Denise A. D. Bedford, Ph.D.
Goodyear Professor of Knowledge Management
Kent State University
Kent Ohio
dbedfor3@kent.edu
Denise.Bedford4 (Skype)

SEMANTIC ANALYSIS OF THE POLITICAL DISCOURSE IN THE 2012 PRESIDENTIAL CAMPAIGN
Research Context

- Political scientists and linguists assume different perspectives when studying the language of politics

- Political scientists, according to Schaffner (2010) focus primarily on the consequences of language for decisions and actions, and the political realities that language creates

- Linguists are interested in the way that language is used to communicate the political message

- From both of these perspectives there is a common perception that the political discourse of the 2012 presidential campaign was emotional, negative, hostile, and aggressive in tone

- Perception is subjective – it is difficult to validate – that is the research challenge that we’ve taken on and report on today
Defining Negativity

• Today we are sharing with you the preliminary results of our research – we are still learning more and expanding the research project

• The major challenge we faced in this project was defining the nature of hostility, emotionality, aggressiveness and negativity

• We leveraged three well established dimensions:
  – Complex indicator of emotional state including anxiety, hostility, depression, social alienation, hope, etc.
  – Complex indicator of the use of concrete or abstract language
  – Simple indicator of references to specific entities and identifiers
Three Dimensions of Analysis

• We semantically evaluated three dimensions of the political discourse, including:
  – Mental and emotional states of the discourse
  – Extensional and intentional nature of the political language
  – References to concrete entities and identifiers

• Each analysis involved developing a profile or set of business rules grounded on human knowledge or expertise
  – Gottschalk-Gleser scale of psychological development
  – Raymond McLaughlin’s linguistic definitions of intentional and extensional language
  – Standard business rules for people, places, organizations and companies
Semantic Analysis Methods

• By semantic analysis we mean any method that combines human knowledge, knowledge engineering, natural language processing, concept or entity extraction and rule-based categorization

• We used three different tools to support our semantic analyses:
  – PCAD Software for Gottschalk-Gleser Scale
  – SAS Categorization Suite for psychological language characterization
  – Open Calais/ClearForest for entity extraction
Mental and Emotional State

• Analysis is based on the content analysis method originally developed by Gottschalk and Gleser (1969).

• This method measures the magnitude of any mental or emotional state or trait that can be clearly defined and categorized.

• It uses not simply words in isolation to classify content, but rather identifies relationships and attitudes reported by the subject.

• It also refines the score depending on whether the reference is applied to the speaker, to others (animate and inanimate) or if it is a denial of the concept.
Mental and Emotional State Factors

- Anxiety
- Hostility Inward
- Hostility Outward
- Ambivalent Hostility
- Social Alienation

- Depression
- Quality of Life
- Cognitive Impairment
- Human Relationships
Intentional and Extensional Language

• Intentional and extensional language provides an indication of different structures and approaches to conveying knowledge and ideas (Korzybski 1938)

• Semanticists characterize intentional language as more abstract and conceptual, whereas extensional language tends to be more concrete and enumerative of real examples and properties

• In the political context, intentional language may be a sign of high supposition, political storytelling or supposition

• In the political context, extensional language may be a sign of factual-oriented dialog, of focus on specific events, people and things
## Intentional Language Features

<table>
<thead>
<tr>
<th>Semantics Type</th>
<th>Semantic Criteria</th>
<th>Semantic Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allness Language</td>
<td>Allness terms as represented by such terms as all, every, entire, whole, none</td>
<td>200 concepts comprised largely of single word adjectives and adverbs</td>
</tr>
<tr>
<td>Superlative Language</td>
<td>Superlative terms such as best, worst, most, least, only, matchless...</td>
<td>2,061 concepts – adjectives and compound adjectival phrases with superlative connotations</td>
</tr>
<tr>
<td>Identification Predication</td>
<td>“Is of” statements which represent identification of objects. Predication terms such as “roses are beautiful”</td>
<td>39 concepts representing both identification and predication semantics as these are difficult to distinguish by humans and challenging at machine level</td>
</tr>
<tr>
<td>Consciousness of Projection</td>
<td>Terms such as seems, appears, in my opinion....</td>
<td>53 concepts representing single and multi-word terms</td>
</tr>
</tbody>
</table>
# Extensional Language Features

<table>
<thead>
<tr>
<th>Semantic Type</th>
<th>Semantic Criteria</th>
<th>Semantic Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparative Language</td>
<td>Terms such as higher, lower, more, less</td>
<td>241 concepts representing single and multi-word comparators</td>
</tr>
<tr>
<td>Conditional Language</td>
<td>Terms such as if, but, except, perhaps, unless....</td>
<td>114 concepts representing single and multi-word conditionals</td>
</tr>
<tr>
<td>Quantitative Language</td>
<td>Quantifying terms or precise numerical designations such as 60, sixty, second, trillion</td>
<td>97 concepts of both alphabetical and numerical characterizations of quantity</td>
</tr>
<tr>
<td>Pseudo-Quantitative Language</td>
<td>Terms that loosely represent the idea of amount and size such as many, much, few, lots...</td>
<td>73 concepts representing single and multi-word representations of amounts</td>
</tr>
</tbody>
</table>
Entities and Identifiers

• Entities and identifiers are things – people, places, organizations and companies

• References to entities and identifiers are an indication of use of concrete language

• The density of references to concrete things may indicate issues- rather than opinion-based content

• May also reflect popular culture in the intensity of coverage of people and organizations versus ideas or issues
Entity Extraction Methods

• Four entities were tracked:
  – People
  – Places
  – Organizations
  – Companies

• Semantic analysis and entity extraction leverages both natural language processing capabilities and embedded business rules

• Open Calais’s publicly available capability was leveraged – it contains a very robust set of business rules – and supports validation through display and highlighting of extracted entities
RESEARCH QUESTIONS
Research Questions

1. Is there a way to use semantic analysis methods to objectively determine whether the political discourse is emotionally charged, hostile, negative and aggressive?

2. If we can develop objective and verifiable semantic methods, what patterns might we find?
   1. Are patterns consistent across media spectra?
   2. Are patterns consistent across political spectra?
   3. Are patterns consistent across media and politics by spectrum?
Data Sets and Sources

• **Media Sources**  
  *(20 dialogs each = 340)*

  *Source: LexisNexis and Factiva*

  • Al Jazeera (1 source)  
  • BBC (1 source)  
  • CNN (5 Hosts)  
  • FOX (5 Hosts)  
  • MSNBC (5 Hosts)

• **Political Sources**  
  *(20 dialogs each = 200)*

  *Sources: UCSB Presidents Project + CSPAN*

  • Bachman  
  • Cain  
  • Gingrich  
  • Huntsman  
  • Paul  
  • Pawlenty  
  • Perry  
  • Obama  
  • Romney  
  • Santorum
What Did We Find?

- Let’s walk through each of the analyses and consider the research results and the implications of these results
Dimension 1:
MENTAL AND EMOTIONAL STATE ANALYSIS
Mental and Emotional State Results

- Two most notable findings are the consistently high rates of depression and the consistently low rates of quality of life language. No wonder we’re depressed when we listen to the news or the political dialog!

- What was surprising were the higher levels of indicators of hostility in the dialog of politicians – we had expected the levels of hostility to be greater in the media.

- For all other scales there were no noticeable variations between the dialog of politicians and the dialog of the media.

- We did see variations and we believe that using a semantically represented version of a well defined metric like Gottschalk-Gleser is a reliable approach.

- We continue to test this scale against a larger and broader corpus.
Politicians discourse is higher in anxiety than is the media discourse.

Bill O’Reilly appears to be the outlier amongst the media sources here.
Politicians discourse is higher outward projected hostility than is the media discourse.

Bill O’Reilly appears to be the outlier amongst the media sources here.
Politicians discourse is much higher in inward projected hostility than is the media discourse.

Brett Beier appears to be the outlier amongst the media sources here.
Politicians discourse is higher in ambivalent hostility than is the media discourse.

Bill O’Reilly and Brett Beier appear to be the outliers amongst the media sources here.
Politicians and media discourse are fairly consistent though Pawlenty appears to be an outlier.

GG Normal Score = 4
5 = Slightly High
6 = Moderately High
7 = Very High
Politicians and media discourse are both at the high end of the depression scale. This is a very interesting result.
Politicians and media discourse are both at the extreme low end of the Quality of Life scale. This is a very interesting result.

GG Normal Score = 4

5 = Slightly High
6 = Moderately High
7 = Very High
Dimension 2:

ABSTRACT VERSUS CONCRETE LANGUAGE PATTERNS
Extensional Language Indicators

• Significant differences were observed for three indicators
  – Comparative language use
  – Conditional language use
  – Quantifying language use

• In each significantly different instance, the Media demonstrated higher rates of use of Extensional Language.

• No significant differences were observed for pseudo-quantifying indicator
Significantly Different Averages

Politicians = 18.48
Media = 23.42

Again High level usage by Gingrich, Obama, Pawlenty & Romney. Obama aligned with Gingrich.

High level usage by “Left” and CNN. Low level usage by “Right”.
Again, high level usage by Gingrich, Obama, Pawlenty & Paul. Obama aligned with Gingrich.

High level usage by “Left” low level usage by “Right”
Pseudo-Quantifying Language

Gingrich again at the high end, with Paul.

High rates of use by CNN
Low rates of use by FOX
Mid-level use by MSNBC

Not Significantly Different Averages

Politicians = 6.54
Media = 8.36
Significantly Different Averages

Politicians = 8.05
Media = 19.78

Again High level usage by Gingrich, Obama, Pawlenty & Paul. Obama aligned with Gingrich.

High level usage by “Left” and CNN in the mid-range. Low level usage by “Right”
Intentional Language Use

• Every indicator demonstrates a significant difference between the media and the politicians.

• In the use of intentional language, the media demonstrates higher rates of use than do politicians.
Significantly Different Averages

Politicians = 11.42
Media = 14.56

Again High level usage by Gingrich and Obama.

High level usage by “Left” and CNN. Low level usage by “Right”.
High level usage by Gingrich, Bachman and Cain.

High level usage by “Left” and CNN. Low level usage by “Right”
High usage by Gingrich, Obama, Pawlenty & Paul. Obama aligned with Gingrich again here.

High level usage by “Left” and CNN. Low level usage by “Right”
Significantly Different Averages

Politicians = 18.52
Media = 24.43

Again High level usage by Gingrich, Obama and Romney.

High level usage by CNN, the “Left” in the middle, and the “Right” at the low range.
Dimension 3: REFERENCES TO IDENTIFIERS AND ENTITIES
Significantly Different Averages

Politicians = 11.59
Media = 16.68

Gringrich is very high in his references to people. But no politician comes close to the media.

On the Media site, the “Left” and CNN are on the high side. “Right” media is on the low use side.
Significantly Different Averages

Politicians = 18.52
Media = 24.43

Gringrich is very high in his references to places with Pawlenty, Romney and Obama close behind.

On the Media site, the “Left” and CNN are on the high side. “Right” media is on the low use side.
Almost identical patterns noted for Organizations.

Significantly Different Averages

Politicians = 6.18
Media = 10.00
Significantly Different Averages

Politicians = 1.51
Media = 9.87

Generally consistent patterns observed here. Exception is that politicians reference to companies is extremely low. This makes intuitive sense.
RESEARCH RESULTS AND OBSERVATIONS
Review of Research Questions

1. Are patterns evident across media spectra?
   By and large yes. We saw consistent trends by Media perspective.

2. Are patterns evident across political spectra?
   Because we had only the presidential contenders to analyze and this represented only one person from the Democratic Party versus many more from the Republican side, we don’t have a full picture.

3. Are patterns evident across media and politics by spectrum?
   By and large, the results seem to align across all three dimensions
Semantic Analysis is Viable Approach

• Semantic analysis methods when modeled on strong and validated models of human knowledge can help us test our perceptions of the nature of discourse

• In the past we’ve applied this approach to the religious context, the human resource management approach and to diplomatic information

• The approach also appears to hold for political discourse

• The research we’ve shared today is preliminary – we continue to test the methodology in a larger corpus and with additional examples derived from both Media and Politicians
Delivery is The Same But Content Varies

• Mental and emotional aspects suggest that politicians are more extreme than are the media. If the political dialog is hostile, depressing and the source of anxiety, it is more likely due to the discourse of politicians than it is to the discourse of the media.

• The Media’s use of both intentional (abstract) and extensional (concrete) language is more intense than Politicians.

• Superlative, allness, comparator and conditional language use registered higher levels across the two groups than did quantitative, pseudo-quantitative, identification-predication or projection language. These may align well with an appeal to our emotions.

• References to concrete entities and identifiers is much richer in the Media discourse than it is in the Politicians’ discourse. This was an unexpected result. We had expected the politicians’ discourse to contain more substance but that was clearly not the case.
Questions and Discussions

THANK YOU!