The First History of the 2008 US Presidential Campaign
Modeling and Measuring Election Discourse

Karl Grossner, PhD Candidate
UC Santa Barbara Department of Geography
Geography and History

• Geography and history are different ways of looking at the world, but they are so closely related that neither one can afford to ignore or even neglect the other. (Baker, 2003 p.2)

• Many more historical geographers than ‘geographical historians’ but there is a ‘spatial turn’ under way.

• HGIS is a growing field, as is ‘digital humanities’

Towards a *digital* historical atlas
Modeling and Measuring

Modeling

- Representing complex social events in databases, to support
  - Discovery
  - Analysis
  - Reasoning
  - Mapping
  - Other visualizations
- For historical scholarship and education
  - Digital historical atlases

Measuring

- Variation in “issue aboutness” of election discourse
  - By region
  - By individual
  - By party
  - By media type
An Interactive Search Environment for Domain-Specific Knowledge Acquisition via Multimedia Data

Karl Grossner, Geography
Jonathan Ventura, Computer Science
Ben Adams, Computer Science
Angus Forbes, Media Arts & Technology
Monica Bulger, Education
Swapna Joshi, Elec & Computer Engineering
# Representing temporality in GIS

(a radically oversimplified distinction)

## CA COUNTIES

<table>
<thead>
<tr>
<th>name</th>
<th>shape (polygon)</th>
<th>t1</th>
<th>t2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alameda</td>
<td>0101000020E6100000F2B5679604F05D</td>
<td>1867</td>
<td>1909</td>
</tr>
<tr>
<td>Alameda</td>
<td>00000F2B56796DC0E3620E6100001000</td>
<td>1910</td>
<td>2008</td>
</tr>
</tbody>
</table>

## EVENTS

<table>
<thead>
<tr>
<th>name</th>
<th>shape (polyline)</th>
<th>t1</th>
<th>t2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Napoleon’s Russian campaign - advance</td>
<td>0101000020E6100000F2B5679604205</td>
<td>1812-06-23</td>
<td>1812-09-14</td>
</tr>
<tr>
<td>Napoleon’s Russian campaign - retreat</td>
<td>B56796042001010000F257C0E36BCF2</td>
<td>1812-10-19</td>
<td>1812-12-20</td>
</tr>
</tbody>
</table>
Napoleon’s Campaigns – In Fact

Armies
- Corps (#’s, nationalities)
- Generals

Events
- Campaigns
- Battles (date, locs, outcomes)
- Fire
- River crossings

Marches
- Flow (duration, #’s)

States/conditions
- Bivouac

Weather
- Temperature
Primitives of geographic knowledge
Golledge (1995)

needed: semantic reference system
cf. Kuhn, Raubal, Janowicz et al

<table>
<thead>
<tr>
<th>name</th>
<th>shape (polyline)</th>
<th>t1</th>
<th>t2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Napoleon’s Russian campaign - advance</td>
<td>01010000020E6100000F2B5679604205</td>
<td>1812-06-23</td>
<td>1812-09-14</td>
</tr>
<tr>
<td>Napoleon’s Russian campaign - retreat</td>
<td>B56796042001010000F257C0E36BCF2</td>
<td>1812-10-19</td>
<td>1812-12-20</td>
</tr>
</tbody>
</table>

Napoleon’s Russian campaign - advance
01010000020E6100000F2B5679604205
1812-06-23 1812-09-14

Napoleon’s Russian campaign - retreat
B56796042001010000F257C0E36BCF2
1812-10-19 1812-12-20

25 November 2008 spatial@ucsb brown bag
To describe an event...

- **participants**
  - individuals
  - groups
  - legal entities
  - wasPresent
    - wasParticipant
    - wasLead

- **products**
  - hadResult

- **influence**
  - influencedBy
    - causedBy
    - required

- **motivation**

- **location**
  - place

- **EVENT**
  - occurredAt
  - hadTimespan
  - hasType
  - hadPurpose

- **purpose**
  - general
  - specific

- **identity**
  - name

- **other events**
  - artifacts
  - ideas
  - people

25 November 2008
spatial@ucsb brown bag
Ultimately modeling processes

- **Process** (series of related events/activities)
  - beginning, formation, development, emergence
  - ending, collapse, destruction, disintegration
  - recovery
  - transition, change (of attributes, of identity), reshaping
  - expansion, growth (spatial or not)
  - contraction, diminishment
  - dispersal (abandon)
  - diffusion (spread from)
  - ascendency (of power, importance), rise, resurgence
Spatial Extents of Election 2008
(partial)

speeches
- Democrat
- Republican

other appearances
- Democrat
- Republican

debates
- Democratic Debate
- Republican Debate
Temporal Components of Election 2008
(partial)
RDF/OWL

Classes
- Actor
  - Person
  - Group
- Information object
  - linguistic object
    - speech text
    - news report
    - blog entry
- Temporal entity
  - Event
    - political activity
    - fundraiser
- Place
  - administrative division (adm)
  - populated place (ppl*)

Properties (relations)
- isMemberOf/hasMember
- isCreationOf/hasCreated
- hadParticipant/participatedIn
- hasPart/isPartOf
- tookPlaceAt/witnessed

<subject,predicate,object>
<Obama, isMemberOf, DemocraticParty>
<DemConvention, hasPart, nomination 
  tookPlaceAt, Denver.CO.US>

tookPlaceAt
domain: Event range: Place

history demands: attestedBy; assertedBy
### CIDOC-CRM

**85 class declarations**

**148 property declarations**

<table>
<thead>
<tr>
<th><strong>Property Name</strong></th>
<th><strong>Entity - Domain</strong></th>
<th><strong>Entity - Range</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>E1 CRM Entity</td>
<td>E41 Appellation</td>
</tr>
<tr>
<td>P2</td>
<td>E1 CRM Entity</td>
<td>E55 Type</td>
</tr>
<tr>
<td>P3</td>
<td>E1 CRM Entity</td>
<td>E62 String</td>
</tr>
<tr>
<td>P4</td>
<td>E2 Temporal Entity</td>
<td>E52 Time-Span</td>
</tr>
<tr>
<td>P5</td>
<td>E3 Condition State</td>
<td>E3 Condition State</td>
</tr>
<tr>
<td>P7</td>
<td>E4 Period</td>
<td>E53 Place</td>
</tr>
<tr>
<td>P8</td>
<td>E4 Period</td>
<td>E19 Physical Object</td>
</tr>
<tr>
<td>P9</td>
<td>E4 Period</td>
<td>E4 Period</td>
</tr>
<tr>
<td>P10</td>
<td>E5 Event</td>
<td>E77 Persistent Item</td>
</tr>
<tr>
<td>P11</td>
<td>E5 Event</td>
<td>E39 Actor</td>
</tr>
<tr>
<td>P12</td>
<td>E4 Period</td>
<td>E70 Thing</td>
</tr>
<tr>
<td>P15</td>
<td>E7 Activity</td>
<td>E19 Physical Object</td>
</tr>
<tr>
<td>P16</td>
<td>E9 Move</td>
<td>E4 Period</td>
</tr>
<tr>
<td>P25</td>
<td>E11 Modification</td>
<td>E4 Period</td>
</tr>
<tr>
<td>P31</td>
<td>E24 Physical Man-Made Thing</td>
<td>E29 Design or Procedure</td>
</tr>
<tr>
<td>P32</td>
<td>E7 Activity</td>
<td>E19 Physical Object</td>
</tr>
<tr>
<td>P33</td>
<td>E7 Activity</td>
<td>E4 Period</td>
</tr>
<tr>
<td>P92</td>
<td>E63 Beginning of Existence</td>
<td>E77 Persistent Item</td>
</tr>
<tr>
<td>P93</td>
<td>E64 End of Existence</td>
<td>E77 Persistent Item</td>
</tr>
<tr>
<td>P15</td>
<td>E7 Activity</td>
<td>E1 CRM Entity</td>
</tr>
<tr>
<td>P19</td>
<td>E7 Activity</td>
<td>E71 Man-Made Thing</td>
</tr>
<tr>
<td>P20</td>
<td>E7 Activity</td>
<td>E7 Activity</td>
</tr>
<tr>
<td>P32</td>
<td>E7 Activity</td>
<td>E55 Type</td>
</tr>
<tr>
<td>P43</td>
<td>E70 Thing</td>
<td>E54 Dimension</td>
</tr>
<tr>
<td>P44</td>
<td>E18 Physical Thing</td>
<td>E3</td>
</tr>
<tr>
<td>P45</td>
<td>E18 Physical Thing</td>
<td>E57 Material</td>
</tr>
<tr>
<td>P46</td>
<td>E18 Physical Thing</td>
<td>E18 Physical Thing</td>
</tr>
</tbody>
</table>

**E1** CRM Entity

**E2** Temporal Entity

**E3** Condition State

**E4** Period

**E5** Event

**E7** Activity

**E8** Thing

**E70** Man-Made Thing

**E24** Physical Man-Made Thing

**E22** Man-Made Object

**E25** Man-Made Feature

**E78** Collection

**E28** Conceptual Object

**E73** Information Object

**E29** Design or Procedure

**E31** Document

**E33** Linguistic Object

**E36** Visual Item

**E55** Type

**E39** Actor

**E74** Group

**E21** Person

**E41** Appellation

**E42** Identifier

**E44** Place Appellation

**E49** Time Appellation

**E75** Conceptual Object Appellation

**E82** Actor Appellation
CIDOC-CRM  an ontology of cultural heritage information

Figure 3: A qualitative metaschema of the CIDOC CRM

CIDOC Conceptual Reference Model (ISO 21127:2006); Doerr, Crofts, Gill, Stead, Stiff
CIDOC-CRM

• Rudimentary reasoning:
  – sub-classes and sub-properties (isA inheritance)
  – property holds for domain with valid values in range
  – inverse and symmetric properties
  – cardinality constraints

<table>
<thead>
<tr>
<th>Property Name</th>
<th>Entity – Domain</th>
<th>Entity - Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>is identified by (identifies)</td>
<td>E1 CRM Entity</td>
</tr>
<tr>
<td>P2</td>
<td>has type (is type of)</td>
<td>E1 CRM Entity</td>
</tr>
<tr>
<td>P11</td>
<td>had participant</td>
<td>E5 Event</td>
</tr>
<tr>
<td>P15</td>
<td>was influenced by (influenced)</td>
<td>E7 Activity</td>
</tr>
<tr>
<td>P20</td>
<td>had specific purpose</td>
<td>E7 Activity</td>
</tr>
<tr>
<td>P127</td>
<td>has broader term (has narrower term)</td>
<td>E55 Type</td>
</tr>
</tbody>
</table>
US Election 2008

- Ontology developed in RDF/OWL using Protégé
- On a CIDOC-CRM scaffold
  - some problems, e.g. activity vs. action; overly complex time constructs
- Instantiated in a RDBMS (PostgreSQL)
  - a model isn’t useful if not used: compatible with GIS
  - dissimilar logics: relational algebra vs. description logic
  - numerous RDBMS tools: procedural languages; 3rd-party functions (spatial, tsearch, recursion); custom data types (e.g. temporal)
A more general model

**temporal entities**
- temps
  - PK t_id
  - t_type
  - alias
  - timespan_id

- events
  - t_id
  - event_type_id
  - t_type
  - event_name
  - timespan_id
  - place_id

**persistent items**
- actors
  - PK actor_id
  - actor_type
  - alias
  - tag

- concepts
  - PK concept_id
  - alias
  - concepttype_id

- issues
  - PK concept_id
  - issue_num
  - tags
  - category

- infobjs
  - PK infobi_id
  - type
  - title
  - author
  - timespan_id
  - source

**space-time_primitives**
- timespans
  - PK timespan_id
  - begin
  - end
  - interval
  - confidence

- places
  - PK place_id
  - name
  - alternate_names
  - latitude
  - longitude
  - timezone

**types**
- event_type
  - PK etype_id
  - type_name

- grouptype
  - PK gtype_id
  - type_name

- concepttype
  - PK ctype_id
  - type_name

- infotype
  - PK itype_id
  - type_name

**associations**
- assoc
  - PK assoc_id
  - subject
  - prop_id
  - object
  - timespan_id
  - attestedBy

**properties**
- properties
  - PK prop_id
  - name
  - domain
  - range
  - parent
  - inverse

One or several association tables store the bulk of knowledge represented, e.g.:
- similarity measures for documents
- group membership, for time-spans, with roles
- actor participation in occurrences
- authorship
- topics in information objects

Properties hold for a domain class (and sub-classes) with possible values in a range class (and sub-classes). Restrictions, including values and cardinality (some, always, at least, etc.) may be assigned. Properties represent the predicates in assertion tuples, e.g. (<group>, hasMember, <person>).
Measuring Election Discourse

- Candidate speeches
  - 227 Democrat (126 Obama)
  - 121 Republican (77 McCain)
- Debate transcripts
  - 21 Democratic, 19 Republican, 4 post-primaries
- Wire service news reports (geography-free? \(\rightarrow\) AP, Reuters, etc.)
  - \(~12,000\) between Nov 2006 and now
- Regional newspaper reports
  - \(~22,600\) reports from 36 newspapers in 9 census regions from Jan 2008
- Blog text since the conventions
  - Liberal: *Huffington Post, Daily Kos*
  - Conservative: *Hot Air, Right Wing News*
- Candidate issue statements
  - 577 statements in 58 issue categories \(\rightarrow\) 34 issues
Issue Signatures (1)

Service to America
Thank you. It's good to be back in Meridian. As you might know, I was once a flight instructor here at the air field named for my grandfather during my long past and misspent youth. And it's always good to be in Mississippi, which you could call my ancestral home. Generations of McCains were born and raised in Carroll County, on land that had been in our family since 1848. The last McCain to live on the property, which the family called Teoc, was my grandfather's brother, Joe McCain. I spent a couple summers here as a young boy, and enjoyed it immensely. I had never had a permanent address because my father's naval career required us to move frequently. But here, in the care of my very likeable Uncle Joe, I could imagine, with a little envy, what it must have been like for the McCains who came before me to be so connected to one place; to be part of a community and a landscape as well as a family.
Issue Signatures (2)

$$\text{Sim}(A, B) = \cos \theta = \frac{A \cdot B}{|A||B|} = \frac{x_1^*x_2 + y_1^*y_2}{(x_1^2 + y_1^2)^{1/2} (x_2^2 + y_2^2)^{1/2}}$$

### Issue Signatures (3)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>33</th>
<th>34</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NA</td>
<td>0.035628140</td>
<td>0.06556735</td>
<td>0.05571609</td>
<td>0.067639191</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.035628140</td>
<td>NA</td>
<td>0.01688337</td>
<td>0.01625446</td>
<td>0.016060756</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.065567346</td>
<td>0.016883374</td>
<td>NA</td>
<td>0.13380258</td>
<td>0.256157692</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.055716088</td>
<td>0.016254475</td>
<td>0.13380258</td>
<td>NA</td>
<td>0.293547250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.067639191</td>
<td>0.016060756</td>
<td>0.25615769</td>
<td>0.29354725</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.077649316</td>
<td>0.015998291</td>
<td>0.22398443</td>
<td>0.08083047</td>
<td>0.139628198</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.032271659</td>
<td>0.013927477</td>
<td>0.13693093</td>
<td>0.11941534</td>
<td>0.208675252</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.042118609</td>
<td>0.007685005</td>
<td>0.20621152</td>
<td>0.15201624</td>
<td>0.209649865</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.144460229</td>
<td>0.033069846</td>
<td>0.21138198</td>
<td>0.11920396</td>
<td>0.160019931</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Cosine Similarity

<table>
<thead>
<tr>
<th>issue 1</th>
<th>issue 2</th>
<th>...</th>
<th>issue 34</th>
<th>doc</th>
</tr>
</thead>
<tbody>
<tr>
<td>speech1</td>
<td>0.214</td>
<td></td>
<td>0.641</td>
<td></td>
</tr>
<tr>
<td>speech2</td>
<td>0.052</td>
<td>0.432</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>

#### Mean

| mean | 0.133 | 0.218 | 0.322 |

---

25 November 2008

spatial@ucsb brown bag
25 November 2008  spatial@ucsb brown bag
Results (3)

81 Democratic Swing State Speeches

55 Republican Swing State Speeches

25 November 2008
spatial@ucsb brown bag
## Swing state issue emphasis by party/region

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>families</td>
<td>27_0 0.05402</td>
<td>the middle class 05_6 0.02198</td>
<td>families 27_0 0.05317</td>
</tr>
<tr>
<td>education</td>
<td>06_0 0.044</td>
<td>global warming 08_1 0.02167</td>
<td>economy 05_0 0.04449</td>
</tr>
<tr>
<td>the middle class</td>
<td>05_6 0.0417</td>
<td>poverty 05_3 0.01635</td>
<td>the middle class 05_6 0.04395</td>
</tr>
<tr>
<td>poverty</td>
<td>05_3 0.04051</td>
<td>economy 05_0 0.01436</td>
<td>health care 09_0 0.04123</td>
</tr>
<tr>
<td>economy</td>
<td>05_0 0.03548</td>
<td>families 27_0 0.01119</td>
<td>poverty 05_3 0.04006</td>
</tr>
<tr>
<td>spending</td>
<td>05_1 0.02647</td>
<td>women's rights 26_0 0.01078</td>
<td>education 06_0 0.03254</td>
</tr>
<tr>
<td>taxes</td>
<td>05_2 0.02568</td>
<td>iraq war 18_0 0.00883</td>
<td>veterans 13_1 0.02852</td>
</tr>
<tr>
<td>service</td>
<td>30_0 0.02077</td>
<td>agriculture 29_0 0.00831</td>
<td>social security 05_5 0.02306</td>
</tr>
<tr>
<td>health care</td>
<td>09_0 0.01898</td>
<td>veterans 13_1 0.00801</td>
<td>global warming 08_1 0.02216</td>
</tr>
<tr>
<td>agriculture</td>
<td>29_0 0.0157</td>
<td>service 30_0 0.00726</td>
<td>women's rights 26_0 0.02209</td>
</tr>
<tr>
<td>voting reform</td>
<td>23_0 0.01417</td>
<td>energy policy 07_0 0.00618</td>
<td>taxes 05_2 0.02085</td>
</tr>
<tr>
<td>research spending</td>
<td>05_9 0.01373</td>
<td>education 06_0 0.00548</td>
<td>agriculture 29_0 0.01652</td>
</tr>
<tr>
<td>social security</td>
<td>05_5 0.01262</td>
<td>faith and politics 24_0 0.00378</td>
<td>spending 05_1 0.01619</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Issue</th>
<th>Environment 08_0 -0.01094</th>
<th>Defense Policy 21_0 -0.00911</th>
<th>Global Image 16_0 -0.00828</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy Policy</td>
<td>07_0 -0.01497</td>
<td>Political Reform 14_0 -0.01179</td>
<td>War on Terror 19_0 -0.00853</td>
</tr>
<tr>
<td>Defense Policy</td>
<td>21_0 -0.03475</td>
<td>Homeland Security 10_0 -0.01209</td>
<td>Abortion 01_0 -0.00918</td>
</tr>
<tr>
<td>Homeland Security</td>
<td>10_0 -0.03777</td>
<td>Health Care 09_0 -0.01435</td>
<td>Homeland Security 10_0 -0.00965</td>
</tr>
<tr>
<td>Iraq War</td>
<td>18_0 -0.04939</td>
<td>Disabilities 32_0 -0.01682</td>
<td>Defense Policy 21_0 -0.01255</td>
</tr>
<tr>
<td>Veterans</td>
<td>13_1 -0.06461</td>
<td>Spending 05_1 -0.03112</td>
<td>Foreign Policy 20_0 -0.01529</td>
</tr>
<tr>
<td>Foreign Policy</td>
<td>20_0 -0.08161</td>
<td>Taxes 05_2 -0.04628</td>
<td>Judicial Activism 12_0 -0.04434</td>
</tr>
</tbody>
</table>
The Final Four
McCain/Palin & Obama/Biden

MEASURING

25 November 2008 spatial@ucsb brown bag
Appalachia

MEASURING

25 November 2008 spatial@ucsb brown bag
MEASURING

Salient terms (stemmed) from newspaper articles, extracted and weighted with Latent Dirichlet Allocation (LDA) clustering algorithm

Northern Appalachia
29 Elmira Star-Gazette
82 Binghamton Press & Sun-Bulletin
535 Erie Times-News
125 Scranton Times-Tribune
503 Charleston Gazette

Southern Appalachia
299 Knoxville News Sentinel
80 Asheville Citizen-Times
137 Winston-Salem Journal
21 Huntsville Times
34 dimensions $\rightarrow$ 2 w/MDS

77 speeches

44 Obama, B.
3 Biden
2 Obama, M.
20 McCain
8 Palin
Travels through (semantic) space

77 speeches
44 Obama, B.
3 Biden
2 Obama, M.
20 McCain
8 Palin
Rudolf Flesch (1948); *A new readability yardstick*, Journal of Applied Psychology, Vol. 32, pp. 221-233
Flesch-Kincaid (2)

Convention Speeches Flesch-Kincaid Readability

BY GRADE LEVEL
- Gettysburg Address
- Obama’s “Race Speech”
- Roosevelt 3rd Inaugural
- Lincoln 2nd Inaugural
- Kennedy Inaugural
- GW Bush 2nd Inaugural
- Clinton 2nd Inaugural
- M L King “I Have a Dream”
- Nixon Resignation

Democratic (27 speeches)  Republican (18 speeches)  Various

25 November 2008  spatial@ucsb brown bag
Flesch-Kincaid (3)

Average grade level
Democrat 10.3
Republican 10.6
Conclusions (Election 2008)

• There is geographic variation in issue aboutness insofar as candidates are concerned
• Democratic and Republican speech occupies distinct, marginally overlapping regions of semantic space
• Candidates discuss issues more than news reports do
• The Obama campaign was in fact rhetorically strategic and disciplined (and/or sitting on a lead)
• Some of my conceits were wrong (cf. McCain speechwriters’ reading level)
• Some were right
• The internet has complicated geography
Further Questions

• What are the actual issues discussed, e.g. change, leadership, Wright & Ayers, VP competency
  – analyze news reports, blogs

• Does an ontology-based data model have value for focused geo-historical analyses?
  – (not just interoperability, sharing, encyclopedic systems)
Comments, questions?

And thanks to NSF
IGERT Grant # DGE-0221713