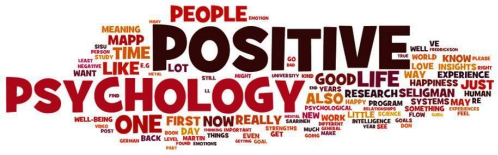
The Psychological Meaning of Words: LIWC and Computerized Text Analysis Method

Yla R Tausczik, James W Pennebaker

Journal of Language and Social Psychology · March 2010

Presented by: Bin Xie, Wenting Tan



Why important?

- **citation:** 2879
- **main problems:**
  - computerized text analysis methods
  - how Linguistic Inquiry and Word Count (LIWC) was created and validated

> We are in the midst of a technological revolution whereby, for the first time, researchers can link daily word use to a broad array of real-world behaviors... to detect meaning in a wide variety of experimental settings, including to show attentional focus, emotionality, social relationships, thinking styles, and individual differences.

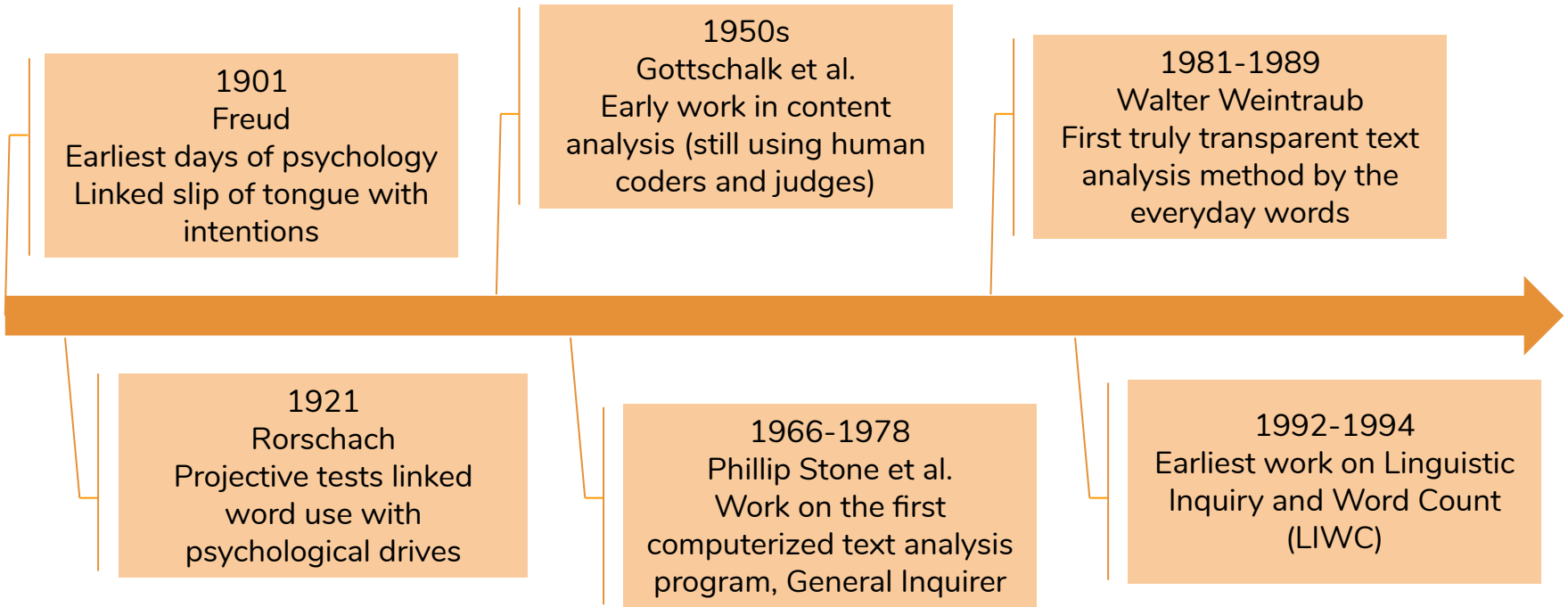
Computerized Text Analysis

Not one method or approach, but many

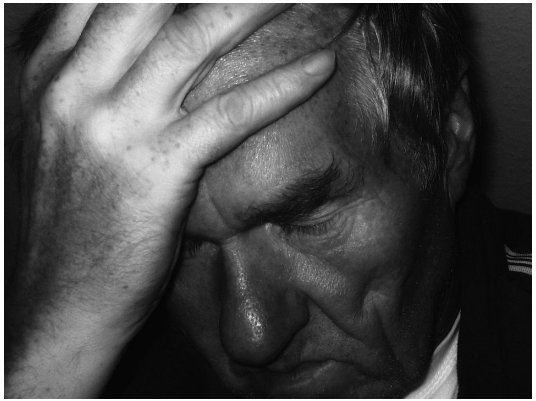
A Swiss Army Knife for studying language use and text data

Any automatized process for categorizing or uncovering latent meaning in word use within or across files with computer-readable formats

Text Analysis History



LIWC - Problems with judges



LIWC

LIWC

LIWC RESULTS

251 WORDS LONG, PROFESSIONAL OR SCIENTIFIC WRITING

Your text sample is 251 words. The LIWC2015 analysis of the text sample you entered is below. If you entered more than 500 words, only the first 500 words were analyzed. Note that LIWC2015 actually produces about 90 different output dimensions. Always remember that the more text you have, the more trustworthy the results.

TRADITIONAL LIWC DIMENSION	YOUR DATA	AVERAGE FOR PROFESSIONAL OR SCIENTIFIC WRITING
I-WORDS (I, ME, MY)	0.0	0.63
SOCIAL WORDS	8.7	7.62
POSITIVE EMOTIONS	2.8	2.32
NEGATIVE EMOTIONS	2.0	1.45
COGNITIVE PROCESSES	13.0	7.52
SUMMARY VARIABLES		
ANALYTIC	95.2	92.57
CLOUT	73.6	68.17
AUTHENTICITY	39.3	24.84
EMOTIONAL TONE	39.9	43.61

LIWC - logic and development

- > 80 categories
- objective (eg. article)
- subjective (eg. emotion word)
  - groups of 3 judges
  - > 93% agreement

LIWC - The Psychometrics of Word Usage

Verifying the **validity** and **reliability** of word usage is trickier!!!

psychometrics of questionnaire

- a factor analysis of the item
- computes the test–retest reliability
- validation tests to see if the questionnaire correlates with or predicts real-world behaviors

psychometrics of word usage

- rarely normally distributed
- low base rate
- people don't repeat themselves (i.e: rarely gets good split-half reliability)

LIWC - Content Versus Style Words

e.g: **It was a dark and stormy night**

content words

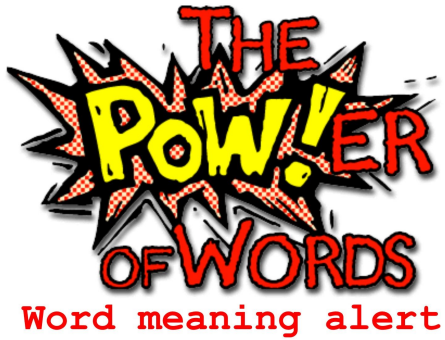
- convey the content of a communication
- nouns, regular verbs, and many adjectives and adverbs
- what they are saying

style words

- intertwined through the content words
- pronouns, prepositions, articles, conjunctions, auxiliary verbs, and a few other esoteric categories
- how people are communicating

LIWC - Social and Psychological Meaning of Words

- Attention focus
- Emotionality
- Social relationships
  - status, dominance, and hierarchy
  - coordination and group processes
  - honesty and deception
  - close relationships
- Thinking styles
- Individual differences



Attention Focus

- Internal vs External
  - eg. tennis & injury
- Pronous - subject of attention (eg. teasing)
- Verb Tense - temporal focus (eg. disclosed event)
- ‘We’, ‘us’, ‘ours’
  - the Royal We

Emotionality

- positive emotion words (eg. love, nice, sweet)
- negative emotion words (eg. hurt, ugly, nasty)
- degree of immersion (eg. physical pain in traumatic event)
- correlation
    - - articles, prepositions, relativity words
    - + pronouns, auxiliary verb, negation

Status, dominance, and social hierarchy

e.g: We are going to get over this difficulty! ***Higher***

Maybe I can finish it before the deadline. ***Lower***

Higher-status

- larger total word count
- more statements that involve others
- more first-person plural
- fewer question markers

Lower-status

- smaller total word count
- more self-focused and tentatives
- more first-person singular
- more question markers

Social coordinaiton and Group Processes

More communication, more unity, and positive feedback may promote better group performance.

1. amount of communication ⟶ ***word count***
2. unity ⟶ ***first-person plural***
   - "We can do it" VS "We ask you to do it"
3. positive feedback ⟶ ***assents & question marks***
   - Early stage: fewer assents and more questions
   - Late stage: more assents and less questions

Honesty and Deception

- + descriptiveness
  - word count
  - negative emotion, motion words
- - diversity and complexity
  - exclusive words (eg. but, without, exclude)
  - ‘I’
  - third-person pronous

Lie to me*

Close Relaitonships

- ‘we’ vs ‘you’
  - eg. relatives of patients
- ‘we’ vs ‘I’
  - eg. romantic partners

Thinking styles

Cognitive complexity: a richness of two components of reasoning

Exclusive words

- make distinctions
- used at higher rates among people telling the truth

e.g: but, without, exclude

conjunctions

- join multiple thoughts together
- important for creating a coherent narrative

e.g: and, also, although

Thinking Style

Prepositions: more complex and concrete information about a topic

e.g: The keys are ***in*** the box ***by*** the lamp ***under*** the painting.

cognitive mechanisms: lead to greater health improvements

- casual words(e.g: because, hence) VS insight words(e.g: think, consider)

tentative language and **filler words:** uncertain about the topic

- tentative language(e.g: maybe, guess): no story for the topic

- filler words(e.g: I mean, you know): younger use more than the elder

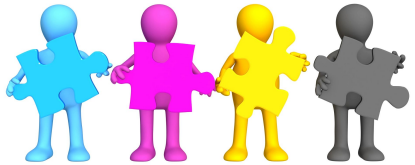
Individual Difference

The self-focus, cognitive complexity, social references, and emotional tone inherent in language differ with age, sex, personality, and mental health.

Age ⟶ the elder are less self-focused, refer more to the moment, accept in verbal complexity

Sex ⟶ women use more social words and references to others; men use more complex language.

Mental health ⟶ Depressed and suicidal individuals are more self-focused, express more negative emotion, use more death-related words.



Discussion & Future Work

- words in context (eg. n-gram, language style)
- college-aged students
- higly contrived lab setting
- spoken vs written language
- cultural difference (eg. markers of politeness, formality)

Conclusion

- computerized text analysis
- LIWC
- language and social psychology

"*burn ever brighter and illuminate the universe increasingly from different places*"