

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Poetics

journal homepage: www.elsevier.com/locate/poetic

Capturing distinctions while mining text data: Toward low-tech formalization for text analysis

Ronald L. Breiger^{a,*}, Robin Wagner-Pacifici^b, John W. Mohr^c

^a School of Sociology, University of Arizona, Tucson AZ 85721, United States

^b Department of Sociology, The New School for Social Research, New York NY 10011, United States

^c Department of Sociology, University of California – Santa Barbara, Santa Barbara CA 93106, United States

ARTICLE INFO

Keywords:

Text mining
Hermeneutics
National security
Computational sociology
Big data
Close reading

ABSTRACT

In this article we consider some low-tech approaches to text mining. Our goal is to articulate a RiCH (Reader in Control of Hermeneutics) style of text analysis that takes advantage of the digital affordances of modern reading practices and easily deployable computational tools while also preserving the primacy of the interpretive lens of the human reader. In the article we offer three analytical interventions that are suitable to the low-tech formalizations we propose: the first and most developed intervention tracks the (normally computationally ignored) “stop” words; the second identifies the use of strategic anxiety terms in the texts; and the third (less developed in this article) introduces the grammatical features of modality (including modalization statements of probability and usuality, and modulation statements regarding degrees of obligation and inclination). All three analytical interventions provide a productive tracking of various modes and degrees of strategic decisiveness, contradiction, uncertainty and indeterminacy in a corpus of recent U.S. National Security Strategy reports.

1. Introduction

*Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura
ché la diritta via era smarrita*

At the mid-point of the path through life, I found
Myself lost in the wood so dark, the way
Ahead was lost.

-Dante Alighieri, Inferno

Recent computational advances such as natural language processing and topic modeling allow massively huge amounts of text to be processed and patterns discovered. These spectacularly efficient pattern-detection methods do not, however, capture the nuance that a reading attentive to semiotics (see [Wagner-Pacifici, 2010, 2017](#)) would uncover, and in this sense the mechanized procedures are distant from the text. Bound up with the enterprise by which texts set forth descriptions, ideas, facts, and arguments (in ways that automated text analysis procedures are increasingly superb at uncovering, and at massive scale), these texts also stutter, make points on the basis of omission and silence, play with grammatical constraints and rules of sentence structure, build upon strategic

* Corresponding author at: School of Sociology, University of Arizona, Social Sciences Bldg. Room 400, Tucson, AZ 85721-0027 USA.
E-mail address: breiger@arizona.edu (R.L. Breiger).

<https://doi.org/10.1016/j.poetic.2018.02.005>

Received 6 July 2017; Received in revised form 11 February 2018; Accepted 12 February 2018
0304-422X/ © 2018 Elsevier B.V. All rights reserved.

uncertainties, and articulate new and unforeseen meanings while switching across or articulating multiple modalities.¹ In these ways the mechanized analytical procedures are distant from the text, and in particular they miss crucial understandings that also challenge and puzzle human readers, who may often find themselves “lost in the wood so dark,” in the words that open Dante’s *Divina Commedia* and that we have chosen to begin our article. Engaging with such challenges and puzzles, arriving at tentative understandings, and then modifying those understandings on the basis of further reading, thinking, and reconnoitering—all this is implicated in the active engagement with texts that is undertaken by socially competent human readers (if only competent readers could absorb the massive amount of text that can be analyzed by computers).

Mohr, Wagner-Pacifici, Breiger, & Bogdanov (2013) outline a “third way” for text analysis, a new context within which to implement some of the main tools of contemporary “Big Data” analysis by asking how these tools might be used to advance concerns inspired by literary critic Kenneth Burke’s (1945) concept of a grammar of motives. Our paper is also concerned with identifying a “third way” for text analysis but we take a different approach toward formulating a middle way between close and distant readings.

We emphasize a family of text analysis procedures that share two principal qualities: they are low-tech, and they are aimed at uncovering—not resolving or disambiguating—strategic uncertainties in texts. As to the first point, we as readers wish to remain more in charge of the interpretive process than is often allowed in computational modeling. A philosopher (Searle, 1990: 29) once pondered the point that, if the brain is like a computer then who is the user? We respect computational tools, and we think the advantage of low-tech tools is that we can deploy them while remaining “the user,” that is, remaining in charge of the interpretive process. We insist that the reader has to be in control at each phase in the process: at the outset as a skilled reader of texts, knowledgeable about texts and their ways; in the middle as a trainer of the computer’s reading of texts; and at the end as an actively engaged reader of the computer’s readings. We thus put forward a framework that affirms a Reader in Control of Hermeneutics (RiCH) reading. We remain aspirational about folding both close and distant readings into a “middle way” by insisting that these different types of reading sustain dialogue with each other. If our approach is indeed helpful, it should in the end add new insight to the existing high-tech approaches, helping them to become even more successful. We motivate our approach to RiCH readings in Section 2.

As to the second shared quality of the low-tech techniques we will introduce, we don’t want the computer to resolve substantive ambiguity or contradictions in texts, but rather to help us uncover these features and understand the functions they may perform. Our substantive concern is how to read official *National Security Strategy* reports (NSSRs) issued under the imprimatur of the US President. We introduce these reports and our interest in them in Section 3. Even at this early stage, however, we can indicate the role of imprecision in conveying their meaning. As two officials involved in the process describe it,

“[T]here is always the issue of imprecise language. Just what is national security strategy, as opposed to grand strategy, or defense strategy, or even national military strategy? And what are the distinguishable elements of power of the United States, and the boundaries between them? How can national security strategy subsume foreign policy as the [Goldwater-Nichols] Act [the law that mandates the President to issue NSSRs] seems to imply by its language? Obviously, there was, and is, no real consensus on this language either in academia, where the public servants in Washington earlier took their training, or in Washington where they practice their arts.” (Snider & Nagl, 2001: 129).

Much as John Padgett and Christopher Ansell (1993: 1310) sought to understand the functioning of state power by identifying and analyzing “credibly imparted multivocal meanings,” we seek techniques that help readers to identify and understand the functioning of strategic ambiguity in texts. These techniques are introduced beginning in Section 4.

2. Reader in control of hermeneutics (RiCH) readings

Often analysts use various kinds of procedures in their work. Rigorous field studies, statistical models, and automated text analysis are three examples from a vastly wider set of methods, tools, and techniques. Analysts often like to think of their research results as being produced by the application of these procedures to empirical data. But really, there is always some sort of interaction between, for example, the way a procedure is set up (including the choice of procedure to be used) and the interpretation of the meter readings (or regression coefficients, or random forests, or what have you) that follow (Breiger, 2000: 109; Mohr and White, 2008: 493; Wagner-Pacifici, 2010: 1373). So the extent to which one might say that a given result is “produced” by a procedure, or instead results from an analyst, forms a continuum. Results of a method are never (if honestly employed) produced solely by the whims of an analyst. Nor are such results ever able “to speak (entirely) for themselves,” if only because the research was designed by the analyst within some sort of community of practice. So the continuum ranges from what we might call “procedure-largely-in-control” to “analyst-in-the-driver’s-seat.” (Another version of this continuum, distant versus close readings, as well as the often surprisingly fuzzy nature of the boundary between them, is apparent in the case of Big Data analysis as revealed in many of the essays in a recent special issue of *Big Data & Society* concerning the use of Big Data in social science; Mohr, Breiger, & Wagner-Pacifici, 2015.) Without giving a name to the poles of this continuum, Andrew Abbott nonetheless clearly recognizes it in his 1997 essay on the continuing relevance of the Chicago School. Abbott relates that

¹ As an example of articulating a new meaning across multiple modalities, we put forward Sewell’s (2005: 342) discussion of the origin of the concept of “revolution” in eighteenth-century France: “The articulation between the semiotics of urban crowd behavior and the semiotics of the theory of popular sovereignty changed the meanings and potentialities of both, reinforcing at once the power of the crowd and the ideology of popular sovereignty. This articulation, which created the new political category ‘revolution,’ turned out to be irreversible.”

Download English Version:

<https://daneshyari.com/en/article/7538085>

Download Persian Version:

<https://daneshyari.com/article/7538085>

[Daneshyari.com](https://daneshyari.com)