Custom Dictionaries in Computer-Aided Text Analysis

Aaron F. McKenny

Computer-Aided Text Analysis

- Word choice provides valuable information in the context of a particular organizational narrative.

- Process of hundreds of documents quickly with extremely high reliabilities.

- Complementary to manual coding

- Many tools
  - CAT Scanner
  - DICTION
  - LIWC
  - …

How CATA Works

- Dictionary-based coding
  - Completely automated, computer does the coding
  - Dictionaries (lists of words) are created and validated prior to analysis
- Computer looks for words from the dictionary in the narratives being analyzed
  - When it finds a word, it increments the value for that dictionary by 1

Example of Dictionary-based coding

- Simple dictionary: *Innovativeness*
  - "Innovative" "Innovation" "Innovate" "Research" "Inventions" "Inventive" "Creative" "Creativity"

- Simple narrative to analyze:
  - "The creativity of our research and development team make this organization one of the most innovative in the industry, with patents on over 2,300 inventions."

- Computer-aided text analysis result:
  - Innovativeness: 4

Standard vs. Custom Dictionaries

- Standard dictionaries
  - Dictionaries that ship with the software
  - Developed/validated by others

- Custom dictionaries
  - You provide the list of words to tabulate
  - As valid as you make it

Developing A Valid Custom Dictionary

- Same forms of validity apply to CATA as to most other measures
  - Internal
  - Discriminant
  - Content, etc.

- Two-step approach
  - Deductive
  - Inductive

Developing A Valid Custom Dictionary

- Use multiple raters to determine what words to include
  - Assess interrater reliability
  - Construct vs method experts?

- Narrative matters
  - Context of words
  - Level of analysis

Useful Dictionary Creation Tools

- Text file cleaner

```
PAST, PRESENT AND FUTURE
The second half of fiscal 2009 and a promising
future wouldn t have been possible without
tremendous work by talented teams fully engaged in
```

- CAT Scanner (built-in functionality)

- Inductive word list generation
  - DICTION 5 - list of insistence words
    - In DICTION 6 you can't copy-and-paste the list
  - CAT Scanner – generate inductive word list

Useful Dictionary Creation Tools

- Dictionary Judging Sheet/Reliability Calculator
  - You enter the words
  - Judges identify whether words fit construct or not
  - It calculates interrater reliability
  - Support for up to 3 judges, but easy to extend.
  - Available at: http://www.amckenny.com/CATScanner/resources.php

CATA Toolkit

- CAT Scanner
  - Free
  - http://www.amckenny.com/CATScanner/
- DICTION
  - $179
  - http://www.dictionsoftware.com/
- LIWC
  - $30 (lite); $90
  - http://www.liwc.net/

Useful Publications

- ** Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods, 13*, 320-347.

- ** McKenny AF, Short JC, Payne GT. (In Press). Using CATA to elevate constructs in organizational research: validating an organizational-level measure of psychological capital. *Organizational Research Methods*. doi: 10.1177/1094428112459910.

- Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A content analysis of the content analysis literature in the organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods*, 10, 5-34.

** - Related to CATA dictionary creation

Useful Publications

- Pennebaker, J., Mehl, M., & Niederhoffer, K. (2003). Psychological aspects of natural language use: Our words, our selves. Annual Review of Psychology, 54, 547-577.

- Morris, R. (1994). Computerized content analysis in management research: A demonstration of advantages & limitations. *Journal of Management*, 20, 903-931.

- Short, J. C., & Palmer, T. B. (2008). The application of DICTION to content analysis research in strategic management. *Organizational Research Methods*, 11, 727-752.

- Kabanoff, B. (1997). Computers can read as well as count: Computer-aided text analysis in organizational research. *Journal of Organizational Behavior*, 18(S1), 507-511.

Other Useful Resources

- CAT Scanner site
  - http://www.amckenny.com/CATScanner
- University of Georgia content analysis site
  - http://www.terry.uga.edu/management/content analysis/
- CARMA Short Courses
  - Dr. Jeremy Short gives a short course on CATA dictionary creation.