Natural Language Processing for Social Sciences

Rada Mihalcea

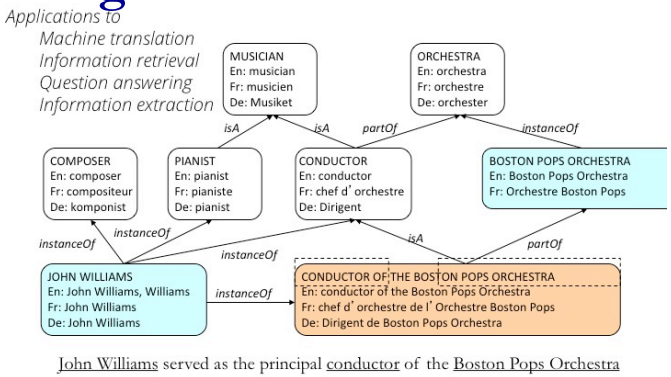
Computer Science and Engineering

University of Michigan

Language and Information Technologies

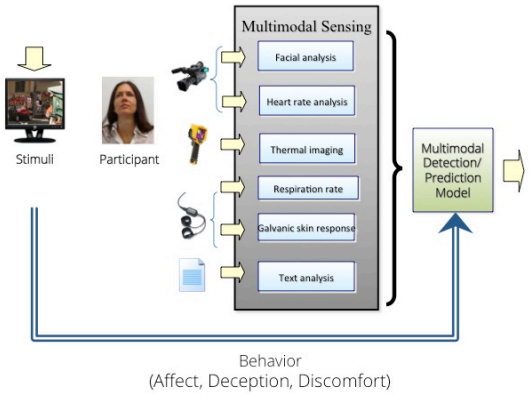
Faculty: Rada Mihalcea, **seven PhD students**: Mahmoud Azab, Aparna Garimella, Shibamouli Lahiri, Konstantinos Pappas, Steve Wilson, Charlie Welch, Laura Wendlandt, **three postdoctoral fellows**: Mohamed Abouelenien, Carmen Banea, Veronica Perez-Rosas, **M.S., undergrad students**.

Text semantics and linguistic resources



Big social media data

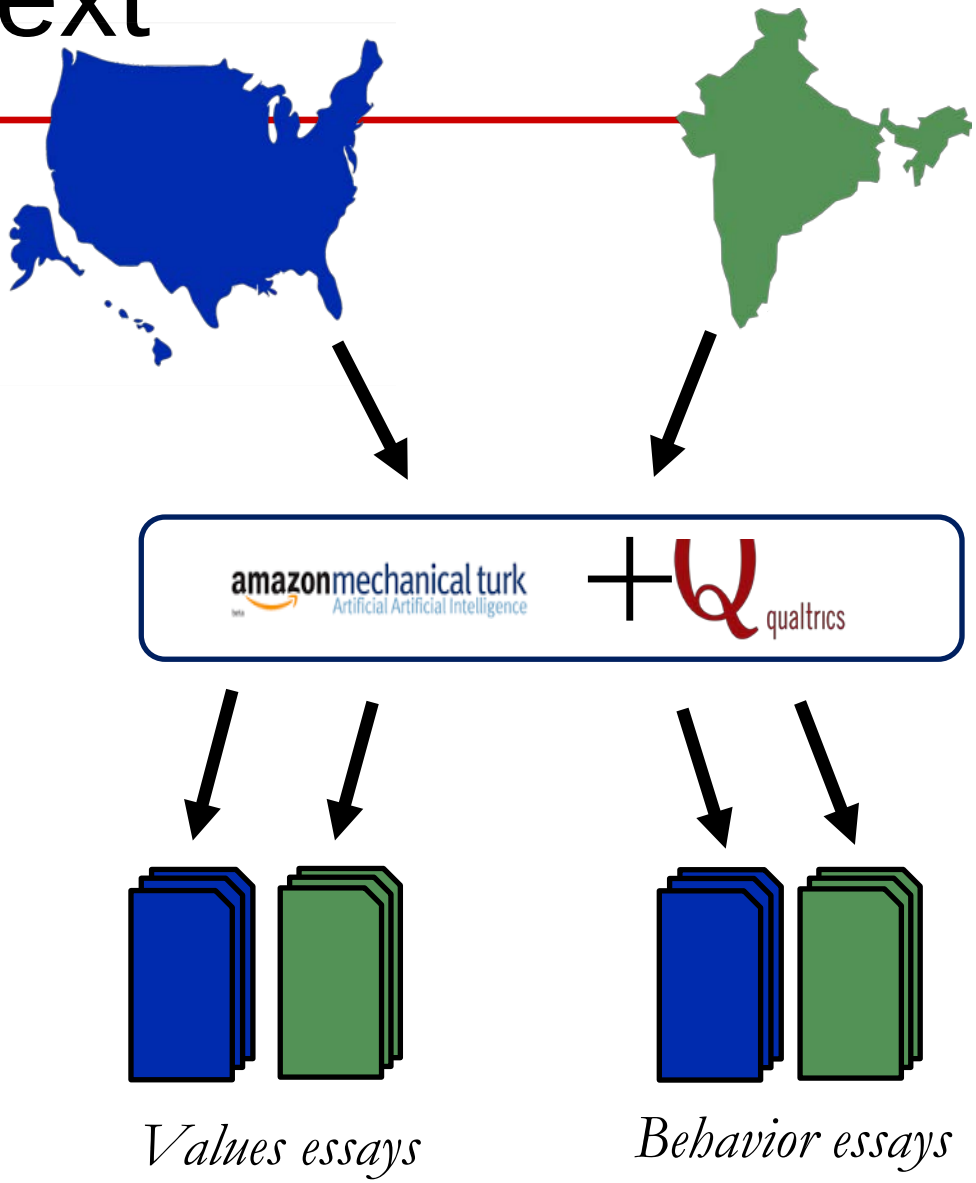
Multimodal models



Research in: **text semantics (**word/text similarity, large semantic networks), **behavior analysis** (multilingual opinion analysis, multimodal models for deception detection, affect recognition, alertness detection), **big data for cross-cultural analysis** (geotagging, understanding cross-cultural differences and worldview), **educational applications** (digital advising, PEGASE: pedagogical search engine, automatic short answer grading)

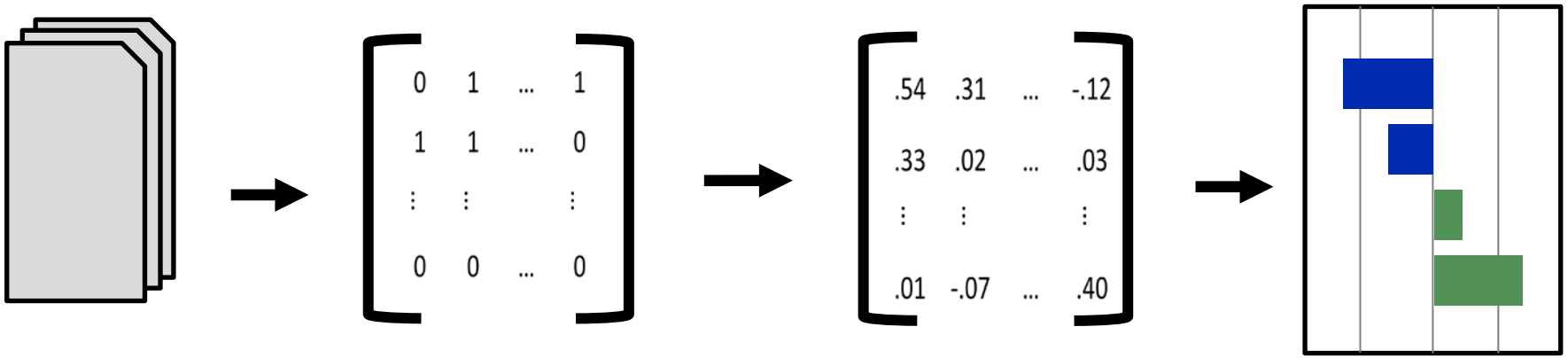
Values from Text

- What do people talk about when asked about their **values** and **everyday behaviors**?
- Can we learn about values and behaviors in different cultures from language use?



Methodology: topic models

- Find out what general topics, or themes, people mentioned in either the value or behavior essays, e.g., the **Meaning Extraction Method** (Chung and Pennebaker, 2008).



Written essays

Term-document matrix

PCA + varimax rotation

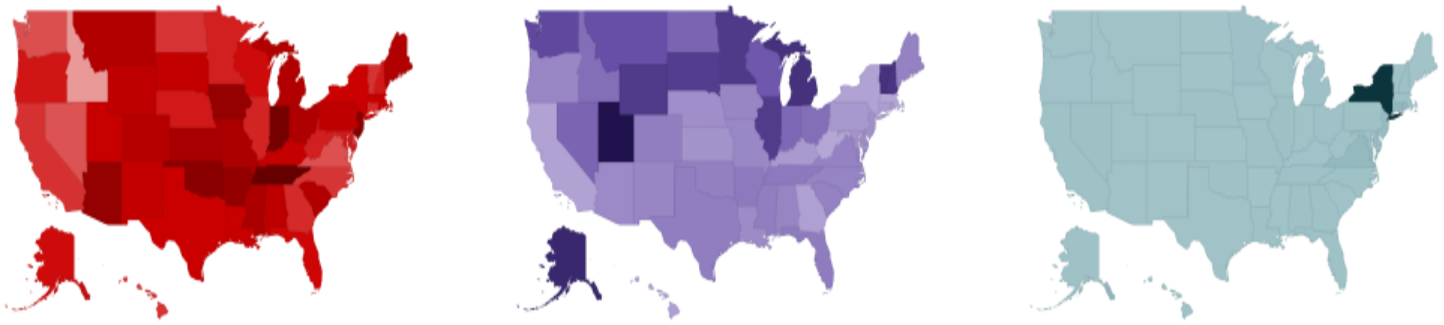
Quantified theme usage

Value Themes + Example Words

Theme	Example Words	LCR
Hard Work	job, work, hard	
Financial	spend, money, time	
Respect	moral, respect, person	
Faith	faith, god, belief	
Understanding	understand, teach, right	
Honesty	honest, truth, sure	
Familial Love	love, family, child	
Relationships	friend, family, people	
Sharing Thoughts	see, speak, mind	
Caring	compassion, kind, equal	
Positivity	happy, feeling, great	
Rule Following	rule, follow, society	
Family Support	support, provide, husband	
Decision Making	decision, future, consider	
Peaceful	human, peace, respect	

Location from Text

Hypothesis: the geographical variability of language can be exploited in order to construct improved geolocation models.



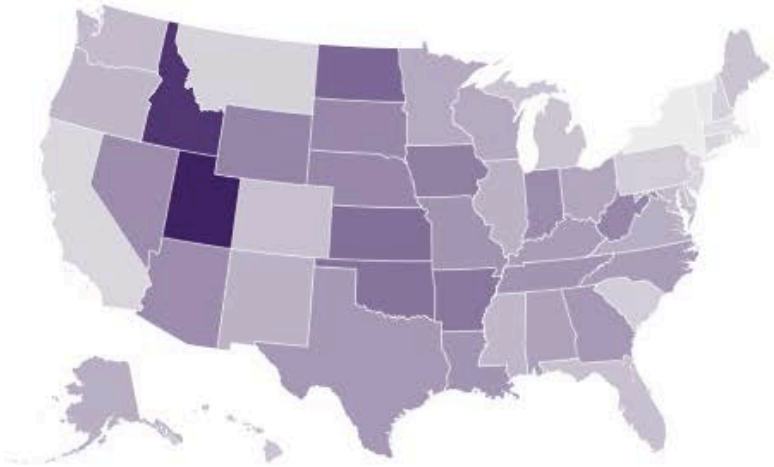
Distribution of three selected words across the 50 U.S. states.

$$WLH(w) = \max_{s \in S} \frac{P(w|s)}{P(w)}$$

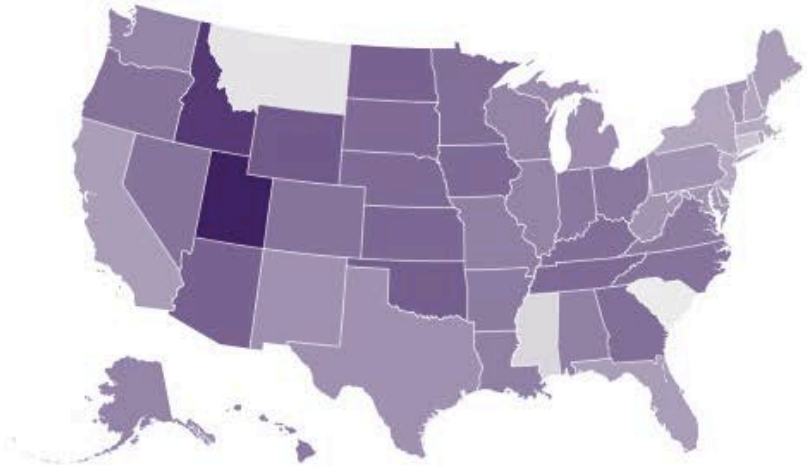
Psycho-linguistic Maps

Correlation: 0.84

FAMILY



POSITIVE EMOTION



Cross-cultural Analysis of Worldview

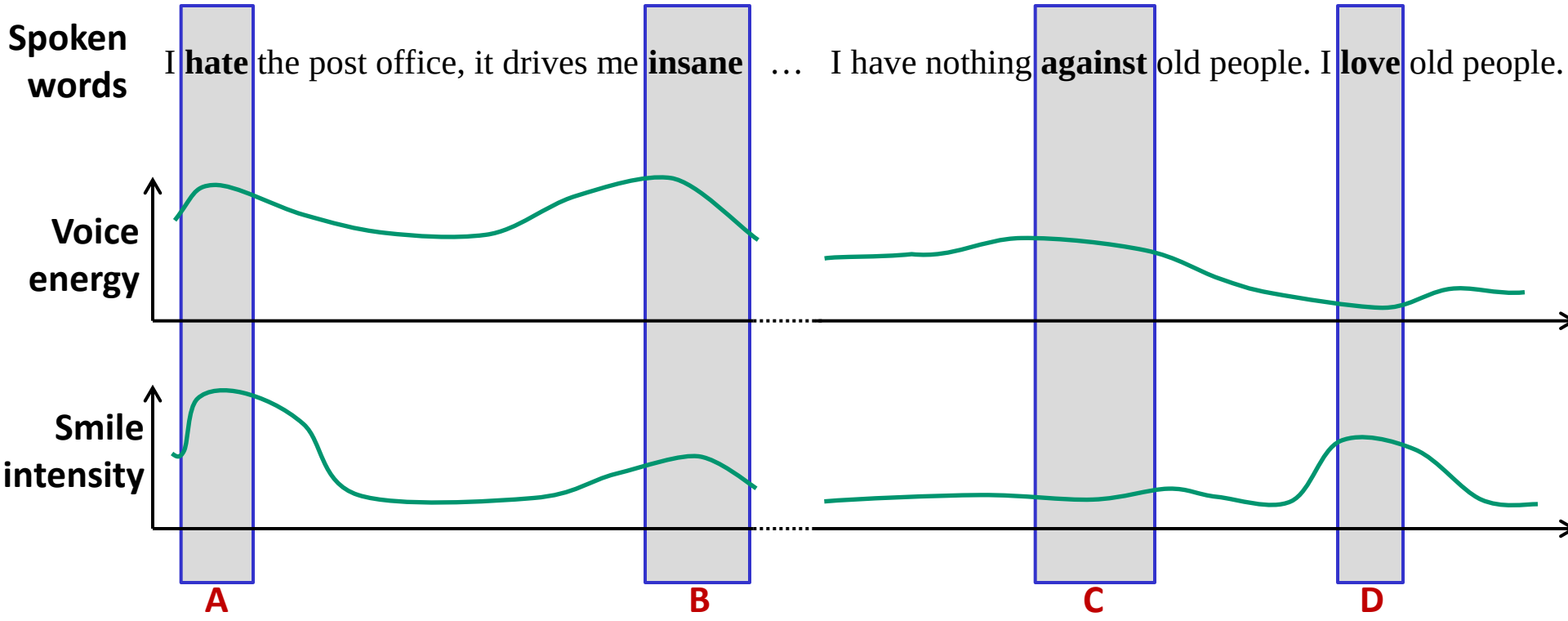
(Mexicans)
The Sun is:
strong
big

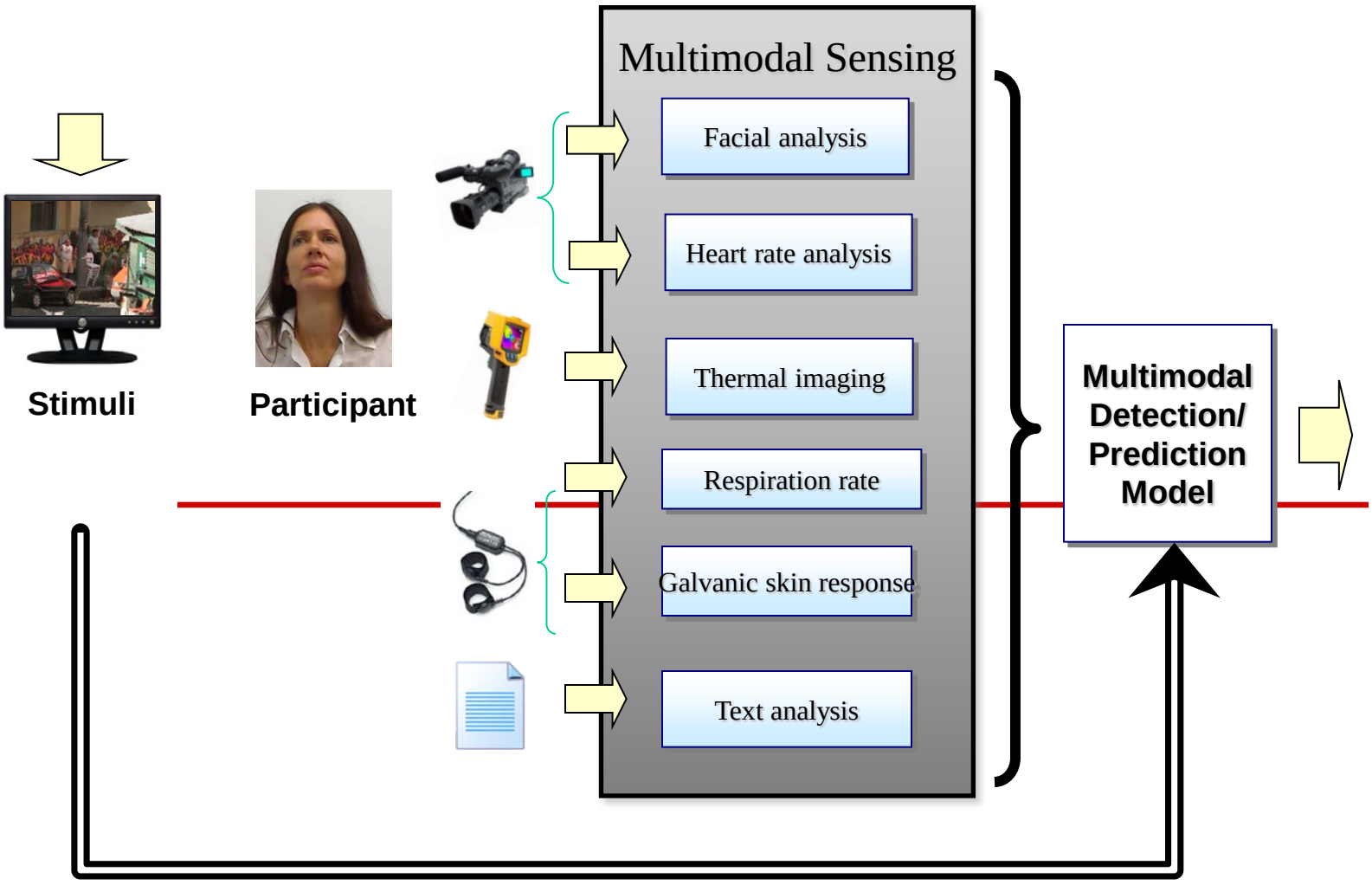
(Germans)
The Sun is:
warm
nice

- Analysis of linguistic expressions to identify differences in worldview
- Machine learning framework on word usage instances
  - Are there differences in how Mexicans and Germans perceive the Sun?
  - How is achievement regarded in Singapore versus U.S.?

Multimodal Behavior Analysis




Multimodal Sensing
Facial analysis
Heart rate analysis
Thermal imaging
Respiration rate
Galvanic skin response
Text analysis
Stimuli
Participant
Multimodal Detection/ Prediction Model
Behavior
(Affect, Deception, Stress, Discomfort)