Extending the Utility of Content Analysis via the Scientific Method

Presentation to the Workshop on Social Theory and Social Computing
May 23, 2010

Kimberly A. Neuendorf, Ph.D. & Paul D. Skalski, Ph.D.
Cleveland State University
My history with Content Analysis

- Practitioner
- Methodological Generalist—CA part of a diversified toolbox
- Interdisciplinary
- Surveyor, educator
Presentation Outline

1-Content Analysis—Definitions and Main Options
2-Standards (a challenged methodology)
3-The Scientific Method and Content Analysis
   - Description/Prediction/Explanation
   - Validity
4-Focus on Computer-aided Text Analysis (CATA) Techniques
5-Focus on Social Networks (and Other Online Communications)
6-Translations/Conclusions
Early Origins of Quantitative Content Analysis

- Content of films (1930s, Dale)
- Political speech (1940s, Lasswell & Leites)
- Berelson’s (1952) defining work
- Violence on television (1960s)
- Psychometrics via content analysis (McClelland and others; 1940s on)
- CATA--Philip Stone’s General Inquirer at Harvard, 1965
Quantitative Content Analysis Defined

Content analysis is a summarizing, quantitative analysis of messages that relies on the scientific method, including attention to objectivity/intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and (deductive) hypothesis testing (Neuendorf, 2002).

It is not limited as to the type of messages that may be analyzed, nor as to the types of variables that might be measured (i.e., not just text)

- Nonverbals
- Paralanguage
- Images
- Audio/video streams
- Etc.
Two Main Options

Two main options exist for the execution of quantitative content analysis.

- **Human coding/Judge-based.** The most commonly used type is that which uses human coders to analyze message characteristics. An a priori coding scheme is devised by researchers, and the instrument is applied to message content by trained, reliable coders.

- **CATA.** The main alternative, using computer applications to analyze text (called CATA, or Computer-Aided Text Analysis), introduces a growing set of options for automated analyses.
However, human coding procedures are important for:

- The *origination* of content analytic schemes that eventually become CATA algorithms
- The measurement of highly *latent* constructs
- The ongoing *validation* of CATA measures
The Scientific Method and Content Analysis: Setting Standards

Content analysis has not always been held to the same standards as other quantitative methods

- Poor intercoder reliability assessment and reportage (most recently, 38% of health media human-coded content analyses included no reference to reliability whatsoever; Neuendorf, 2009)
- Analyses often limited to univariate, descriptive statistics
- Validation spotty for both human coding schemes and CATA
- Samples not always representative (poor generalizability)

Textbooks attempting to bring rigor to CA (Krippendorf, 1980; 2004; Neuendorf, 2002; Riffe, Lacy, & Fico, 2005)
The Functions of Science and CATA:

1. Description

- Occurrence, context and location of occurrence, co-occurrence, network analysis
2. Prediction

- Ability to statistically predict past or future conditions from message characteristics
2. Prediction

Source Characteristics → Message Characteristics → Receiver Outcomes
2. Prediction

Source Characteristics → Message Characteristics → Receiver Outcomes

Outcomes → Message Characteristics → Source Characteristics
For Prediction
- Collect source data or receiver data along with content analysis data on messages

For Later Inference
- Establish a “worn path” between message data and source or receiver data over replications
Prediction

e.g., Text-based (CATA) profiles of language have predicted deception on the part of the source (e.g., Newman et al., 2003; Zhou et al. 2004)
Prediction

e.g., Stylistics of chat (e.g., word length, punctuation, vocabulary richness) may predict authorship, source gender, internet domain (Kucukyilmaz et al. (2008))
Inference

e.g., Well-worn path predicting source
gender from particular language choices
(e.g., Chung & Pennebaker, 2007; 2009)
In a linking of content analysis and experimental findings, Franium, Seelfelt, and Vandello (2008) studied the prevalence of rape myth endorsements in online newspaper headlines regarding the 2003-2004 Kobe Bryant case, and then conducted an experiment that found male subjects to hold higher rape-supportive attitudes after exposure to myth-endorsing headlines identified via content analysis.
3. Explanation

- Introduction of *theory* – what are the underlying mechanisms driving relationships among variables?

- Debates in philosophy of science over independence of prediction and explanation
  
  - e.g., Symmetry thesis:
    
    - every successful explanation is a potential prediction, every successful prediction a potential explanation
      
      (Hempel, 1965; Ruben, 1990)
3. Explanation

Prediction without explanation—a hollow victory?

- e.g., Naccarato & Neuendorf’s (1998) significant prediction of print ad recall, with 59% of variance “explained”; predictors not consistent with prevailing advertising literature
- e.g., Predictions of gender from language use
CATA
Measurement in CATA

Three choices:

- Custom Dictionaries
  - Complicated, time-consuming

- Standard Dictionaries
  - A task of matching one’s conceptualization to someone else’s operationalization—sometimes a scavenger hunt
  - Similar to the challenge of finding an appropriate scale for a survey

- "Emergent" Coding—outcome based on language patterns that emerge (e.g., CATPAC)
Validity and CATA

- Validation part of development of CATA system (e.g., Lin et al., 2009—genres of online discussion threads)
- Validation of thematic CA (psychometrics) against self-report—rare and uncertain (e.g., McClelland et al., 1992)
- A comprehensive model for assessing content, external, and predictive validity when using CATA—Short, Broberg, Cogliser, Brigham (2010) as applied to “entrepreneurial orientation“:
  - Content validity—an inductive/deductive combo
  - External validity—use multiple sampling frames
  - Predictive validity—measure non-CATA variables that should relate
Validity of Standard Dictionaries

- Trusting the Standard Dictionary—an issue of face validity
  - Few CATA programs reveal the full dictionary lists (e.g., Diction, General Inquirer)
  - None reveal the full algorithm (including disambiguation (e.g., well, pot, leaves))
  - None account for negation

- Construct and Criterion Validity
  - Rod Hart’s Diction—“normed” rather than validated
  - Gottschalk and Bechtel’s PCAD—validated against standard psychiatric diagnoses
# Quantitative CATA Programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Author</th>
<th>Original Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBPro</td>
<td>M. Mark Miller</td>
<td>Newspaper articles</td>
</tr>
<tr>
<td>Yoshikoder</td>
<td>Will Lowe</td>
<td>Political documents</td>
</tr>
<tr>
<td>WordStat</td>
<td>Normand Peladeau</td>
<td>Part of SimStat, a statistical analysis package</td>
</tr>
<tr>
<td>General Inquirer</td>
<td>Philip Stone</td>
<td>General mainframe computer application (1960s)</td>
</tr>
<tr>
<td>Profiler Plus</td>
<td>Michael Young</td>
<td>Communications of world leaders</td>
</tr>
<tr>
<td>LIWC 2007</td>
<td>Pennebaker, Booth, &amp; Francis</td>
<td>Linguistic characteristics &amp; psychometrics</td>
</tr>
<tr>
<td>Diction 5.0</td>
<td>Rod Hart</td>
<td>Political speech</td>
</tr>
<tr>
<td>PCAD 2000</td>
<td>Gottschalk &amp; Bechtel</td>
<td>Psychiatric diagnoses</td>
</tr>
<tr>
<td>WORDLINK</td>
<td>James Danowski</td>
<td>Network analysis/communication</td>
</tr>
<tr>
<td>CATPAC</td>
<td>Joseph Woelfel</td>
<td>Consumer behavior/marketing</td>
</tr>
</tbody>
</table>
# Quantitative CATA Programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Type</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBPro</td>
<td>Word count/custom dictionaries only</td>
<td>N/A—all custom dictionaries</td>
</tr>
<tr>
<td>Yoshikoder</td>
<td>Word count/custom dictionaries only</td>
<td>N/A—all custom dictionaries</td>
</tr>
<tr>
<td>WordStat</td>
<td>Word count/custom dictionaries only</td>
<td>N/A—all custom dictionaries</td>
</tr>
<tr>
<td>General Inquirer</td>
<td>Word count with dictionaries</td>
<td>Dictionaries adapted from Harvard IV, Lasswell values, other standard linguistic and socio-psychological scales</td>
</tr>
<tr>
<td>Profiler Plus</td>
<td>Word count with dictionaries</td>
<td>Proprietary</td>
</tr>
<tr>
<td>LIWC 2007</td>
<td>Word count with dictionaries</td>
<td>Some dimensions have been validated against assessments by human judges</td>
</tr>
<tr>
<td>Diction 5.0</td>
<td>Word count with dictionaries</td>
<td>No—based on R. Hart’s substantive work</td>
</tr>
<tr>
<td>PCAD 2000</td>
<td>Word count with dictionaries</td>
<td>Long history of development of a human-coded scheme; both human &amp; CATA heavily validated against clinical diagnoses</td>
</tr>
<tr>
<td>WORDLINK</td>
<td>Word co-occurrence</td>
<td>N/A—emergent dimensions</td>
</tr>
<tr>
<td>CATPAC</td>
<td>Word co-occurrence</td>
<td>N/A—emergent dimensions</td>
</tr>
</tbody>
</table>
Online Communication
Online Communication

Online message content—an unprecedented amount of often naturally-occurred “speech”/text

“Cues filtered out” vs. “hyperpersonal” models of online communication (Walther, 2004; 2007)

A “place of refuge” for marginalized communicators (e.g., children (Kim et al., 2007), those with autism (Newton, Kramer, & McIntosh, 2009))
Online Communication

However, there have been found striking differences between computer-mediated communication (CMC) and non-CMC text (Abbasi & Chen, 2008):

- CMC richer in interaction
- CMC less topical
- CMC incorporates novel language varieties (including web-centric symbolic cues such as emoticons; Neviarouskaya, Predinger, & Ishizuka, 2007; 2009)
- CMC may be one-to-one or mass in intent (unique in communication)
Content Analyses of Online Communication

Early CATA studies examined web sites, blogs, postings

- Attempting to infer source affect (e.g., Abbasi et al., 2008; Lieberman & Goldstein, 2006; Neviarouskaya et al., 2009)
- Using stylometrics to identify authorship (e.g., Kucukyilmaz et al., 2008; Opoku et al., 2007)
Content Analyses of Social Networking

Until recently, most were human-coded

Studies of MySpace and Facebook, analyzing, e.g.:

- Personal disclosure (Jones et al., 2008)
- Pictorial self-presentation (Kane, 2008)
- Cross-national differences (Lunk, 2008)
- Ethnic and racial identities (Grasmuck et al., 2009)
- Use of social networking sites by 2008 U.S. presidential candidates (Compton, 2008)
CATA and Social Networking

- Few examples to date
  - Kucukyilmaz et al. (2008) used CATA to predict the identities of 100 online authors from language use in chat logs...but they had to use some human intervention to achieve 99.7% hit rate

- Barriers:
  - Social Networking sites incorporate images, video, etc.
  - Fluidity of content
  - Hidden features (e.g., long discussions on Facebook)
Translations

Scientific inquiry and advancement of knowledge in the area of computer processing of text is impeded by separate literatures among engineering, systems science, psychologists, communication scientists, linguists.

Terminology translations

- e.g., Lin et al. (2009) establish “coherence” between their automated classification system for online discussion threads in a school system and human ratings. . . it’s what I would call “validation”

Deductive/inductive mutual process (data mining, etc.)—not routinely accepted by social/behavioral scientists
Conclusion

The challenges to producing useful applications of content analysis that employ the standards of scientific inquiry, particularly within the context of automated analyses, are substantial.

However, the benefits of a scientific approach are even more substantial, including greater confidence in knowledge and the ability to explain findings and predict future outcomes in other contexts.

As Carl Sagan famously stated, science is a “candle in the dark” (Sagan, 1997).
end
Yoshikoder
About Yoshikoder

Created by Will Lowe at Harvard’s Department of Government

Can be downloaded free at www.yoshikoder.org

A cross-platform, multi-lingual CATA program

Must run one case at a time

Assumes the researcher will create dictionaries

Can import external dictionaries

Exports results into Excel
Yoshikoder: KWIC and Concordance

Documents

ID 1 I consider myself a contradiction. My outside does not match my inside... On the outside I am this older looking person with gray hair. But on the inside I feel half my age. I am not sure how much time has passed. I don't know if I want to get older. I spent my whole working life doing jobs I hated, even though I did them extremely well. Now I am finally doing something I love. Enhancing my education as well as helping people in whatever way I can. Something I could not do in my youth because of too many responsibilities. I have three lists that I keep. The first consists of things I have to do. The second is a list of things I want to do. And the third is a list of things I would like to do some time in the future, if possible. One of the goals on list number two was to finish college and I was fortunate enough to accomplish this in 1980 with an undergraduate degree. I am now attempting graduate school. I would like to go on for a doctorate if the college I attend has one to offer. Since I have returned to school I find, not how much I know, but how much I don't yet know. People have the tendency to look at me and think I am an old study-wetty because I dress and act conservatively. When they actually get to know me, they are surprised at some of my accomplishments, my hobbies, my studies, and my interests (of which I have many). It shocks them to find out like George Carlin, jazz music and Willy Nelson or that I like to read books that are written about the mind of serial killers along with historical novels about the Roman empire. They also find it surprising that my favorite TV shows include Frazier as well as NYPD Blue and Saturday Night Live. I guess you could call me a conservative liberal or maybe a liberal conservative. Two things about my life make me sad. One is the fact that I know I will never live long enough to achieve all of the goals on my three lists. And the second is that I was never fortunate enough to have a family of my own. Although I have a wide array of many friends, it is still not the same as family.

ID 2 I am a graduate student at OSU. I am working toward a master's degree in communication. I plan on continuing my education and pursuing a doctorate degree so I can eventually become a professor. I live in Lakewood, Ohio, a suburb of Cleveland. I have been in Lakewood for 4 years and I really love the area. I am happily married. My husband and I do not have children yet but hope to soon. I have 3 siblings and we all live within 50 miles of each other. We see each other often. I have 5 nieces and nephews and I love being an aunt. My husband and I were just asked to be the godparents of our niece.

ID 3 Since the question itself is vague, I will describe myself in two different ways, physically and mentally. I am an 18-year-old female. I don't reach my "full potential" for height, so I am only 5'2". Maybe I should have listened to my mom when she told me to eat my vegetables. I have dark brown hair that some would consider black. When I was younger I hated my hair because all the boys in my 2nd and 3rd grade classes preferred blondes, so naturally I
LIWC
About LIWC

• Created by Pennebaker, Booth, & Francis

• Intended to measure both affective and cognitive constructs

• 84 Output Variables (standard dictionaries):
  • 17 Standard linguistic dimensions (e.g., number of pronouns)
  • 25 Word categories (e.g., “psychological constructs – affect, cognition”)
  • 10 Time categories (e.g., “space, motion”)
  • 19 Personal concerns (e.g., “home”)
## Linguistic Inquiry and Word Count

Table 1: LIWC2007 Output Variable Information

<table>
<thead>
<tr>
<th>Category</th>
<th>Abbrev</th>
<th>Examples</th>
<th>Words in category</th>
<th>Validity (Judges)</th>
<th>Alpha: Binary/ Raw</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic Processes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word count</td>
<td>wc</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>words/sentence</td>
<td>wps</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dictionary words</td>
<td>dic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words&gt;6 letters</td>
<td>sixtr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total function words</td>
<td>funct</td>
<td></td>
<td>464</td>
<td>.97/.40</td>
<td></td>
</tr>
<tr>
<td>Total pronouns</td>
<td>pronun</td>
<td>I, them, itself</td>
<td>116</td>
<td>.91/.38</td>
<td></td>
</tr>
<tr>
<td>Personal pronouns</td>
<td>ppromn</td>
<td>I, them, her</td>
<td>70</td>
<td>.88/.20</td>
<td></td>
</tr>
<tr>
<td>1st pers singular</td>
<td>I</td>
<td>I, me, mine</td>
<td>12</td>
<td>.52</td>
<td>.62/.44</td>
</tr>
<tr>
<td>1st pers plural</td>
<td>we</td>
<td>We, us, our</td>
<td>12</td>
<td>.66/.47</td>
<td></td>
</tr>
<tr>
<td>2nd person</td>
<td>you</td>
<td>You, your, thou</td>
<td>20</td>
<td>.73/.34</td>
<td></td>
</tr>
<tr>
<td>3rd pers singular</td>
<td>shehe</td>
<td>She, her, him</td>
<td>17</td>
<td>.75/.52</td>
<td></td>
</tr>
<tr>
<td>3rd pers plural</td>
<td>they</td>
<td>They, their, they'd</td>
<td>10</td>
<td>.50/.36</td>
<td></td>
</tr>
<tr>
<td>Impersonal pronouns</td>
<td>ipron</td>
<td>It, it's, those</td>
<td>96</td>
<td>.78/.46</td>
<td></td>
</tr>
<tr>
<td>Articles</td>
<td>art</td>
<td>A, an, the</td>
<td>3</td>
<td>.14/.14</td>
<td></td>
</tr>
<tr>
<td>[Common verbs] a</td>
<td>verb</td>
<td>Walk, went, see</td>
<td>383</td>
<td>.97/.42</td>
<td></td>
</tr>
<tr>
<td>Auxiliary verbs</td>
<td>auxverb</td>
<td>Am, will, have</td>
<td>144</td>
<td>.91/.23</td>
<td></td>
</tr>
<tr>
<td>Past tense a</td>
<td>past</td>
<td>Went, ran, had</td>
<td>145</td>
<td>.79</td>
<td>.94/.75</td>
</tr>
<tr>
<td>Present tense a</td>
<td>present</td>
<td>To, do, be</td>
<td>150</td>
<td>.61/.74</td>
<td></td>
</tr>
</tbody>
</table>
The LIWC Interface
LIWC Output: Data Matrix
(Each row is a case/text, each column a dictionary)
About Diction

- Created by Roderick P. Hart, University of Texas, originally for the purpose of analyzing political discourse
- To measure “semantic features”, uses a series of 31 standard dictionaries and five “Master Variables” (scales constituted of combinations of the 31):
  - Activity
  - Optimism
  - Certainty
  - Realism
  - Commonality
The Diction Interface

Diction 5.0 - Diction1

Input Files
- Clinton.txt
- Dole.txt
- Nixon.txt
- Unabomer.txt

Total Words Analyzed: 0
Total Characters Analyzed: 0
Average Word Size: 0.00
Number of Different Words: 0
Alphanumeric Identifier: 
Descriptive Identifier: 

Options
Active Custom Dictionaries:
Character Counts: [none]
View Character Counts: No
View Word Counts: No
Small File Option: Report extrapolations
Large File Option: Averaged [Analyze maximum 500,000 words]
Numeric File Name: C:\Program Files\dict\data\Research.num
Use Comma Separator: Yes
Print Input Text: No
View Input Text: No

Normative Values
Class: All

For Help, press F1
## Diction Output: Dictionary Totals with Normative Values

### Normative Values
Class: Entertainment  
Type: TV Dramas

### Standard Dictionary Totals

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>% of Words Analyzed</th>
<th>Normal Range Low</th>
<th>Normal Range High</th>
<th>Standard Score</th>
<th>Out of Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical Terms</td>
<td>0.75</td>
<td>0.15</td>
<td>0.36</td>
<td>6.41</td>
<td>-0.94</td>
<td></td>
</tr>
<tr>
<td>Ambivalence</td>
<td>24.16</td>
<td>4.83</td>
<td>14.25</td>
<td>26.31</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>Self-reference</td>
<td>39.00</td>
<td>7.80</td>
<td>22.59</td>
<td>41.77</td>
<td>3.94</td>
<td></td>
</tr>
<tr>
<td>Tenacity</td>
<td>48.60</td>
<td>9.72</td>
<td>40.59</td>
<td>57.59</td>
<td>2.88</td>
<td></td>
</tr>
<tr>
<td>Leveling Terms</td>
<td>7.50</td>
<td>1.50</td>
<td>5.14</td>
<td>12.31</td>
<td>-0.36</td>
<td></td>
</tr>
<tr>
<td>Collectives</td>
<td>2.25</td>
<td>0.45</td>
<td>0.40</td>
<td>3.72</td>
<td>-1.34</td>
<td></td>
</tr>
<tr>
<td>Praise</td>
<td>7.05</td>
<td>1.41</td>
<td>5.12</td>
<td>12.77</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>6.25</td>
<td>1.25</td>
<td>4.49</td>
<td>12.83</td>
<td>1.86</td>
<td></td>
</tr>
<tr>
<td>Inspiration</td>
<td>0.62</td>
<td>0.12</td>
<td>0.06</td>
<td>3.17</td>
<td>-1.20</td>
<td></td>
</tr>
<tr>
<td>Blame</td>
<td>2.00</td>
<td>0.40</td>
<td>1.44</td>
<td>6.38</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>Hardship</td>
<td>8.50</td>
<td>1.70</td>
<td>2.08</td>
<td>8.60</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>Aggression</td>
<td>4.50</td>
<td>0.90</td>
<td>0.31</td>
<td>4.79</td>
<td>-0.21</td>
<td></td>
</tr>
<tr>
<td>Accomplishment</td>
<td>5.67</td>
<td>1.33</td>
<td>2.51</td>
<td>8.40</td>
<td>-0.82</td>
<td></td>
</tr>
</tbody>
</table>
Diction Output: Managing Normative Values

### Normative Values

<table>
<thead>
<tr>
<th>Class</th>
<th>Entertainment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>TV Dramas</td>
</tr>
</tbody>
</table>

### Standard Dictionary Totals

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>% of Words</th>
<th>Normal Range</th>
<th>Standard Score</th>
<th>Out of Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical Terms</td>
<td>0.75</td>
<td></td>
<td></td>
<td>-0.94</td>
<td></td>
</tr>
<tr>
<td>Ambivalence</td>
<td>24.16</td>
<td></td>
<td></td>
<td>1.78</td>
<td></td>
</tr>
<tr>
<td>Self-reference</td>
<td>39.00</td>
<td></td>
<td></td>
<td>3.94</td>
<td></td>
</tr>
<tr>
<td>Tenacity</td>
<td>48.60</td>
<td></td>
<td></td>
<td>2.08</td>
<td></td>
</tr>
<tr>
<td>Leveling Terms</td>
<td>7.50</td>
<td></td>
<td></td>
<td>-0.36</td>
<td></td>
</tr>
<tr>
<td>Collectives</td>
<td>2.25</td>
<td></td>
<td></td>
<td>-1.34</td>
<td></td>
</tr>
<tr>
<td>Praise</td>
<td>7.05</td>
<td></td>
<td></td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>6.25</td>
<td></td>
<td></td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>Inspiration</td>
<td>0.62</td>
<td></td>
<td></td>
<td>-1.20</td>
<td></td>
</tr>
<tr>
<td>Blame</td>
<td>2.00</td>
<td>0.40</td>
<td>1.44</td>
<td>8.38</td>
<td>-0.85</td>
</tr>
<tr>
<td>Hardship</td>
<td>0.50</td>
<td>1.70</td>
<td>2.00</td>
<td>0.60</td>
<td>0.57</td>
</tr>
<tr>
<td>Aggression</td>
<td>4.50</td>
<td>0.90</td>
<td>0.31</td>
<td>4.79</td>
<td>-0.21</td>
</tr>
<tr>
<td>Accomplishment</td>
<td>6.67</td>
<td>1.33</td>
<td>2.61</td>
<td>8.40</td>
<td>-0.82</td>
</tr>
</tbody>
</table>

Changes made here affect only the selected files for the active project. The project must be saved to make the changes permanent. The current project Properties still apply to new files that are added.
Diction: Custom Dictionaries as Simple .txt Files
PCAD
About PCAD

Developed by Gottschalk & Bechtel, using scales developed by Gottschalk & Gleser for human-coding in 1960s

Diagnostic—assesses one text at a time

Intended for naturally-occurring speech or writing, minimum 80 words

Measures states of neuropsychiatric interest such as:

- Anxiety
- Hostility
- Cognitive impairment
- Depression
- Schizophrenia
- Achievement Strivings
- Hope
The PCAD Interface
PCAD Interface-2

Multi-Sample Output Options

Output from all samples in this file will be sent to the screen. Please indicate any other destinations for outputs.

- Copy output to printer
- Copy output to file
- Copy (summary) output to spreadsheet

OK
PCAD Output: 4 Types (Clauses, Summaries, Analyses, Diagnoses)
PCAD Output: Analyses

Analysis for C:\Program Files\PCAD\K1033F.sam

giving a description of the typical trait-like characteristics of
an individual is more stable if at least three speech samples are
used. Scoring is not reliable on samples of less than 80 words.
This evaluation consists of a single verbal sample containing
122 words.

SPECIFIC ANALYSIS

This individual has a moderately elevated total anxiety score. A
total anxiety level of this magnitude may occur infrequently in
normal individuals. It appears more often in people with acute
or chronic mental disorders.

Certain anxiety subscale scores are moderately elevated. These
are:

-separation anxiety

Certain anxiety subscale scores are mildly elevated. These are:

-mutilation anxiety

-shame anxiety

Those anxiety subscale scores which are elevated provide clues
concerning the dynamics of the contributions to the total
anxiety. This information offers guidelines to the clinician for
psychosocial or biomedical areas meriting more inquiry and
investigation.
If the clinician is in the process of making a neuropsychiatric diagnosis, the DSM-IV diagnostic classifications to consider are:

For Adults:

Axis I:

- Panic Disorder or Anxiety Disorder, with agoraphobia (300.21) or without agoraphobia (300.01).
- Adjustment Disorder with anxious mood, moderate (309.24)
- Adjustment Disorder with mixed anxiety and depressed mood (309.28)

For Children:

Axis I:

- Separation Anxiety Disorder (309.21)
- Anxiety Disorder not otherwise specified (300.00)
About CATPAC

Created by Joseph Woelfel, Communication scientist at University of Buffalo

Part of the GALILEO suite of softwares that analyze and display various types of networks

CATPAC uses a neural network approach, identifying the most frequent words and determining patterns of connection based on co-occurrence

A scanning window is used to measure the association/co-occurrence

Uses cluster analysis to present results of this co-occurrence procedure
The show's two hour long pilot episode begins on St. Patrick's Day with nurse Lydia Wright waking Dr. Mark Greene and telling him that he has a patient. This patient turns out to be Dr. Doug Ross, a good friend and co-worker of Mark's. Doug is drunk and sleeps it off in exam three. This irresponsible behavior of showing up to work drunk shows exactly what Doug's personality is and how he acts in future episodes. After having gone back to sleep after this disturbance, Mark is again woken up two times until he finally wakes for his shift at 6:30 A.M.

At 6:15 a.m., a building had collapsed during construction, leaving two dead and twelve injured. Dr. Peter Benton walks into his shift this morning to find out that Cook County General will be receiving these patients, seven of them being critical. At this time, there is a small staff available and they gather everybody available to help with the mass trauma.

The first patient is a Mr. Wilson who the E.R. frantically brings to the trauma room. The man's right hand is horribly injured; orthopedics is notified of the patient and an O.R. is booked to save his hand. Other patients include Mr. Jackson, who is gasping for air and an elderly woman who coughs up blood as the recently awakened Doug works to help her. Meanwhile, Benton, a surgical resident, wishes to perform the orthopedic surgery, however, is told he is too inexperienced for this case. Instead, Benton comes to help the elderly woman whose condition continues to worsen; in the end she lives.

Mark now must inform Mr. Pinnell, who is father had a heart attack, that he died. This man is extremely...
CATPAC Output:
Descending Frequency List, Alphabatically Sorted List

<table>
<thead>
<tr>
<th>WORD</th>
<th>FREQ</th>
<th>PCNT</th>
<th>FREQ</th>
<th>PCNT</th>
<th>WORD</th>
<th>FREQ</th>
<th>PCNT</th>
<th>FREQ</th>
<th>PCNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>BENDIN</td>
<td>15</td>
<td>10.1</td>
<td>92</td>
<td>55.7</td>
<td>ARRIVES</td>
<td>4</td>
<td>2.1</td>
<td>26</td>
<td>14.2</td>
</tr>
<tr>
<td>CARTER</td>
<td>17</td>
<td>9.0</td>
<td>76</td>
<td>45.1</td>
<td>BENDIN</td>
<td>19</td>
<td>10.1</td>
<td>92</td>
<td>55.3</td>
</tr>
<tr>
<td>MARK</td>
<td>16</td>
<td>6.5</td>
<td>79</td>
<td>47.2</td>
<td>CAROL</td>
<td>6</td>
<td>3.2</td>
<td>26</td>
<td>14.7</td>
</tr>
<tr>
<td>TELLS</td>
<td>13</td>
<td>6.9</td>
<td>74</td>
<td>44.0</td>
<td>CARTER</td>
<td>17</td>
<td>9.0</td>
<td>76</td>
<td>45.1</td>
</tr>
<tr>
<td>DOUG</td>
<td>12</td>
<td>6.3</td>
<td>66</td>
<td>39.1</td>
<td>CONDITION</td>
<td>4</td>
<td>2.1</td>
<td>26</td>
<td>14.3</td>
</tr>
<tr>
<td>E</td>
<td>11</td>
<td>5.8</td>
<td>58</td>
<td>35.7</td>
<td>DOUG</td>
<td>12</td>
<td>6.3</td>
<td>66</td>
<td>39.1</td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>5.3</td>
<td>57</td>
<td>35.1</td>
<td>DR</td>
<td>5</td>
<td>2.8</td>
<td>28</td>
<td>15.3</td>
</tr>
<tr>
<td>CAROL</td>
<td>6</td>
<td>4.2</td>
<td>42</td>
<td>25.0</td>
<td>E</td>
<td>10</td>
<td>5.3</td>
<td>57</td>
<td>35.1</td>
</tr>
<tr>
<td>HELP</td>
<td>6</td>
<td>3.2</td>
<td>39</td>
<td>23.1</td>
<td>GSW</td>
<td>5</td>
<td>2.8</td>
<td>28</td>
<td>15.3</td>
</tr>
<tr>
<td>MORGENSTERN</td>
<td>6</td>
<td>3.2</td>
<td>31</td>
<td>18.9</td>
<td>HELP</td>
<td>6</td>
<td>3.2</td>
<td>39</td>
<td>23.1</td>
</tr>
<tr>
<td>GOOD</td>
<td>5</td>
<td>2.6</td>
<td>28</td>
<td>16.3</td>
<td>JOB</td>
<td>4</td>
<td>2.1</td>
<td>26</td>
<td>14.3</td>
</tr>
<tr>
<td>NOTHING</td>
<td>5</td>
<td>2.6</td>
<td>11</td>
<td>6.0</td>
<td>MAN</td>
<td>4</td>
<td>2.1</td>
<td>26</td>
<td>14.3</td>
</tr>
<tr>
<td>PATIENT</td>
<td>5</td>
<td>2.6</td>
<td>22</td>
<td>12.0</td>
<td>MARK</td>
<td>10</td>
<td>6.0</td>
<td>79</td>
<td>45.2</td>
</tr>
<tr>
<td>SURGERY</td>
<td>5</td>
<td>2.6</td>
<td>21</td>
<td>12.5</td>
<td>MORGENSTERN</td>
<td>6</td>
<td>3.2</td>
<td>31</td>
<td>18.9</td>
</tr>
<tr>
<td>SUSAN</td>
<td>5</td>
<td>2.6</td>
<td>30</td>
<td>16.4</td>
<td>NOTHING</td>
<td>5</td>
<td>2.6</td>
<td>11</td>
<td>6.0</td>
</tr>
<tr>
<td>TIME</td>
<td>5</td>
<td>2.6</td>
<td>28</td>
<td>16.3</td>
<td>PATIENT</td>
<td>5</td>
<td>2.6</td>
<td>26</td>
<td>14.3</td>
</tr>
<tr>
<td>UPSET</td>
<td>5</td>
<td>2.6</td>
<td>31</td>
<td>18.3</td>
<td>R</td>
<td>11</td>
<td>5.8</td>
<td>58</td>
<td>35.7</td>
</tr>
<tr>
<td>WIFE</td>
<td>5</td>
<td>2.6</td>
<td>25</td>
<td>14.7</td>
<td>SURGERY</td>
<td>5</td>
<td>2.6</td>
<td>26</td>
<td>14.3</td>
</tr>
<tr>
<td>WELL</td>
<td>5</td>
<td>2.6</td>
<td>31</td>
<td>18.3</td>
<td>SUSAN</td>
<td>5</td>
<td>2.6</td>
<td>26</td>
<td>14.3</td>
</tr>
<tr>
<td>ARRIVES</td>
<td>4</td>
<td>2.1</td>
<td>26</td>
<td>14.2</td>
<td>TELLS</td>
<td>4</td>
<td>2.1</td>
<td>14</td>
<td>7.7</td>
</tr>
<tr>
<td>CONDITION</td>
<td>4</td>
<td>2.1</td>
<td>28</td>
<td>16.3</td>
<td>TIME</td>
<td>5</td>
<td>2.6</td>
<td>26</td>
<td>14.3</td>
</tr>
<tr>
<td>EPISODE</td>
<td>4</td>
<td>2.1</td>
<td>14</td>
<td>7.7</td>
<td>UPSET</td>
<td>5</td>
<td>2.6</td>
<td>31</td>
<td>16.9</td>
</tr>
<tr>
<td>JOB</td>
<td>4</td>
<td>2.1</td>
<td>28</td>
<td>16.3</td>
<td>WIFE</td>
<td>5</td>
<td>2.6</td>
<td>26</td>
<td>14.3</td>
</tr>
<tr>
<td>MAN</td>
<td>4</td>
<td>2.1</td>
<td>22</td>
<td>12.6</td>
<td>WELL</td>
<td>5</td>
<td>2.6</td>
<td>31</td>
<td>16.9</td>
</tr>
</tbody>
</table>

Word Method:
ACU TCMW  HSMGSWJNBDMDP  CE  TER  ROPIAEON  SUODRAPAIE  ENSMNFRLRWStUNRTUERI
IDEF  EPG  A  IHTGKEST TEL
WITLSE  NF  N  NO  R
SOY  G  TE
N
E

...
**CATPAC Output: Dendogram**

<table>
<thead>
<tr>
<th>PATIENT</th>
<th>SURGERY</th>
<th>TIME</th>
<th>UPSET</th>
<th>WIFE</th>
<th>WILL</th>
<th>ARRIVES</th>
<th>CONDITION</th>
<th>EPISODE</th>
<th>JOB</th>
<th>MAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 2.6  22 12.0</td>
<td>5 2.6  21 11.5</td>
<td>5 2.6  30 16.4</td>
<td>5 2.6  31 16.9</td>
<td>5 2.6  25 13.7</td>
<td>5 2.6  31 16.9</td>
<td>4 2.1  26 14.2</td>
<td>4 2.1  26 15.3</td>
<td>4 2.1  14 7.7</td>
<td>4 2.1  26 15.3</td>
<td>4 2.1  22 12.0</td>
</tr>
</tbody>
</table>

**WARDS METHOD**

```
ACUTCHWMBNEDMDEPECER
RPPIAJESUIEGOARPAE
RNSMRNFRARWSLBNUR
TIRL
IDEEOEIPGGALRTGKISTL
VITEEENNJOTOEES
BTIRENNNIDR
SITRNGTE
NRE
N

```

```

```
CATPAC Output: 3D Plot