

Latent semantic analysis of a Brexit survey experiment

Annika Fredén

Karlstad University

Society for Political Methodology Annual Meeting

Provo, July 19 2018

When are numbers insufficient for analysing political behavior?

- Split between political sides, threats
- **Brexit 2016**, US presidential election 2016, 2018 Swedish parliamentary election ...
- Relationships between campaign information and decision-making
- Latent semantic analysis - *what is new and what is special?*
- Forthcoming chapter in:
Statistical Semantics: Methods and Applications
edited by Sikström, S., Willander, J. and D. Garcia (Springer)

A Brexit Survey Experiment (Bäck et al., 2017)

- Randomized experiment in an Internet Survey Panel a couple of weeks before referendum 2016 (Prolific Academic), N=400, relatively young and well-educated sample
- Death Treatment (high threat) - Respondents asked to describe their feelings about dying after describing what they think happens to their body when they die (Greenberg et al., 1990)
- Dentist Treatment (control) - Respondents asked to describe their feelings about a dentist visit
- Small differences at the aggregate - a variable concerning social risk taking indicated individual level differences

Method

- **Latent Semantic Analysis:** Statistical analysis of underlying meanings of words (Landauer, T. K., & Dumais, S. T, 1997)
- Google-n-gram semantic space (co-occurrence matrix). Singular value decomposition applied to matrix which transforms text data to normalised vectors (compare factor analysis). A similar procedure applied to the experimental text data.
- **Cluster-method:** k-means algorithm maximises similarities between words in clusters (in this case, k=6)
- Individuals' closeness to cluster centroid transformed to numeric value and included in logistic regression model
- More fine-grained and data-driven than the dictionary-based LIWC

Death Threat Treatment

- " Now we would like to know what you think happens when you die. Below are two questions where we are interested in your thoughts. It is your gut reaction and spontaneous thoughts we are interested in, so please write the first thing that comes to your mind. There is no right or wrong answer. Describe in as much detail as possible what you think will happen physically to your body when you die.... What emotions are you feeling when you think about your own death?"

Dentist Treatment (control)

- "Now we would like to know what you think happens when you go to the dentist. Below are two questions where we are interested in your thoughts. It is your gut reaction and spontaneous thoughts we are interested in, so please write the first thing that comes to your mind. There is no right or wrong answer. Describe in as much detail as possible what you think happens when you go to the dentist.... What emotions are you feeling when you think about going to the dentist?"

Follow-up item death threat (Bäck et al., 2017)

How did you feel when you were answering the previous questions? Please write 5 words to describe this. For example, if you feel positive then write more and stronger words describing this, and if you feel negative then write more and stronger words describing that. Write only one descriptive word in each box:

Word 1

Word 2

Word 3

Word 4

Word 5



Analysing small data from a survey experiment

- Pros
 - Fewer words = less noise
 - Easier to analyse with latent semantics
- Cons
 - Potential loss of meaning and irony
 - Not "naturally occurring" data

Word clusters under Brexit campaign following death threat (illustration from Bäck et al., 2018, estimations performed in www.semanticexcel.com)



Findings

- Small differences between groups at the aggregate
- Large variation within treatment groups
- Feelings of anxiety after threat as revealed by the text analysis increases the likelihood that a voter chooses the status quo option

Concluding Remarks

- Relationship between threat and status quo behavior, given that the individual experiences anxiety
- Example of inductive approach leading to new knowledge
- How does the construction of survey items affect the results?
- When is cluster-based methods preferable to maximum likelihood estimations such as for example structural topic models (Roberts et al., 2014)? When are dictionary-based approaches preferable?