

**“READING FROM A DISTANCE:  
COMPUTATIONAL TEXT ANALYSIS  
WITH VOYEUR AND ITS  
IMPLICATIONS ON LITERARY  
CRITICISM”**

**Lydia Zvyagintseva  
MLIS/HuCo**

# WHAT IS TEXT ANALYSIS?

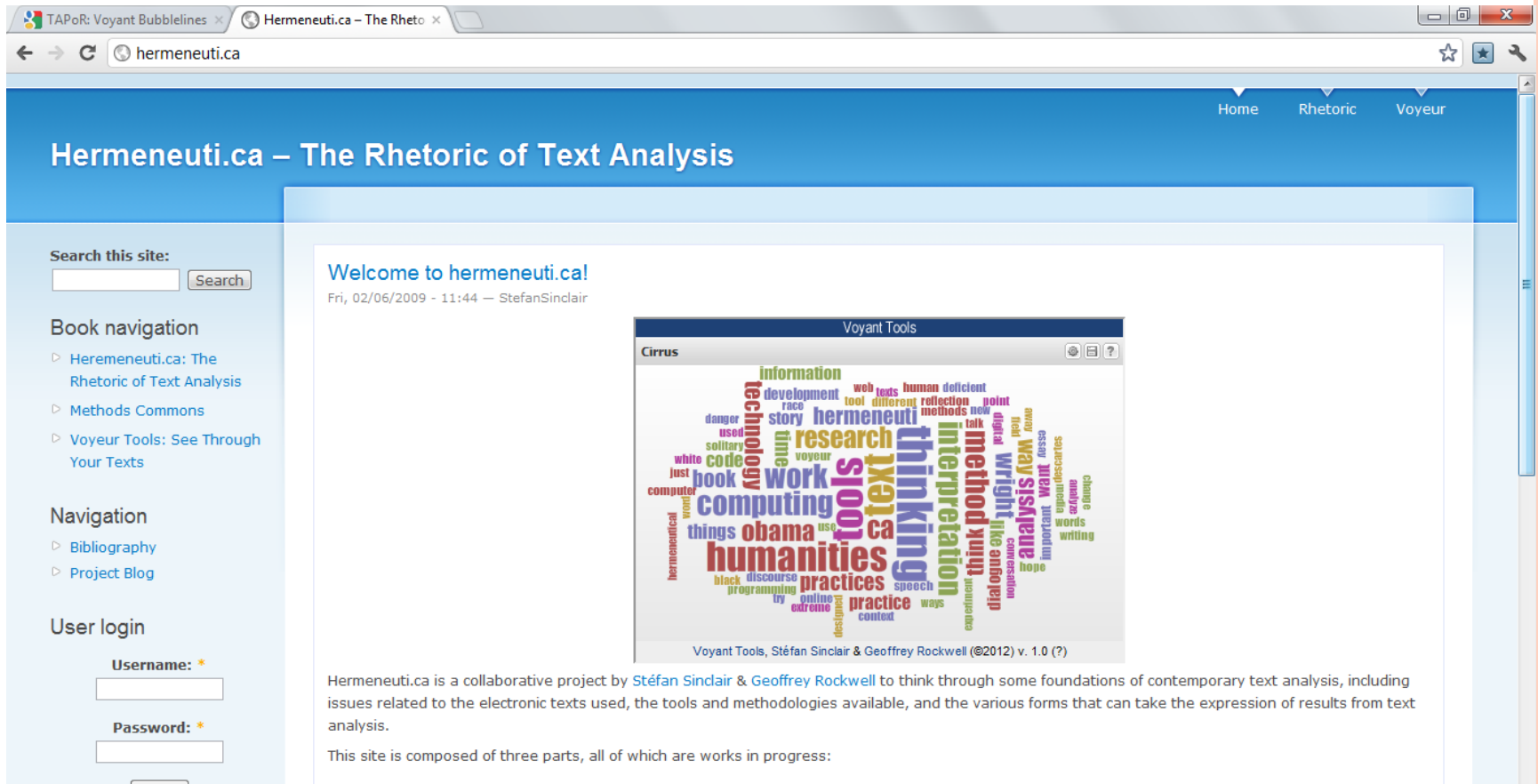


**Father Roberto Busa**  
**1913 - 2011**



# METHODS AND THEORIES

- Cold empiricism vs. interpretive humanities?
- “*process* rather than outcomes”



The screenshot shows a web browser window with the URL [hermeneuti.ca](http://hermeneuti.ca). The page title is "Hermeneuti.ca – The Rhetoric of Text Analysis". The navigation menu includes "Home", "Rhetoric", and "Voyeur". On the left sidebar, there is a search bar, a "Book navigation" section with links to "Hermeneuti.ca: The Rhetoric of Text Analysis", "Methods Commons", and "Voyeur Tools: See Through Your Texts", a "Navigation" section with links to "Bibliography" and "Project Blog", and a "User login" section with fields for "Username:" and "Password:". The main content area features a "Welcome to hermeneuti.ca!" message dated "Fri, 02/06/2009 - 11:44" by "StefanSinclair". Below the welcome message is a "Voyant Tools" window displaying a word cloud. The word cloud contains various terms related to text analysis and digital humanities, with the most prominent words being "thinking", "interpretation", "method", "analysis", "research", "work", "computing", "humanities", "practices", "text", "information", "development", "technology", "code", "book", "writing", "important", "wait", "change", "o", "thing", "words", "writing", "hope", "important", "wait", "analysis", "method", "interpretation", "research", "work", "computing", "humanities", "practices", "text", "information", "development", "technology", "code", "book", "writing", "important", "wait", "change", "o", "thing", "words", "writing", "hope". Below the word cloud, the text reads "Voyant Tools, Stéfán Sinclair & Geoffrey Rockwell (@2012) v. 1.0 (?)".

Hermeneuti.ca is a collaborative project by [Stéfán Sinclair](#) & [Geoffrey Rockwell](#) to think through some foundations of contemporary text analysis, including issues related to the electronic texts used, the tools and methodologies available, and the various forms that can take the expression of results from text analysis.

This site is composed of three parts, all of which are works in progress:

# TAPoR: TEXT ANALYSIS PORTAL FOR RESEARCH

The screenshot displays the TAPoR website interface within a browser window. The browser's address bar shows the URL [portal.tapor.ca/portal/portal](http://portal.tapor.ca/portal/portal). The page title is "Text Analysis Portal for Research".

**Navigation and Login:** The top navigation bar includes the TAPoR logo and a "Main" link. On the right, there is a login section with fields for "Username" and "Password", a "Login" button, and a "Sign Up for an Account Here" link.

**Welcome Section:** A "Welcome" box contains the text: "Welcome to TAPoR. TAPoR is a gateway to tools for sophisticated analysis and retrieval, along with representative texts for experimentation." It lists "Features of TAPoR" (Manage electronic texts, Experiment with online text tools, Learn about digital textuality) and provides a "Sign Up for an Account Here" link.

**Try It Section:** A "Try It" box guides users through three steps: 1. Choose a text to analyze (options: Universal Declaration of Human Rights, Moby Dick (Chapter 1), Hyperliste (Des vilains), French Medieval Poetry, Your Web Page), 2. Choose a tool to run (options: View, List Words, Explore with Voyeur, Find a Word), and 3. Run Tool on Text (with a text input field containing "http://").

**TAPoR Tools:** A "TAPoR Tools" box allows users to browse or search for tools. It includes a search bar and a "Search" button. The tool list includes: Browse by type, Browse by Source Format, New And Popular, and All A-Z.

**TAPoR Texts:** A "TAPoR Texts" box allows users to browse or search for text. It includes a search bar and a "Search" button. The text list includes: Browse by Tags, Browse by File Format, and Browse All.

**TAPoR News:** A "TAPoR News" box features a "General" section with several news items: "TAPoR 2.0 being Developed", "18th Connect and Voyeur", and "Portal Down Occasionally".

**Electronic Texts:** An "Electronic Texts" box provides "Links to other electronic text collections:".

**Credits:** A "Credits" box lists "Partner Institutions".

# WHY VOYEUR, WHY NOW?

- Voyeur is a web-based text analysis environment
- “user-friendly, flexible and powerful”
- “recognize current user needs and anticipate future practices”



# VOYANT

see through your text

Add Texts

Type in one ore more URLs on separate lines or paste in a full text.

reveal

Upload Open

3 words and  
202), of (146),  
nd

Corpus Reader

time." This time we want to talk about the crumbling schools that are stealing the future of black children and white children and Asian children and Hispanic children and Native American children. This time we want to reject the cynicism that tells us that these kids can't learn; that those kids who don't look like us are somebody else's problem. The children of America are not those kids, they are our kids, and we will not let them fall behind in a 21st century economy. Not this time.

This time we want to talk about how the lines in the Emergency Room are filled with whites and blacks a frequency: 18 who do not have health care; who don't have the power on their own to overcome the special interests in Washington, but who can take them on if we do it together.

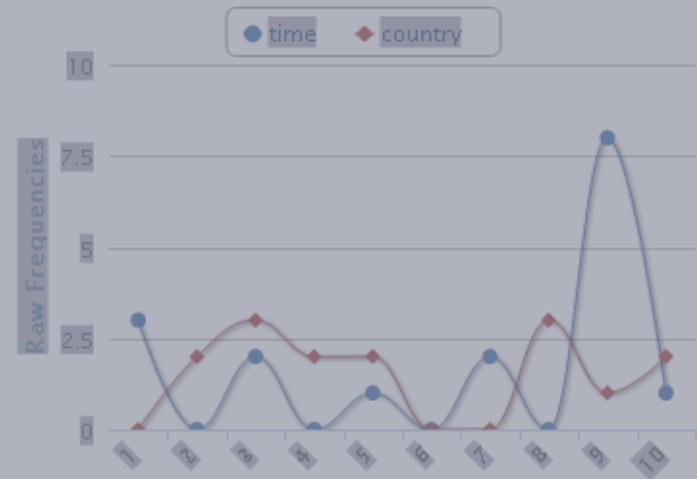
This time we want to talk about the shuttered mills that once provided a decent life for men and women of every race, and the homes for sale that once belonged to Americans from every religion, every region, every walk of life. This time we want to talk about the fact that the real problem is not that someone who doesn't look like you might take your job; it's that the corporation you work for will ship it overseas for nothing more than a profit.

This time we want to talk about the men and women of every color and creed who serve together, and fight together, and bleed together under the same proud flag. We want to talk about how to bring them home from a war that never should've been authorized and never should've been waged, and we want to talk about how we'll show our patriotism by caring for them, and their families, and giving them the benefits they have earned.

time

Corpus

Word Trends



Relative Segments time

Keywords in Context

Left	Keyword	Right
<b>Document: 1) "We the people, in order to fo...</b>		
...together and say, "Not t	time	." This time we want to
say, "Not this time." This	time	we want to talk about
...and Native American ch	time	we want to reject the
...st century economy. No	time	.This time we want to
... economy. Not this time.	time	we want to talk about
we do it together.This	time	we want to talk about
every walk of life. This	time	we want to talk about
more than a profit.This	time	we want to talk about

Page 1 of 1 Context Preview

Words in Documents



- <http://voyant-tools.org/tool/Bubblelines/?corpus=1332385205746.3169&bins=180&type=black&type=white&type=time&type=committed&type=time&type=time&docId=d1332303072504.29a87bf5-8feo-dc12-2070-8da0791e065d&typeFilter=black&typeFilter=white&typeFilter=time&typeFilter=committed>
- <p>Sinclair, S. and G. Rockwell (2012). Word Trends. <i>Voyant</i>. Retrieved March 2, 2012 from <http://voyeurtools.org/tool/TypeFrequenciesChart/></p>
- <http://voyeurtools.org/tool/TypeFrequenciesChart/?corpus=1332383721132.4585&docIdType=d1332303072504.58f1a948-db6a-7360-3107-7154e152e676%3Atime&docIdType=d1332303072504.58f1a948-db6a-7360-3107-7154e152e676%3Acountry&mode=document&freqsMode=raw>







# YING CHEN'S ENTIRE BODY OF WORK THROUGH TA?



# TAPoR 2.0

## TAPoR

TAPoR is for discovering research tools for studying texts.

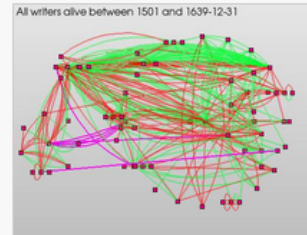
- Browse for tools by type or tag
- Search for tools and use them
- Comment on tools and read reviews

Words	Distribution	Count
great	Shakespeare	290
country	Shakespeare	202
time	Shakespeare	168
people	Shakespeare	141
came	Shakespeare	141
'that		137
little	Shakespeare	135
master	Shakespeare	134
said	Shakespeare	123
took	Shakespeare	122
feet	Shakespeare	114
good	Shakespeare	112
majesty	Shakespeare	111
like	Shakespeare	110
gave	Shakespeare	102

List Words - HTML (TAPoRware)



Wordle



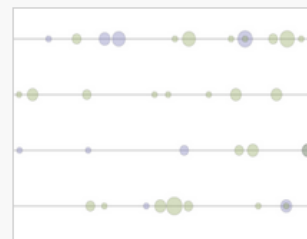
OrlandoVision (OVIS)



Voyant Cirrus



Voyant Links



Voyant Bubblelines

4) 1594 Merchant of Venice	22,310	3,330
5) 1594 Midsummer Night's Dream	17,195	3,045
6) 1594 Taming of the Shrew.txt	22,177	3,309
7) 1598 All's Well That Ends Well.txt	24,500	3,580
8) 1598 As You Like It.txt	22,954	3,321
9) 1599 Merry Wives of Windsor.txt	23,781	3,317
10) 1599 Twelfth Night or What You Will	21,459	3,143
11) 1600 Much Ado about Nothing	22,605	3,013

- **Longest documents** (by words): 1599 Henry 1593 King Richard III (31,663). Shortest documents: 1591 Errors (16,251), 1594 Midsummer Night's... (17,195). All...
- **Highest vocabulary density**: 1611 Tem 1605 Macbeth (186.7). Lowest density: 1593 King Richard I Much Ado about Nothi...; (133.3). All...
- **Most frequent words** in the corpus: the (28,323), and (24,20,468), to (19,220), it (17,626). More...
- **Words with notable peaks in frequency** across the corp... most (1,170), then (2,170), speak (1,176).

### Search

### Sort

- All (77)
- Concording (4)
- Miscellaneous (23)
- Statistical (30)
- Text gathering (5)
- Popular (-1)
- Editing (0)
- Search (33)
- Text cleaning (5)
- Visualization (22)

### Tags

### Comments

26, rue du Labrador A L'Affaire Tournesol L'Affaire Tournesol (film d'animation) Air des bijoux Archives Hergé Arumbayas Le ...

Comparator (HTML) is a free, web-based tool designed to run in a browser window. It is easy to use, designed to compare the text of two documents, either ...

Comparator (XML) is a free, web-based tool designed to run in a browser window. It is easy to use, designed to compare the text of two documents, either ...

Comparator (Plain Text) is a free, web-based tool designed to run in a browser window. It is easy to use, designed to compare the text of two documents, either ...