landing in a tree. My question has always been how did that pirate get in that tree and why was he threatening a bunch of African Americans?

With *slave* turning up in an odd position in the network graph of the corpus of oral legends, I went back through the texts and marked the following for their contents:

- In Broussard 4, pirates shoot a crewman so that his ghost will protect the treasure.
- In LOH 162, a family fortune is first buried in a barrel of flour and then when the money is transported west by a family member, and *two slaves*, the family member dies and is buried with the money by the slaves.
- In LOH 164, a slave is killed after promising to look after the family fortune so that his spirit "would continue to guard the money."
- In LOH 157, a rambling story with little to recommend it beyond the idea of there being gold still buried somewhere, a dead man is attached to the treasure and transportation is involved --a regular feature, by the way, of the *alpha* narratives.

A rough sketch of an ideational network would report the following associations: treasure, tree, dead man, pirates, slaves. (See Figure 2.) Is there, then, a text that allows us to bridge from slaves to pirates in a way that is other than the two occupying the same role within this corpus?

There is another text that points the way: LOH 163 tells the story of famous pirate Jean Lafitte moving to Texas to escape being arrested for all the things he has done wrong, while also referencing what he has done right. Lafitte takes up residence in Galveston, but still conducts business in Louisiana. One day he has to abandon a ship in Sabine Lake, full of treasure. What kind of treasure might be found at the bottom of the lake? As it turns out, there are a number of historical accounts that indicate that pirates were involved in the slave trade, and one of those accounts turned up in the web collection of the corpus:

> By 1817 the privateers of Jean Lafitte and his predecessor, Luis de Aury, were capturing numerous Spanish slavers off the coast of Cuba. The pirate's barracoons, or slave pens, on Galveston Island were often swelled beyond capacity, containing a thousand or more

> African chattels. Many buyers came to the island to buy slaves at $1.00 per pound, and three brothers, John, Rezin, and James Bowie, were among the pirate's best customers. In 1853 John Bowie recorded in "DeBow's Magazine" that the brothers, who channeled their illicit slave trade via Black Bayou on Lake Sabine or via the Calcasieu to Lake Charles, realized a net profit of $65,000 in two years time from the sale of 1,500 Africans in Louisiana. [TN 8]

Such an account as this one reveals a clear association between slaves and treasure, through piracy. Slaves were the treasure, and they were traded in the border areas of the new nation. Without saying it, and perhaps without knowing it directly but feeling the rightness of the connection, Oscar Babineaux maintains a legend with a powerful historical lesson.

For the purposes of demonstrating, and exploring, the possible utility of what might seem like old folkloristic ideas and methodologies like morphologies with new fangled analyses like topical networks, I have used only a small collection of texts here as a playground, and I have gathered them around a particular text in order to throw such processes into relief. The goal in doing so in order was to begin to understand how one might scale up such an enterprise, grounded in the humanistic notion of qualitative inquiry.

But perhaps some are wondering, "Why count words?" One initial answer to the question is simple: I counted words because I wanted to know just how many words were required to invoke a *storyworld* because I want to understand how that happens. Given the size of a great number of literary forms, one thousand words is already amazingly concise, but one hundred words? Each word must pack an incredible amount of power: something made even more amazing when one realizes that only half that number of words are unique in their usage, even in such little texts!

How is this possible? How can such a small subset of words from an already small text make a story go? That is, I think, the real question. Counting words is but one step along the way, but an important one, and one that we, as folklorists, have failed to undertake. Think for a minute of all the texts that are indexed in the great collection projects of the twentieth century. Add to them all the texts we have collected under the auspices of the ethnography of speaking. It's an impressive amount of work, and while we have made some synthetic gestures, we have, by and large, mostly focused on differences. All of those differences are, of course, quite compelling, but in focusing on differences, we have also missed an opportunity to make attempts at larger kinds of claims about human nature and culture.

#

**WORKS CITED**

Abello, James, Peter Broadwell, and Timothy R Tangherlini. 2012. Computational Folkloristics. *Communications of the ACM* 55/7: 60.

Ancelet, Barry Jean. 1994. *Cajun and Creole Folktales: The French Oral Tradition in South Louisiana*. Garland Publishing.

Barthes, Roland. 1981. *Camera Lucida: Reflections on Photography*. Tr. Richard Howard. Farrar,

Strauss, Giroux.
Carron, Pádraig Mac, and Ralph Kenna. 2012. Universal Properties of Mythological Networks. *Europhysics Letters* 99/2: 28002.
Dundes, Alan. 1971. Folk Ideas as Units of Worldview. *Journal of American Folklore* 84/331: 93-103.
Hurley, Gerard T. 1951. Buried Treasure Tales in America. *Western Folklore* 10/3: 197-216.
Laudun, John. 2012. "Talking Shit" in Rayne: How Aesthetic Features Reveal Ethical Structures. *Journal of American Folklore* 125/497: 304–326.
Lindahl, Carl, Maida Owens, and C. Renée Harvison. 1997. *Swapping Stories: Folktales from Louisiana*. University Press of Mississippi.
Mohr, John W, and Petko Bogdanov. 2013. Introduction--Topic Models: What They Are and Why They Matter. *Poetics* 41/6: 545–69.
Mullen, Patrick. 1978. The Folk Idea of Unlimited Good in American Buried Treasure Legends. *Journal of Folklore Institute* 15/3: 209-220.
Saxon, Lyle, Edward Dryer, and Robert Tallant. 1945/1987. *Gumbo Ya-Ya: A Collection of Louisiana Folk Tales*. Reprinted by Pelican Publishing Company.
Tangherlini, Timothy R. 2010. Legendary Performances: Folklore, Repertoire and Mapping. *Ethnologia Europaea* 40/2: 103–15.
Tangherlini, Timothy R. 2013. "The Folklore Macroscope." *Western Folklore* 72/1: 7–27.
Word, Christine. 1988. *Ghosts along the Bayou: Tales of Hauntings in Southwestern Louisiana*. The Acadiana Press.
Victoria Johansson. 2008. Lexical diversity and lexical density in speech and writing: a developmental perspective. *Lund University, Dept. of Linguistics and Phonetics Working Papers* 53: 61-79.

# # #

1. Upon realizing this, that I had so many stories that were so much the same, I double-checked the procedures I had used for acquiring texts: had I left out stories that should be included? Had I intuited earlier than I realized a pattern that then determined what texts I chose to accept into the collection? So far as I can tell, nothing of the sort occurred. I went looking for any and all texts that mentioned treasure in some fashion, whether it was called simply money or gold or coins or anything else. I gradually broadened my scope to include stories that suggested the possibility of an unknown reward or bounty, but such broadening should only have brought more variety to the possible structures of the texts, not a narrowing.

2. Finlayson is not alone. David Elson has focused on building a thematic model of actant-event scenarios in order to discern story similarity and analogy (2012).