Research design in computational text analysis

Jae Yeon Kim

15 April, 2019

Research design

- Understanding the strengths and limitations of computational linguistic (text) analysis
- No data mining; theory-driven questions and measurements

Research question

- Pierre Bourdieu: “opposing similar in structure to those found in cultural practices also appear in eating habits” (distinction)
- Jurafsky et al. “Linguistic Markers of Status in Food Culture: Bourdieu’s Distinction in a Menu Corpus”. *Journal of Cultural Analysis*. 2016.

Computational linguistic approach

- "Language offers a powerful tool for **observing and quantifying the sometimes unconscious way** that our associations and understandings of **culture** reflect our **social attitudes and prejudices**" (Jurafsky et al.)

- Social desirability bias (Everybody lies)
- Preference falsification (by Timur Kuran, Actions don't speak louder)

Distinction

- Authenticity (natural, traditional)
- Educational capital
- Plenty
- Implicit signaling of quality

Corpus

- Downloaded from allmenus.com in 2011 for restaurants in Boston, Chicago, Los Angelos, New York, Philadelphia, San Francisco, and Washington D.C. (N = 45,108)
- Used Yelp to validate restaurants and bars (no delis!) and add more information (e.g., price range) (N = 6511)

Some interesting findings

- Experience restaurants: natural authenticity, educational capital (a lot of French, Italian, Japanese words), less adjectives
- Cheap restaurants: traditional authenticity, planety (i.e., portion size), more adjectivecs

Studying ideas and culture

- A new systematic way to study ideas and culture
- Building bridges between qualitative and quantitative fields