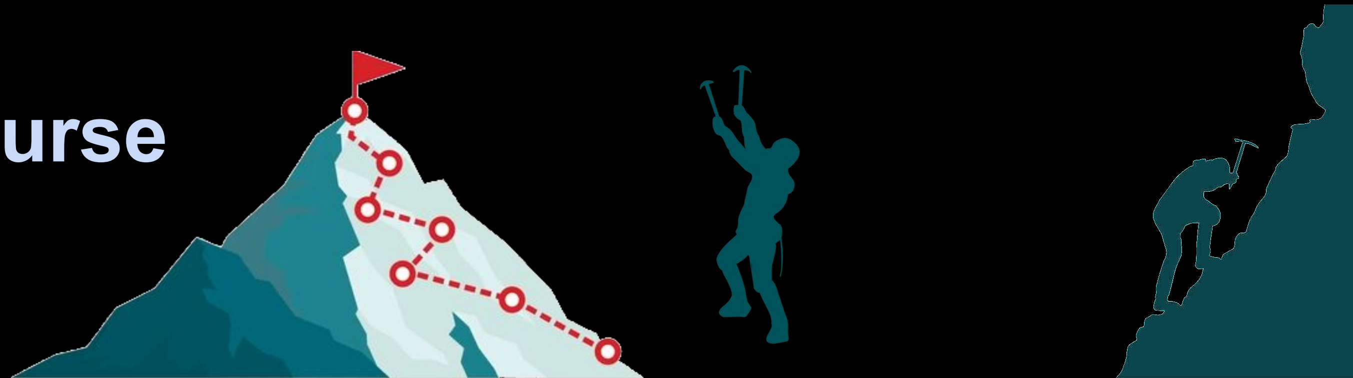
Human-Centered NLP
(Language and Human Psychology)

CSE 538

NLP, The Course



Overall NLP Concept

I. Syntax

II. Semantics

Overall NLP Concept

III. Language Modeling

IV. Applications

NLP, The Course

Overall NLP Concept

I. Syntax

Introduction to NLP; Tokenization; Words Corpora

One-hot, and Multi-hot encoding.
Parts-of-Speech; Named Entities;

Parsing; Verbal Predicates;Dependency Parsing

II. Semantics

Dependency Parsing; Word Sense Disambiguation

Vector Semantics (Embeddings), Word2vec

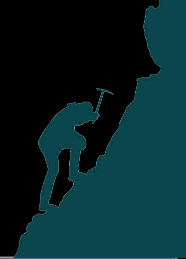
Probabilistic Language Models
Ngram Classifier, Topic Modeling

Overall NLP Concept

III. Language Modeling

IV. Applications

NLP The Course



Overall NLP Concept	Computation or ML
I. Syntax	**Classification**
Introduction to NLP; Tokenization; Words Corpora	Regular Expressions; Edit Distance
One-hot, and Multi-hot encoding. Parts-of-Speech; Named Entities;	Maximum Entropy Classifier (LogReg), Gradient Descent,
Parsing; Verbal Predicates;Dependency Parsing	Cross Validation; Regularization Accuracy Metrics; Shift Reduce
II. Semantics	**Probabilistic Models**
Dependency Parsing; Word Sense Disambiguation	Term Probabilities; N-d Vectors
Vector Semantics (Embeddings), Word2vec	LDA, Skipgram Model
Probabilistic Language Models Ngram Classifier, Topic Modeling	markov assumption, chain rule, smoothing

Overall NLP Concept	Computation or ML
III. Language Modeling	**Transformers**
Ethical Considerations	Model cards, Pred Bias Frmwrk
Masked Language Modeling (autoencoding)	Neural Networks; Backprop Cross-Entropy Loss Self-Attention,
Generative Language Modeling (autoregressive)	Positional encodings The Transformer: Beam Search
Applying LMs	Fine-Tuning, zero-/few-shot, Instruction tuning
IV. Applications	**Custom Statistical or Symbolic**
Language and Psychology (advanced sentiment)	Differential Language Analysis; Adaptive Modeling; Human LMing
Speech and Audio Processing, Dialog (chatbots)	Wave Transforms; RNNs
Question Answering, Translation	Multihop Reasoning

NLP The Course

Overall NLP Concept	Computation or ML
I. Syntax | Classification	
Introduction to NLP; Tokenization; Words Corpora	Regular Expressions; Edit Distance
One-hot, and Multi-hot encoding. Parts-of-Speech; Named Entities;	Maximum Entropy Classifier (LogReg), Gradient Descent,
Parsing; Verbal Predicates;Dependency Parsing	Cross Validation; Regularization Accuracy Metrics; Shift Reduce
II. Semantics | Probabilistic Models	
Dependency Parsing; Word Sense Disambiguation	Term Probabilities; N-d Vectors
Vector Semantics (Embeddings), Word2vec	LDA, Skipgram Model
Probabilistic Language Models Ngram Classifier, Topic Modeling	markov assumption, chain rule, smoothing

Overall NLP Concept	Computation or ML
III. Language Modeling | Transformers	
Ethical Considerations	Model cards, Pred Bias Frmwrk
Masked Language Modeling (autoencoding)	Neural Networks; Backprop Cross-Entropy Loss Self-Attention,
Generative Language Modeling (autoregressive)	Positional encodings The Transformer: Beam Search
Applying LMs	Fine-Tuning, zero-/few-shot, Instruction tuning
IV. Applications | Custom Statistical or Symbolic	
Language and Psychology (advanced sentiment)	**Differential Language Analysis; Adaptive Modeling; Human LMing**
Speech and Audio Processing, Dialog (chatbots)	Wave Transforms; RNNs
Question Answering, Translation	Multihop Reasoning

Natural Language Processing

Psychological & Health Sciences

Extraversion

Schwartz, H. A., Eichstaedt, ... & Ungar. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8(9)*.

Introversion

Schwartz, H. A., Eichstaedt, ... & Ungar. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8(9)*.

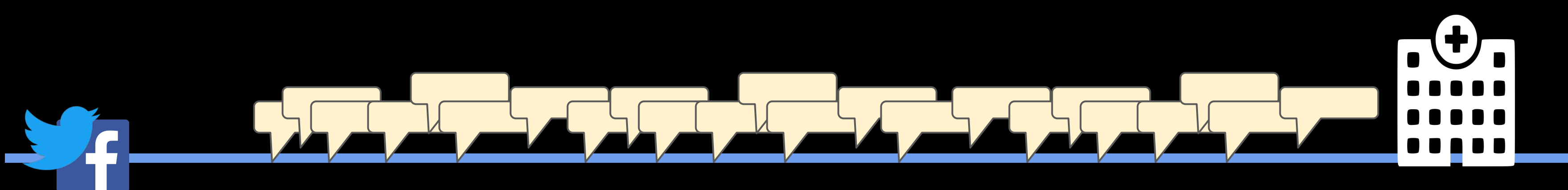
Natural Language Processing

Psychological & Health Sciences

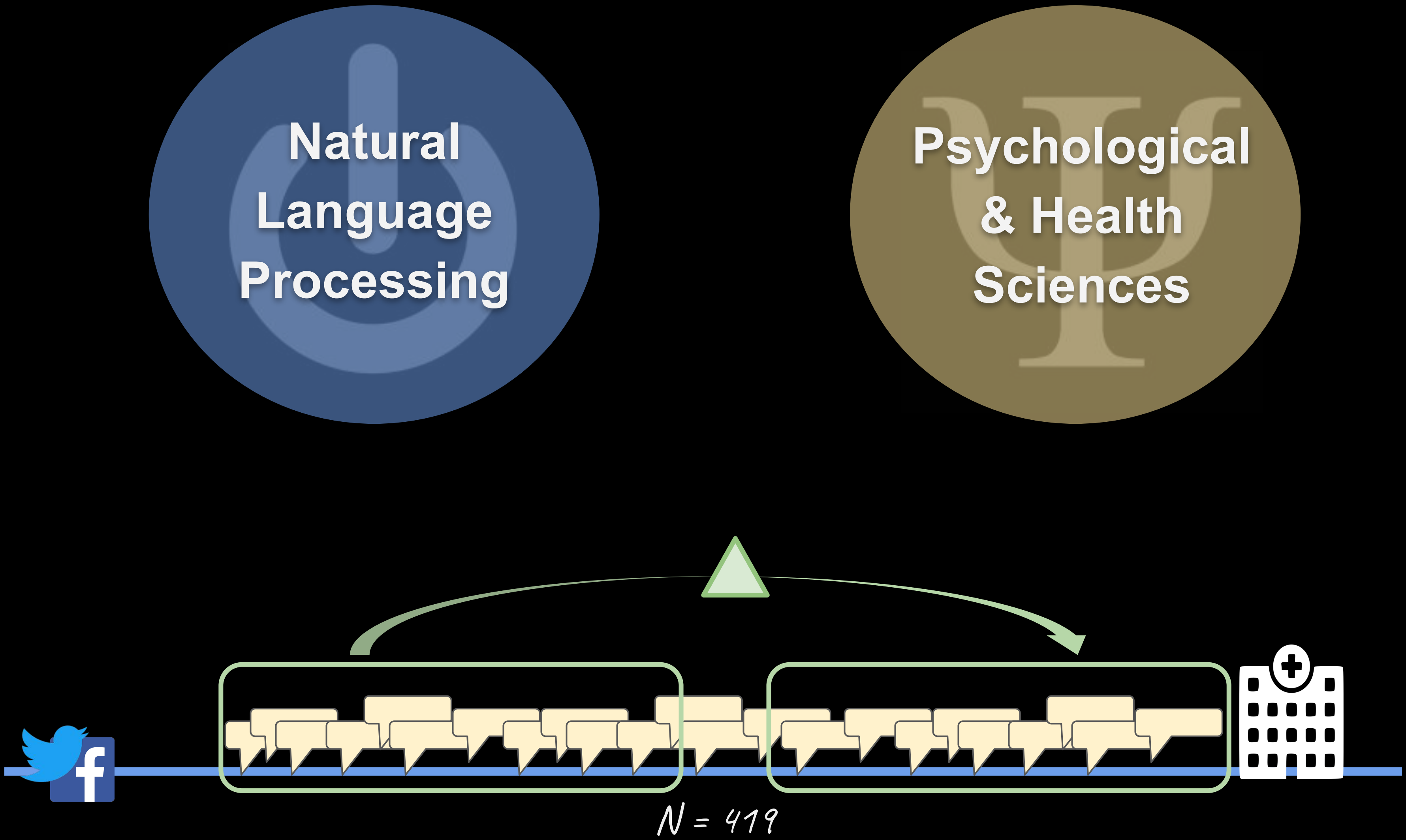
Schwartz, H. A., Eichstaedt, ... & Ungar. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8(9)*.

Natural Language Processing

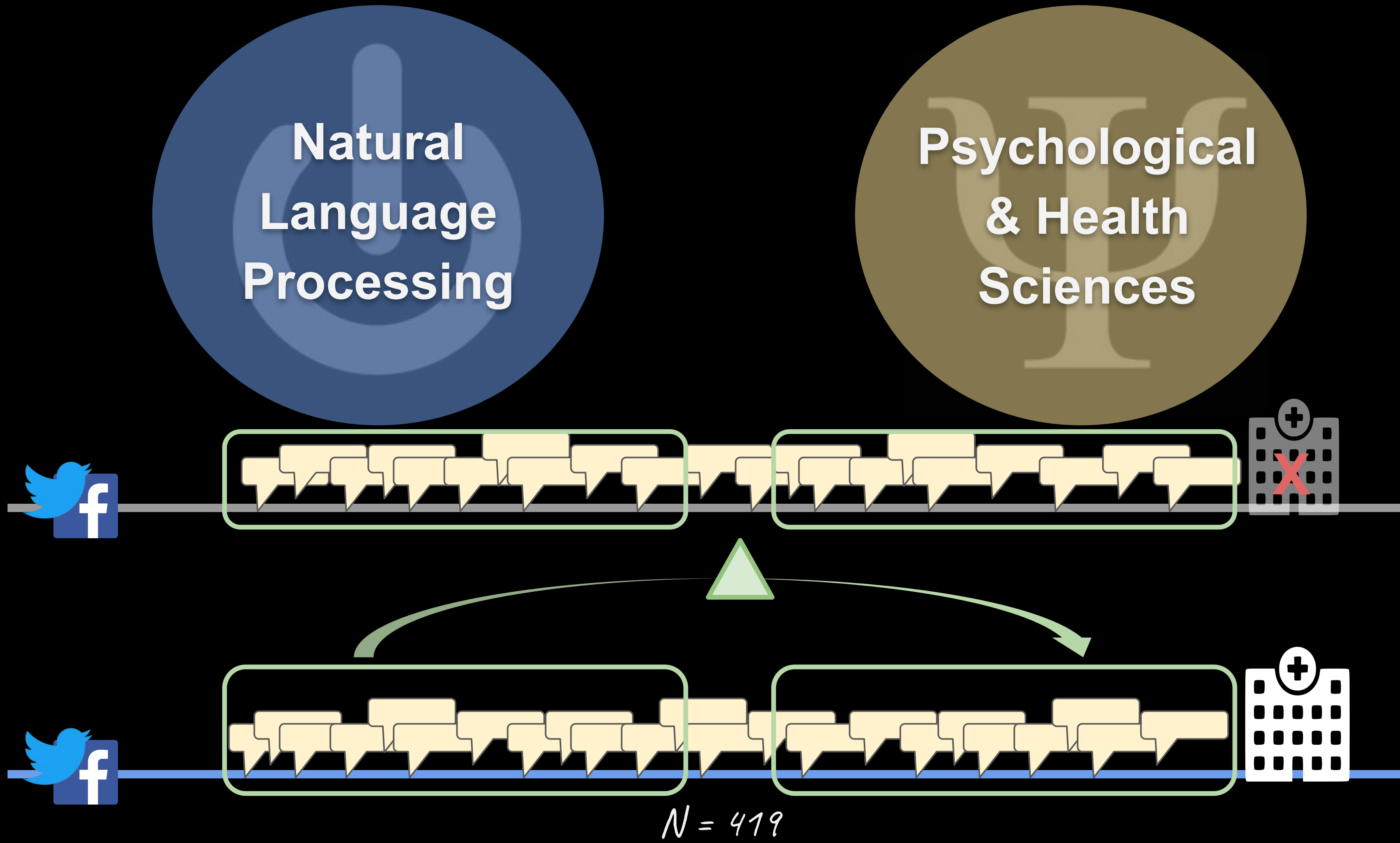
Psychological & Health Sciences



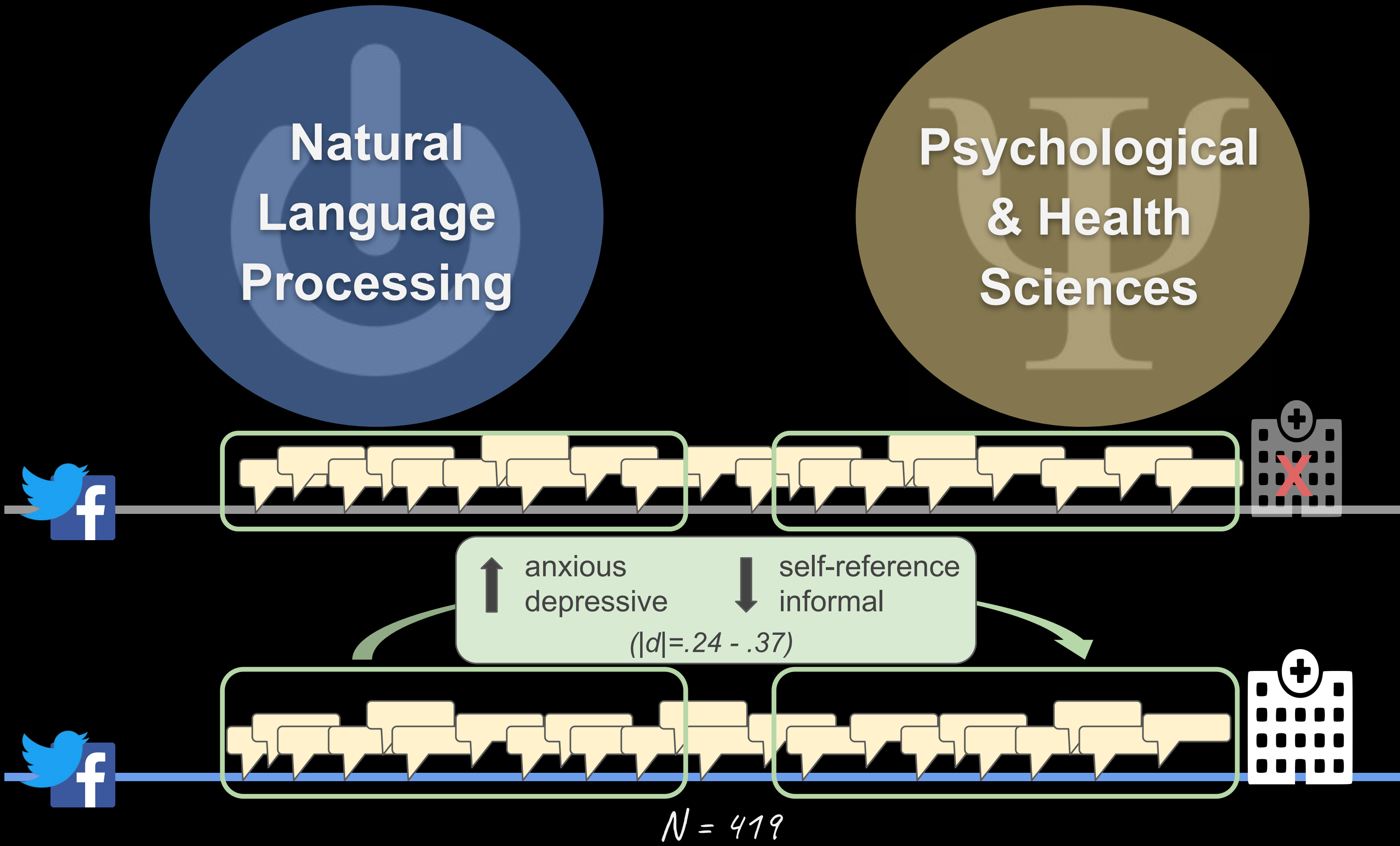
Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.



Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.



Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1),* 1-9.



Guntuku, S. C., Schwartz, H. A., Kashyap, A., Gaulton, J. S., Stokes, D. C., Asch, D. A., ... & Merchant, R. M. (2020). Variability in Language used on Social Media prior to Hospital Visits. *Nature - Scientific Reports*, *10(1)*, 1-9.

Natural Language Processing

Psychological & Health Sciences

Overly Simplified Problem-Statement:

Natural language is written by

Overly Simplified Problem-Statement:

Natural language is written by **people.**

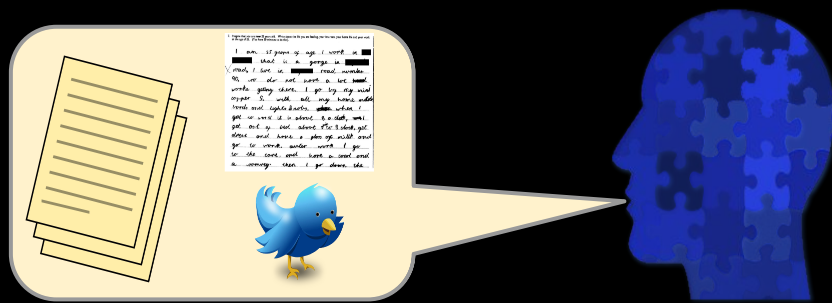
Overly Simplified Problem-Statement:

Natural language is written by **people.**

Problem

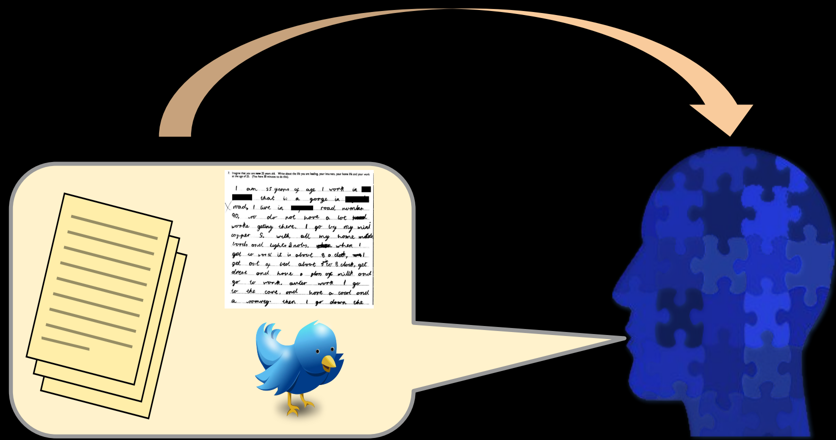
Natural language is written by **people.**

Natural language is generated by people.



People have different beliefs, backgrounds, styles, vocabularies, preferences, knowledge, personalities, …

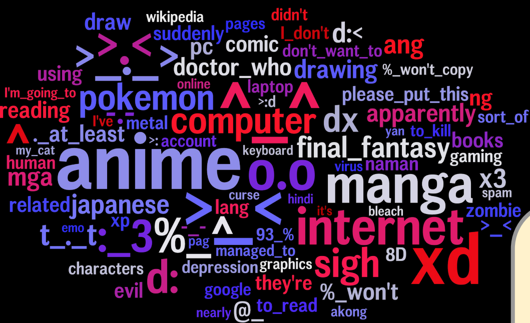
Natural language is generated by people.

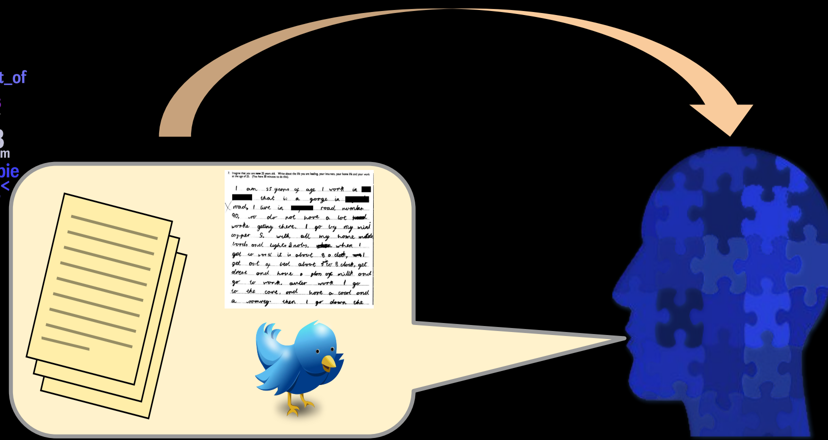


People have different beliefs, backgrounds, styles, vocabularies, preferences, knowledge, personalities, …,

and our language reflects these differences.

Natural language is generated by people.

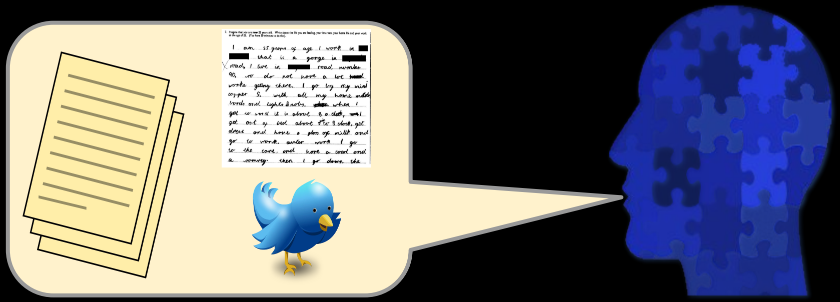




People have different beliefs, backgrounds, styles, vocabularies, preferences, knowledge, personalities, …,

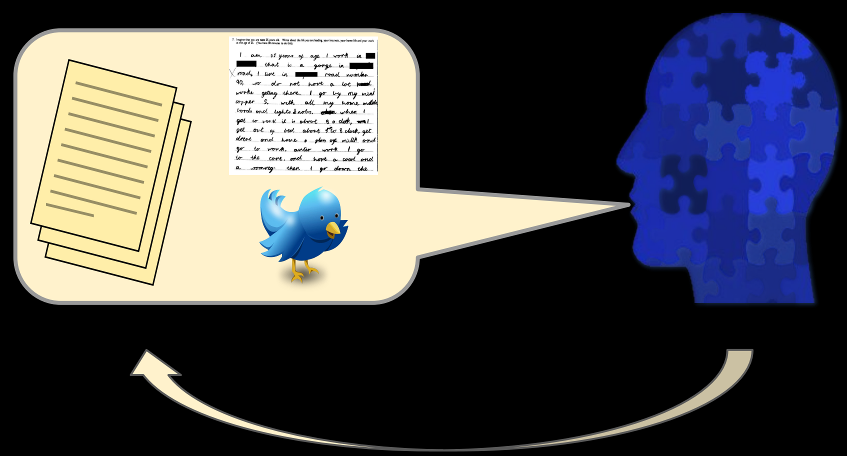
and our language reflects these differences.

Human Centered NLP:



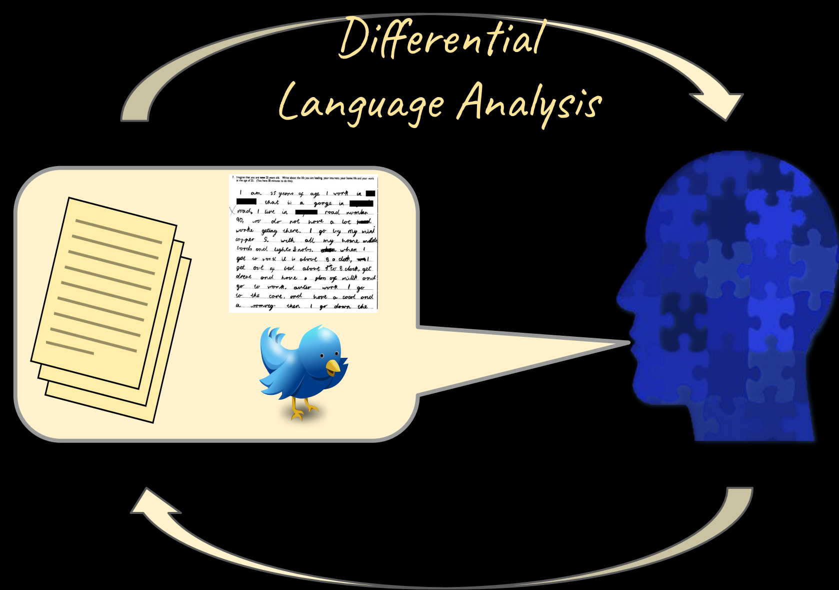
Human Centered NLP:

1. Model language as a human process



Human Centered NLP:

1. Model language as a human process
2. Use language to better understand humans.



Human-Centered NLP – We will cover:

1. Differential Language Analysis
2. Human Factor Adaptation
3. Human Language Modeling

Differential Language Analysis

Input:

Linguistic features

Human or community attribute

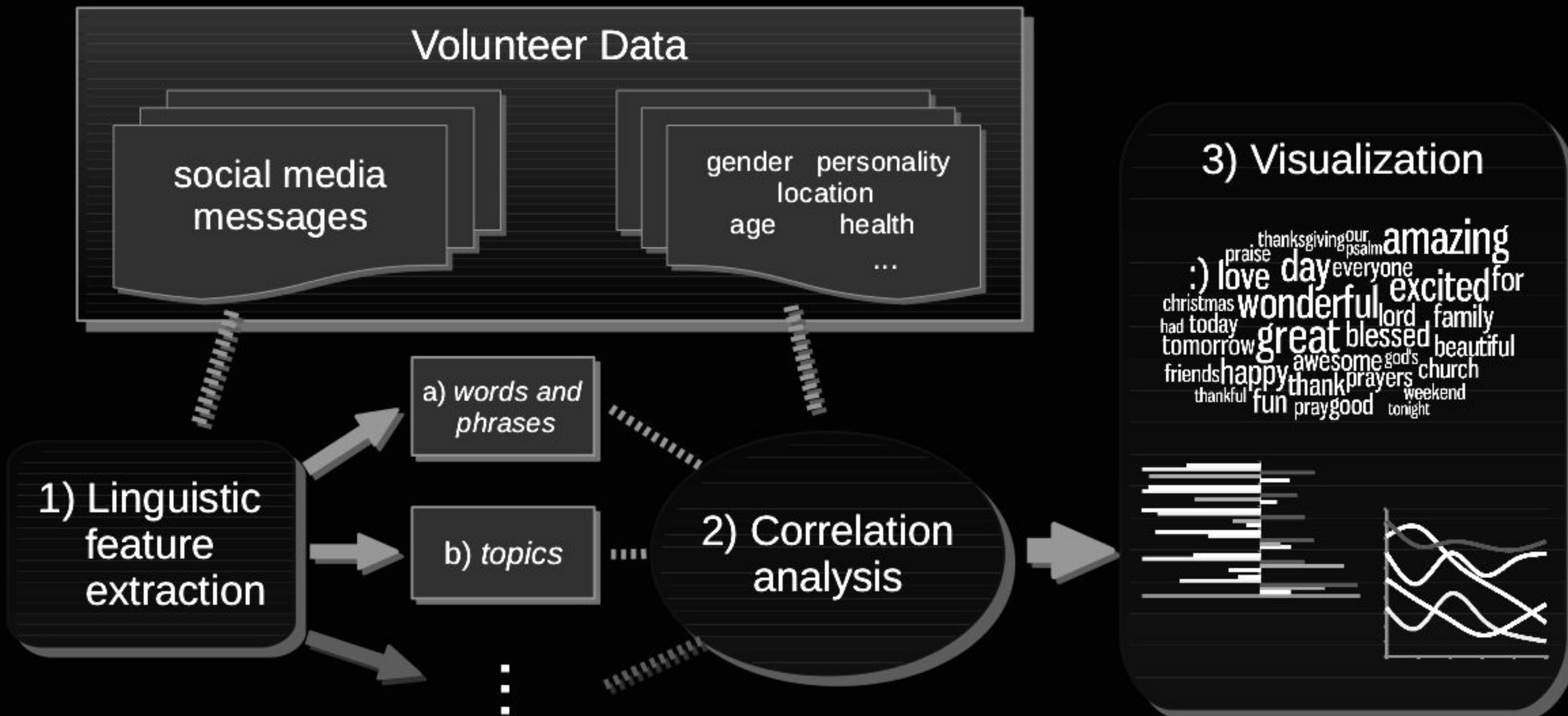
Output:

Features distinguishing attribute

Goal: Data-driven insights about an attribute

E.g. Words distinguishing communities with increases in real estate prices.

Differential Language Analysis



Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

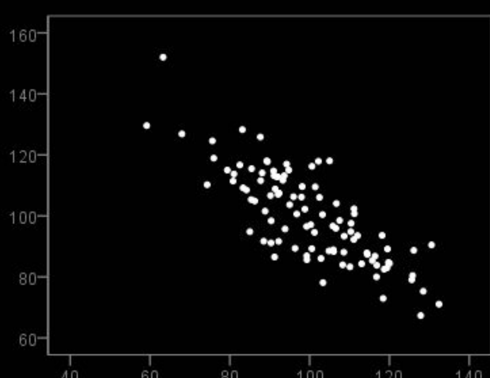
- Pearson Product-Moment Correlation
  - Limitation: Doesn’t handle controls

Differential Language Analysis

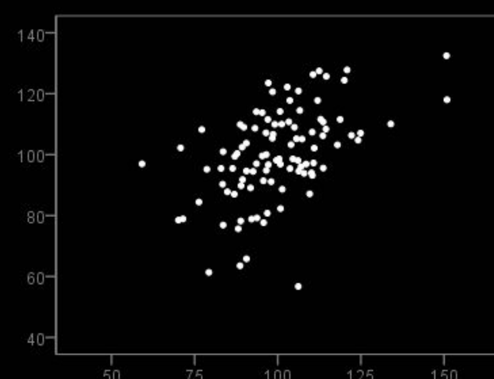
Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

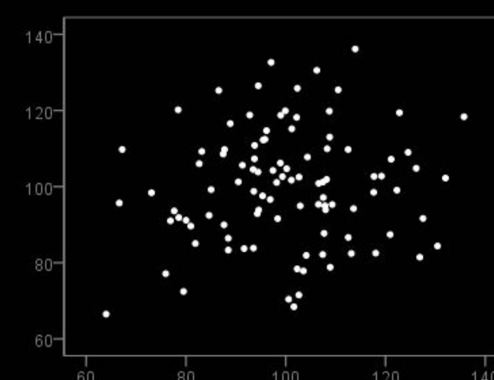
- Pearson Product-Moment Correlation
  - Limitation: Doesn't handle controls



r = -0.8



r = 0.5



r = 0.1

Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  - Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  Fit the model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$

Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation

  Limitation: Doesn't handle controls

- **Standardized** Multivariate Linear Regression

  Fit the model: $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$

  Adjust all variables to have "mean center" and "unit variance":

Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation

  Limitation: Doesn’t handle controls

- **Standardized** Multivariate Linear Regression

  Fit the model: $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$

  Adjust all variables to have “mean center” and “unit variance”:

$$z = \frac{x - \mu}{\sigma}$$

$\mu =$ Mean

$\sigma =$ Standard Deviation

Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  - Limitation: Doesn't handle controls

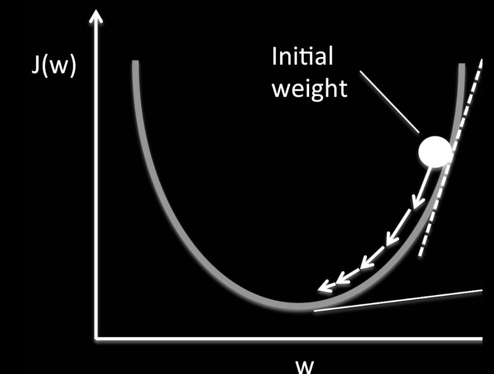
- Standardized Multivariate Linear Regression

  Fit the model:

  $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$

  Option 1: Gradient Descent:

  $J = \sum (y - \hat{y})^2$ -- "Sum of Squares" Error



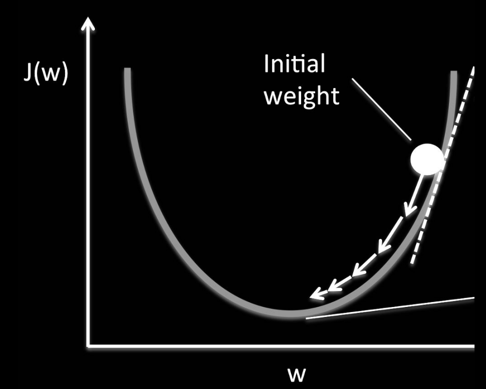
Differential Language Analysis

Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  Fit the model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$
  Option 1: Gradient Descent:
  $J = \sum (y - \hat{y})^2$ -- "Sum of Squares" Error
  Option 2: Matrix model: $Y = X\beta + \epsilon$



Differential Language Analysis

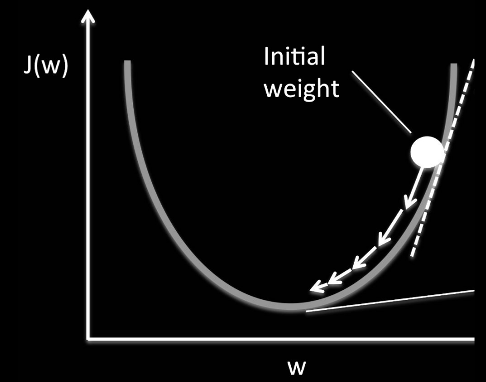
Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  - Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  - Fit the model: $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$
  - Option 1: Gradient Descent:
    - $J = \sum (y - \hat{y})^2$ -- "Sum of Squares" Error
  - Option 2: Matrix model: $Y = X\beta + \epsilon$
  - Matrix Computation Solution:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$



Differential Language Analysis

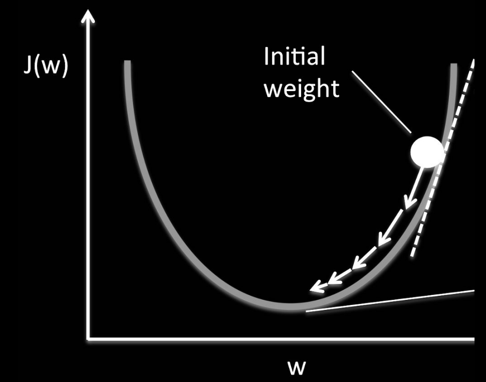
Methods of Correlation Analysis:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- Pearson Product-Moment Correlation
  Limitation: Doesn't handle controls

- Standardized Multivariate Linear Regression
  Fit the model: $$Y_i = \beta_0 + \boxed{\beta_1} X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{m1} + \epsilon_i$$
  Option 1: Gradient Descent:
  $J = \sum (y - \hat{y})^2$ -- "Sum of Squares" Error
  Option 2: Matrix model: $$Y = X\beta + \epsilon$$
  Matrix Computation Solution:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$



Differential Language Analysis

Methods of “Correlation” Analysis for binary outcomes:

- Logistic Regression over Standardized variables

- Odds Ratio

$$\frac{\dfrac{\dfrac{countA(\text{"horrible"})}{NA}}{1-\dfrac{countA(\text{"horrible"})}{NA}}}{\dfrac{\dfrac{countB(\text{"horrible"})}{NB}}{1-\dfrac{countB(\text{"horrible"})}{NB}}}$$

Differential Language Analysis

Methods of “Correlation” Analysis for binary outcomes:

- Logistic Regression over Standardized variables

- Odds Ratio

$$\frac{\dfrac{\frac{countA(\text{"horrible"})}{NA}}{1-\frac{countA(\text{"horrible"})}{NA}}}{\dfrac{\frac{countB(\text{"horrible"})}{NB}}{1-\frac{countB(\text{"horrible"})}{NB}}} \propto log\left(\frac{\frac{countA(\text{"horrible"})}{NA}}{1-\frac{countA(\text{"horrible"})}{NA}}\right) - log\left(\frac{\frac{countB(\text{"horrible"})}{NB}}{1-\frac{countB(\text{"horrible"})}{NB}}\right)$$

Differential Language Analysis

Methods of “Correlation” Analysis for binary outcomes:

- Logistic Regression over Standardized variables
- Odds Ratio

$$\frac{\dfrac{\frac{countA(\text{"horrible"})}{NA}}{1-\frac{countA(\text{"horrible"})}{NA}}}{\dfrac{\frac{countB(\text{"horrible"})}{NB}}{1-\frac{countB(\text{"horrible"})}{NB}}} \propto log\left(\frac{\frac{countA(\text{"horrible"})}{NA}}{1-\frac{countA(\text{"horrible"})}{NA}}\right) - log\left(\frac{\frac{countB(\text{"horrible"})}{NB}}{1-\frac{countB(\text{"horrible"})}{NB}}\right)$$

$$= log\left(\frac{countA(\text{"horrible"})}{NA-countA(\text{"horrible"})}\right) - log\left(\frac{countB(\text{"horrible"})}{NB-countB(\text{"horrible"})}\right)$$

Differential Language Analysis

$$log\left(\frac{countA(\text{"horrible"})}{NA - countA(\text{"horrible"})}\right) - log\left(\frac{countB(\text{"horrible"})}{NB - countB(\text{"horrible"})}\right)$$

- Odds Ratio using Informative Dirichlet Prior

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) - \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right) \qquad (20.9)$$

(Monroe et al., 2010; Jurafsky, 2017)

Differential Language Analysis

$$log\left(\frac{countA(\text{"horrible"})}{NA - countA(\text{"horrible"})}\right) - log\left(\frac{countB(\text{"horrible"})}{NB - countB(\text{"horrible"})}\right)$$

- Odds Ratio using **Informative Dirichlet Prior**

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \boxed{\alpha_w}}{n^i + \boxed{\alpha_0} - (f_w^i + \boxed{\alpha_w})}\right) - \log\left(\frac{f_w^j + \boxed{\alpha_w}}{n^j + \boxed{\alpha_0} - (f_w^j + \boxed{\alpha_w})}\right) \qquad (20.9)$$

(where n^i is the size of corpus i, n^j is the size of corpus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in corpus j, α_0 is the size of the background corpus, and α_w is the count of word w in the background corpus.)

(Monroe et al., 2010; Jurafsky, 2017)

Differential Language Analysis

$$\log\left(\frac{countA(\text{"horrible"})}{NA - countA(\text{"horrible"})}\right) - \log\left(\frac{countB(\text{"horrible"})}{NB - countB(\text{"horrible"})}\right)$$

- Odds Ratio using **Informative Dirichlet Prior**

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \boxed{\alpha_w}}{n^i + \boxed{\alpha_0} - (f_w^i + \boxed{\alpha_w})}\right) - \log\left(\frac{f_w^j + \boxed{\alpha_w}}{n^j + \boxed{\alpha_0} - (f_w^j + \boxed{\alpha_w})}\right) \qquad (20.9)$$

(where n^i is the size of corpus i, n^j is the s … pus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in … is the size of the background corpus, and α_w is the count of word w in … corpus.)

Bayesian term for “smoothing”: accounts for uncertainty as a function of event frequency (i.e. words observed less) by integrating “prior” beliefs mathematically.

(Mo

Differential Language Analysis

$$log\left(\frac{countA(\text{"horrible"})}{NA - countA(\text{"horrible"})}\right) - log\left(\frac{countB(\text{"horrible"})}{NB - countB(\text{"horrible"})}\right)$$

- Odds Ratio using **Informative Dirichlet Prior**

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \boxed{\alpha_w}}{n^i + \boxed{\alpha_0} - (f_w^i + \boxed{\alpha_w})}\right) - \log\left(\frac{f_w^j + \boxed{\alpha_w}}{n^j + \boxed{\alpha_0} - (f_w^j + \boxed{\alpha_w})}\right) \quad (20.9)$$

(where n^i is the size of corpus i, n^j is the s … pus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in … is the size of the background corpus, and α_w is the count of word w in … corpus.)

Bayesian term for “smoothing”: accounts for uncertainty as a function of event frequency (i.e. words observed less) by integrating “prior” beliefs mathematically.
“Informative”: the prior is based on past evidence. Here, the total frequency of the word.

(Mo

Differential Language Analysis

$$log\left(\frac{countA(\text{"horrible"})}{NA-countA(\text{"horrible"})}\right)-log\left(\frac{countB(\text{"horrible"})}{NB-countB(\text{"horrible"})}\right)$$

- Odds Ratio using Informative Dirichlet Prior

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i+\alpha_w}{n^i+\alpha_0-(f_w^i+\alpha_w)}\right) - \log\left(\frac{f_w^j+\alpha_w}{n^j+\alpha_0-(f_w^j+\alpha_w)}\right) \quad (20.9)$$

(where n^i is the size of corpus i, n^j is the size of corpus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in corpus j, α_0 is the size of the background corpus, and α_w is the count of word w in the background corpus.)

(Monroe et al., 2010; Jurafsky, 2017)

Differential Language Analysis

$$log\left(\frac{countA(\text{"horrible"})}{NA - countA(\text{"horrible"})}\right) - log\left(\frac{countB(\text{"horrible"})}{NB - countB(\text{"horrible"})}\right)$$

- Odds Ratio using Informative Dirichlet Prior

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) - \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right) \quad (20.9)$$

(where n^i is the size of corpus i, n^j is the size of corpus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in corpus j, α_0 is the size of the background corpus, and α_w is the count of word w in the background corpus.)

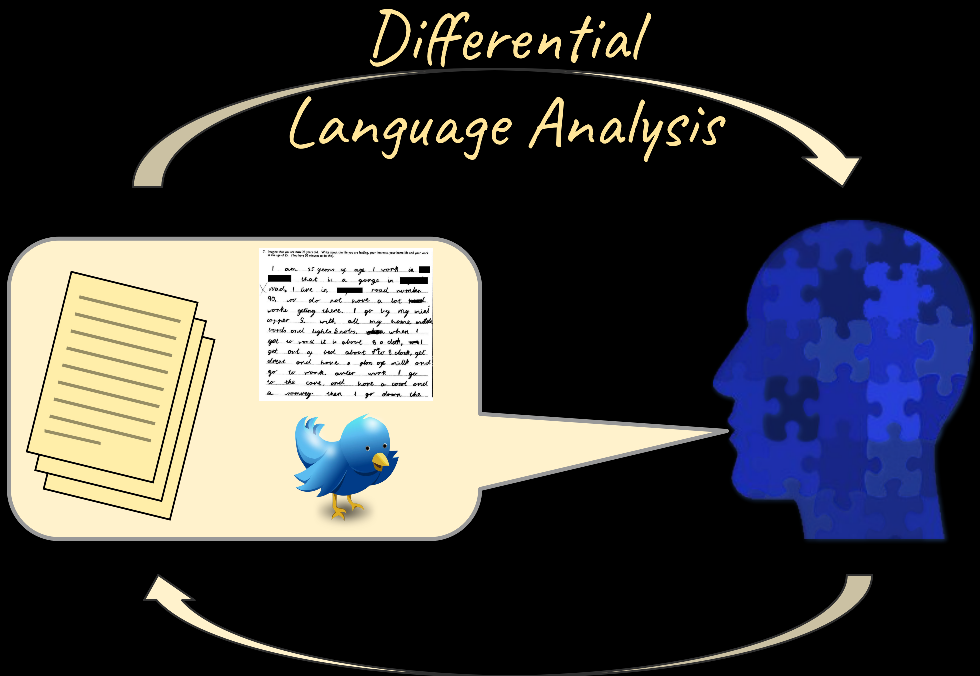
Final score is standardized (z-scored): $\dfrac{\hat{\delta}_w^{(i-j)}}{\sqrt{\sigma^2\left(\hat{\delta}_w^{(i-j)}\right)}}$, where

$$\sigma^2\left(\hat{\delta}_w^{(i-j)}\right) \approx \frac{1}{f_w^i + \alpha_w} + \frac{1}{f_w^j + \alpha_w}$$

(Monroe et al., 2010; Jurafsky, 2017)

DLATK

Python Library, CLI, and Colab for DLA



https://dlatk.github.io/

Getting Started in Colab