

# Taking Campaign Strategy Online

*Using Candidate Websites to Advance the Study of Issue Emphasis*

Andrew Therriault  
*New York University*

American Political Science Association  
*2010 Annual Meeting and Exhibition*

September 3, 2010

# Motivation

How do political candidates choose which issues to emphasize in campaigns?

- Nearly all existing theories predict that candidates will emphasize most favorable issues
- Empirical evidence shows that competing candidates often discuss the same issues
- I propose that while candidates may discuss the same issues, they do so in very different ways
- To show this requires better data: not just which issues are discussed, but also how

# Overview

- Prevailing Theories of Issue Emphasis
- Alternative Hypotheses
- Benefits of Automated Website Coding
- Coding Walkthrough
- Sample Results
- Discussion

# Prevailing Theories of Issue Emphasis

- Issue Ownership (Ansolabehere and Iyengar 1994, Petrocik 1996): Candidates emphasize issues on which their parties have better long-term reputations
- Positional advantages (Sellers 1998, Damore 2004, Sides 2006, Hillygus and Shields 2008): Candidates emphasize issues on which their positions are preferred
- Performance (Sellers 1998, Holian 2004, Vavreck 2009): Candidates emphasize issues on which their records or those of their parties' are particularly good
- Common theme is emphasis of favorable issues to increase salience
- Does not account for evidence of issue convergence (Sigelman and Buell 2004, Kaplan, Park, and Ridout 2006)

# Alternative Hypotheses

- “Riding the Wave” (Ansolabehere and Iyengar 1994): Candidates emphasize salient issues, are punished if not.
- Framing (Holian 2004, Chong and Druckman 2007): Candidates focus voters’ attentions on most favorable issue components
- Voter learning, persuasion, and cueing (implied by other fields): Candidates emphasis issues in order to provide information to voters about themselves or their opponents, or seek to manipulate voter opinion either through argument or partisan cues

# Data Requirements

- Previous research mostly based on analyses of news reports, speeches, and TV ads
- Good for telling *which* issues are emphasized, but not so good for *how* they're emphasized
- Limited quality and quantity of content in an ad or a second-hand report
- Especially a problem for sub-presidential races
- Availability of content across candidates certainly non-random
- Needs: More and better content, useful data on House and Senate races, minimal selection bias

# Why Websites Work

The main benefits of using websites as indicators of issue emphasis stems from their low cost:

- Amount of content on a webpage virtually unlimited, not dependent upon resources
- Produced directly by campaign, good proxy for overall emphasis strategy
- Nearly every candidate maintains a site—not only House and Senate, but also state and local
- Since 2002, almost all of these sites archived by Library of Congress

# Coding Methods

The downside of this bountiful data is that there's so much of it.

- Thousands of candidate sites, with multiple versions of each
- Site structures vary widely, and some take full advantage of the lack of space limitations
- Most useful dataset produced so far hand-codes sites (see Druckman et al 2010), limited by resource constraints

By using automated text analysis, can include far more races, as well as dynamic analyses of changing emphasis over the campaign season.

# Sample Data Coding

To illustrate the process, I code sites of major-party Senate candidates in 2002.

- 31 races, 52 candidates
- Captured October 15 versions of home pages, biographies, issue pages and subpages, news pages, as well as the most recent press releases, speeches, and ad transcripts (979 pages total = average of about 19 pages per candidate)
- Coded emphasis into 4 issue categories (economy, public services, social, and defense)
- Also recorded position/valence framing, party/candidate framing, and levels of detail

# Text Analysis Algorithms

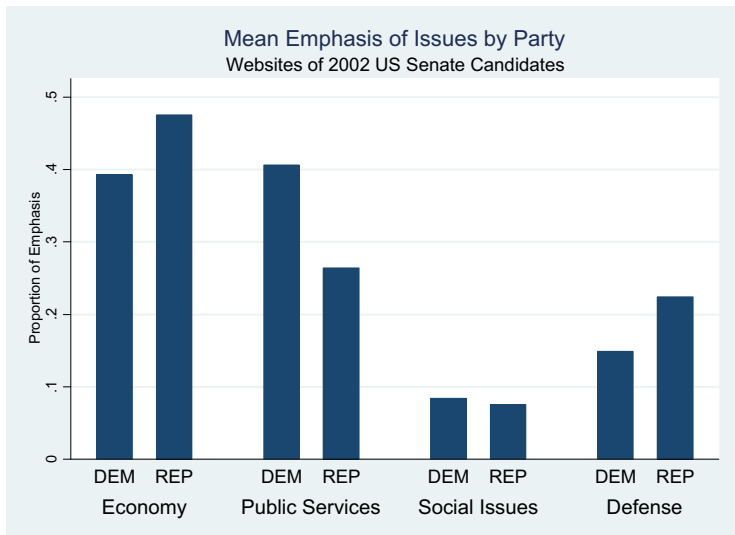
Basic steps (coded using Text Classification Tools developed for Pennsylvania Policy Database Project):

- 1 Collect and organize sites (HTML files)
- 2 Convert HTML to sentence database
- 3 Hand-code training data
- 4 Input training data to build classification model, then automatically classify “virgin” text
- 5 Weight sentences according to location on page and site, then reaggregate to site level

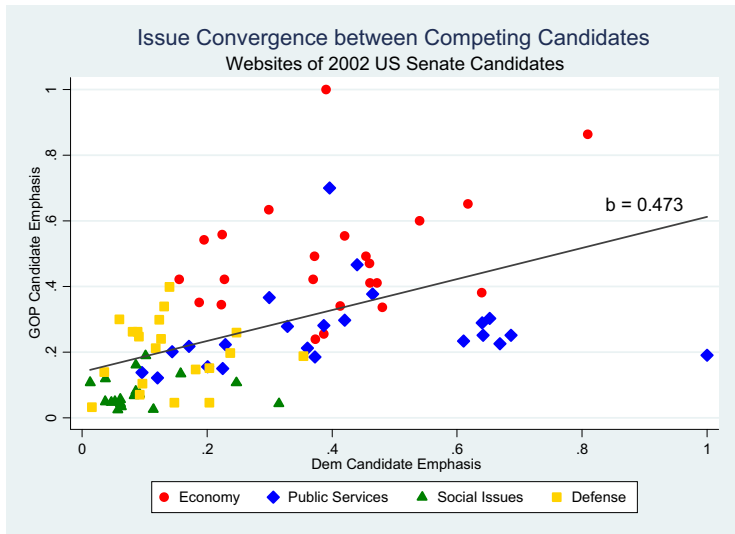
# Accuracy of Sample Coding

- Began with 519 hand-coded samples; used this training data to classify additional cases; 636 more added via iterative coding process (Moody et al, forthcoming); left with 1155 hand-coded observations out of approximately 35k to be coded.
- Rotating sample to use 1000 training observations to predict other 155:
  - 75.3% PCP for issue categories (58.9% PRE)
  - 67.6% PCP for position/valence framing (44.5% PRE)
  - 82.6% PCP for party/candidate framing (63.3% PRE)
  - 64.0% PCP for level of detail (low/med/high) (39.4% PRE)

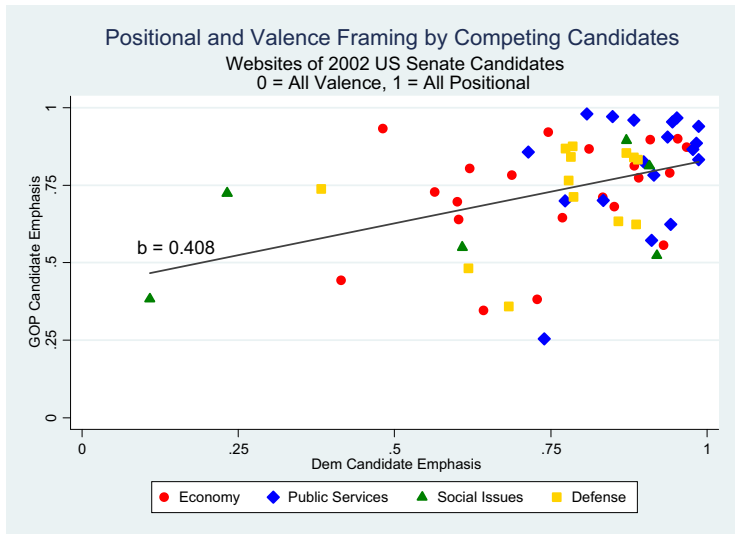
# Issue Emphasis by Party



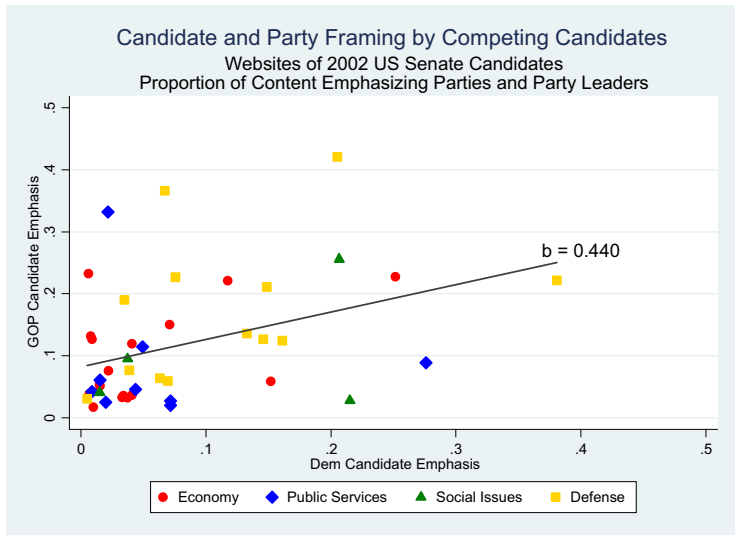
# Issue Emphasis by Competing Candidates



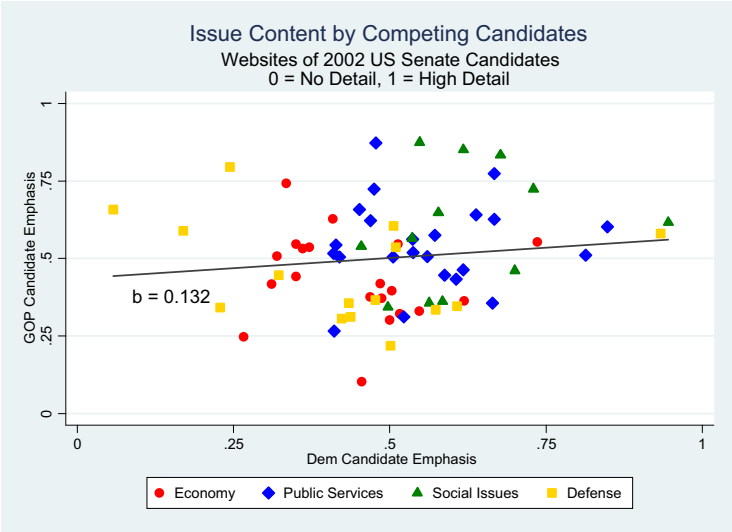
# Positional and Valence Framing



# Candidate and Party Framing



# Differences in Levels of Detail



# Discussion

- To reiterate: results here are only for illustrative purposes
- What they illustrate: potential of automated text analysis to code websites
- But much caution needed before going completely hands-off
- Even with automation, no magic bullet: lots of work required before data is ready for prime-time
- In any case, data not much value without similarly-innovative theory to apply it to.