Automatic Tone Detection Using Improved Linguistic and Machine Learning Methods

IGII ENVERGA
PROFESSOR PEARL
COMPUTATION OF LANGUAGE LAB
OCTOBER 18, 2013

Introduction

- information extraction research
- surface-level semantic content
- subtle information
  - sentiments and intentions
- mental state expression
- wider net
- textual cues
- humans – computers

Mental State
Deception
Politeness
Rudeness
Embarrassment
Confidence
Disbelief
Formality
Persuading

Related Work

- mental states in comparison to moods and emotions
- Mishne et al (2005)
  - Experiments with Mood Classification in Blog Posts
- Keshtkar et al (2009)
  - Using Sentiment Orientation Features for Mood Classification in Blogs
- machine learning improvements

LiVEJOURNAL

Related Work

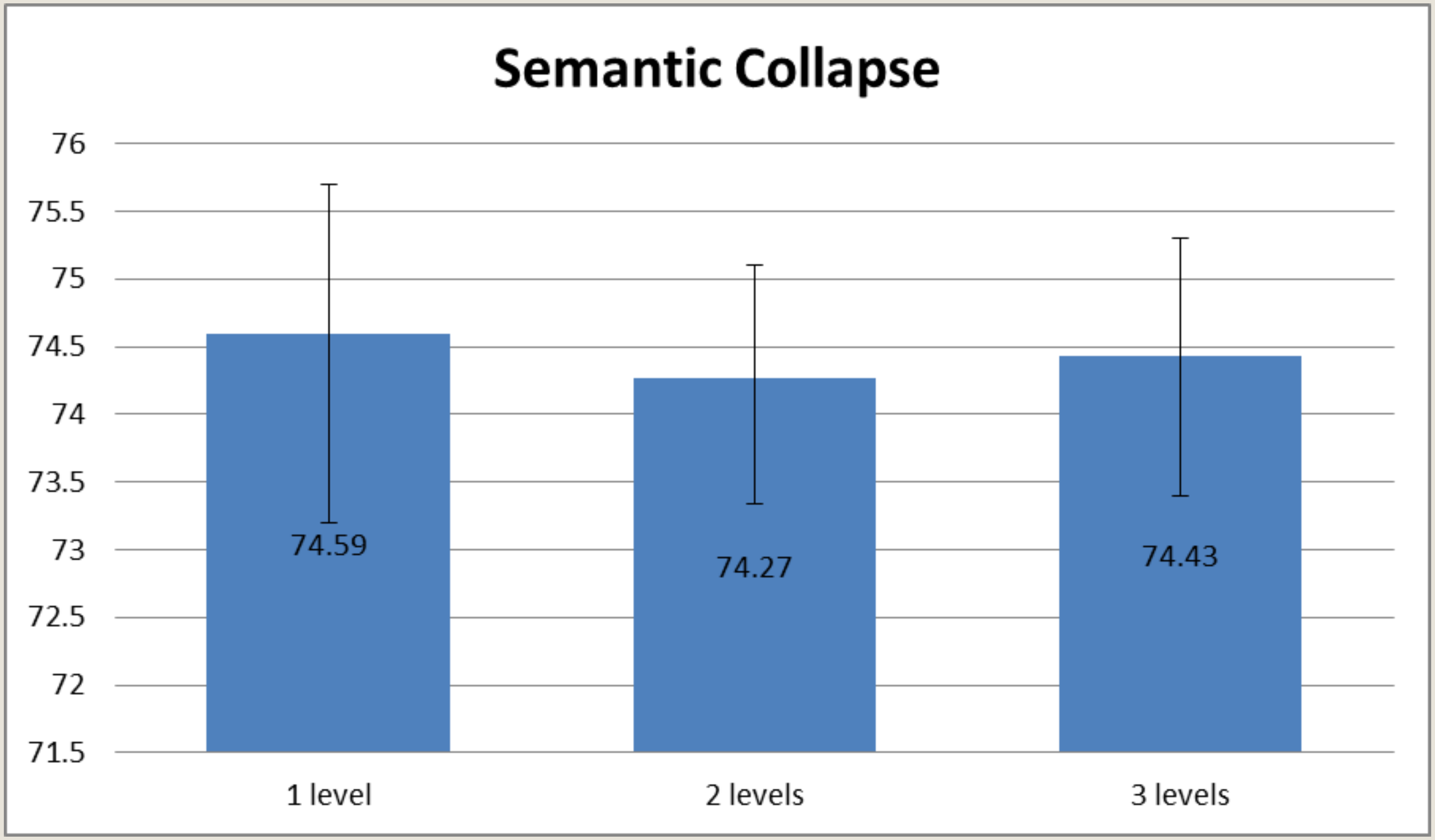
- linguistically informed features
- Danescu et al (2013)
  - A Computational Approach to Politeness with Application to Social Factors
- Pearl & Steyvers (2013)
  - Automatic Identification of Tone from Language Text
- basic content + semantic, syntactic, and valence components

Semantic Collapse

- WordNet
  - collapse words to their hypernyms
- apple -> edible fruit

normal message	level	collapsed message
I think you look really nice in green.	1	I evaluate you look really nice in chromaticcolor.
	2	I think you look really nice in color.
	3	I think you look really nice in visualproperty.

Results

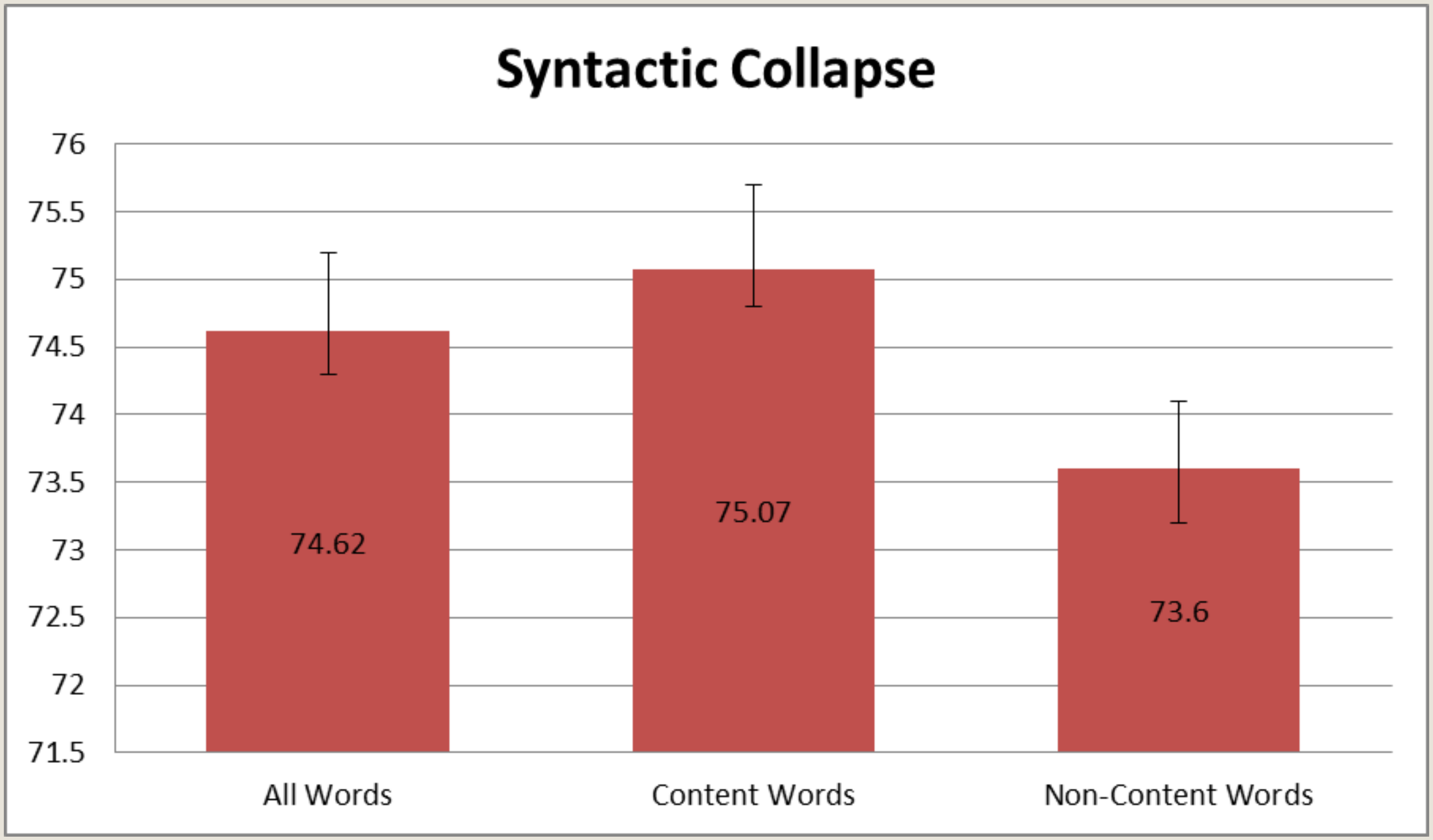

Semantic Collapse
76
75.5
75
74.5
74
73.5
73
72.5
72
71.5
74.59
74.27
74.43
1 level
2 levels
3 levels

Syntactic Collapse

- Stanford's Part of Speech Tagger
  - collapses words into their part of speech
- car -> NN

normal message	type of collapse	collapsed message
Dude she would totally go to prom with me.	all	NNP PRP MD RB VB TO VB IN NN.
	content words	NNP she would RB VB to VB with NN.
	non-content words	Dude PRP MD totally go TO prom IN me.

Results



Syntactic Collapse
76
75.5
75
74.5
74
73.5
73
72.5
72
71.5
74.62
75.07
73.6
All Words
Content Words
Non-Content Words

Valence Collapse

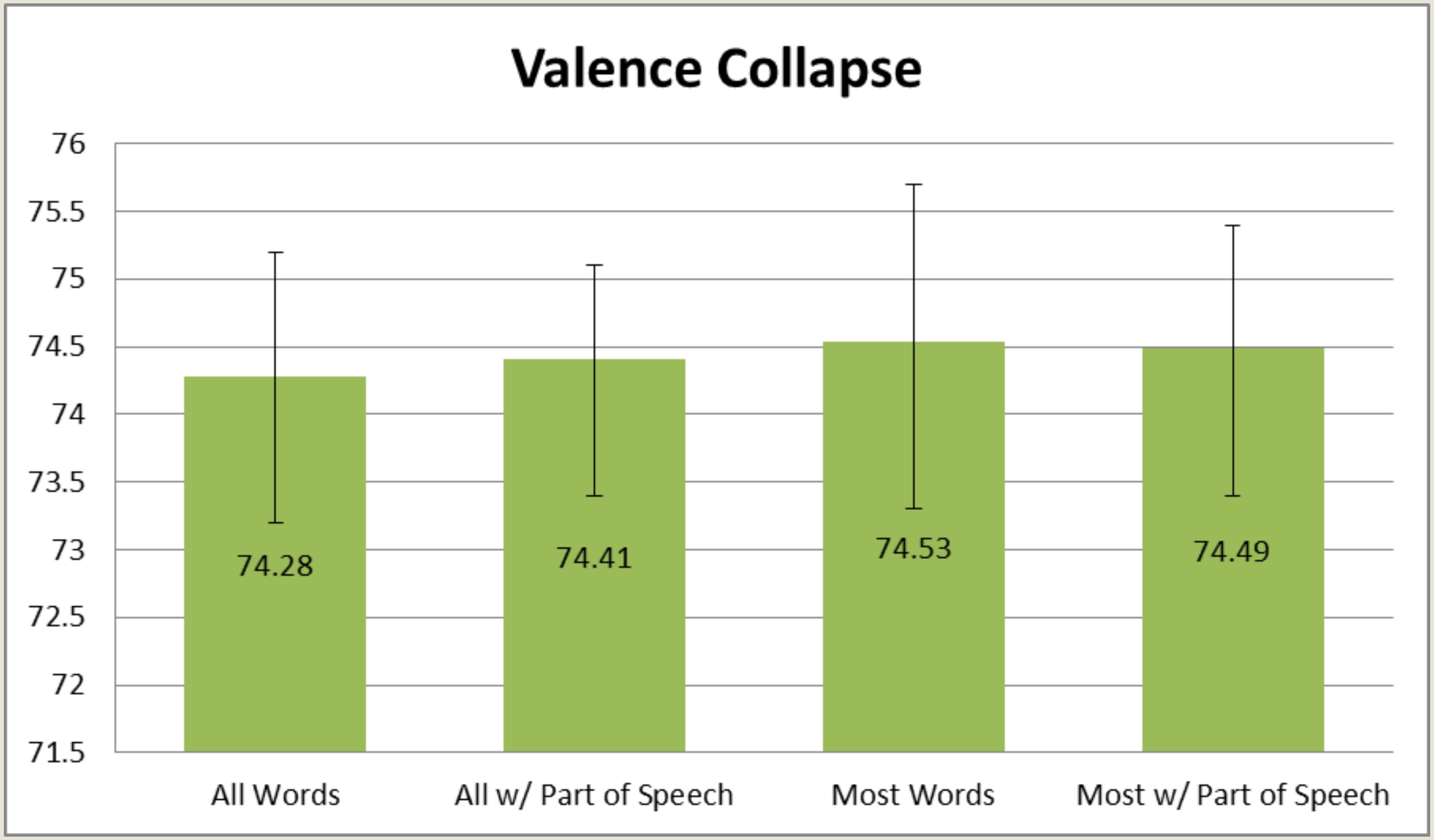
- Affective Ratings from Kuperman et al (2013)
  - collapses words into their valence
- dirty -> negative

8 - positive

0 - negative

normal message	type of collapse	collapsed message
Suck my big toe, you jerk.	all words	negative my positive positive, you negative.
	all w/ part of speech	negverb my posadj posnoun, you negnoun.
	most words	negative my big toe, you negative.
	most w/ part of speech	negverb my big toe, you negnoun.

Results


Valence Collapse
76
75.5
75
74.5
74
73.5
73
72.5
72
71.5
74.28
74.41
74.53
74.49
All Words
All w/ Part of Speech
Most Words
Most w/ Part of Speech

Word Sleuth

GWAP

natural ability of humans to determine tone

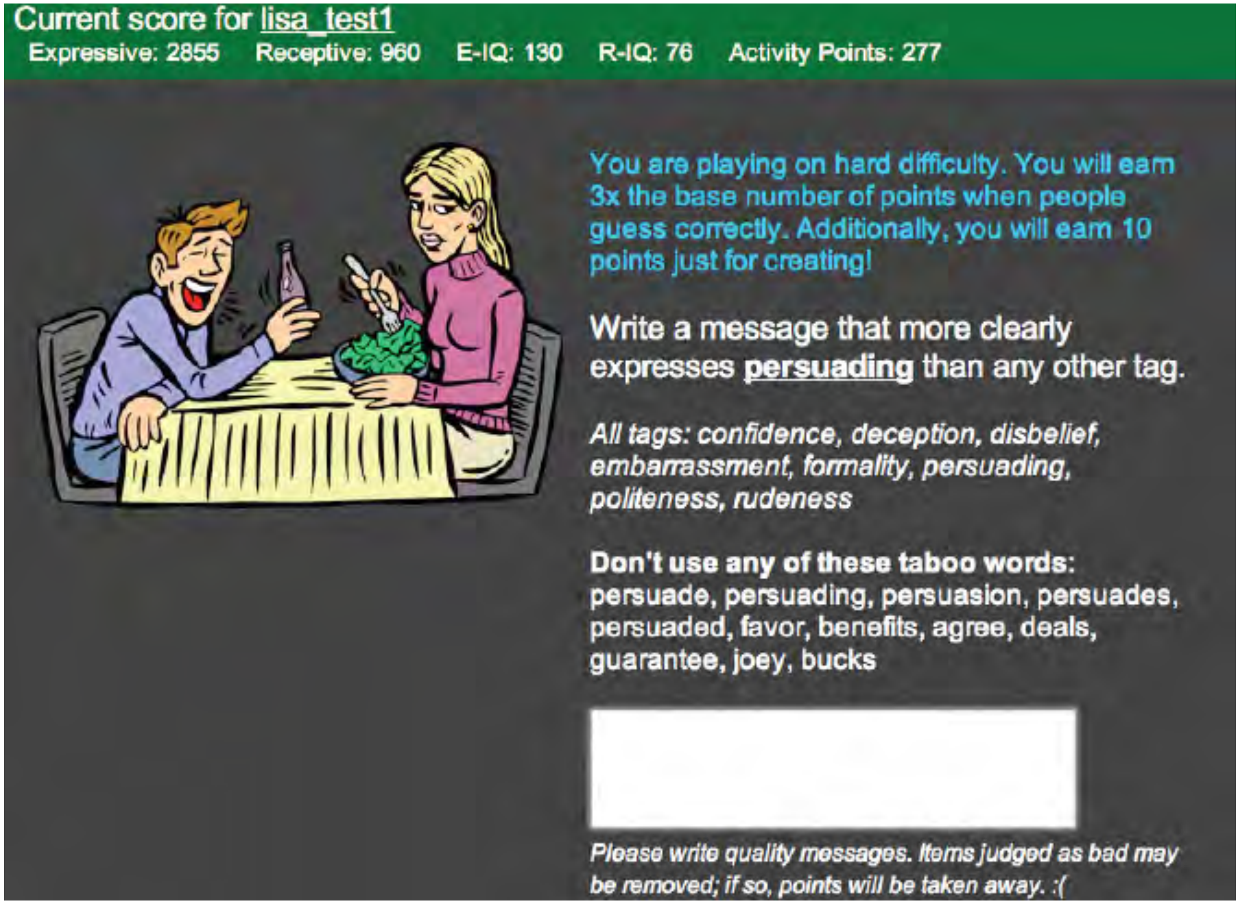
high scores and levels

player rewards

creating messages

labeling messages

8 tones



Expressor Gameplay

Word Sleuth Gameplay

Our Data

	messages	guesses
unfiltered	4839	55577
2+ guesses and >50% accuracy	3349	45312
equal amount for each mental state	1208	15149

mental state	sample bad message
formality	I have to eat you now.
deception	I love the cake you made.

Features

- number of word types
- number of word tokens
- number of sentences
- number of punctuation marks
- average sentence and word length
- word type to word token ratio
- average word log frequency for common words
- valence score

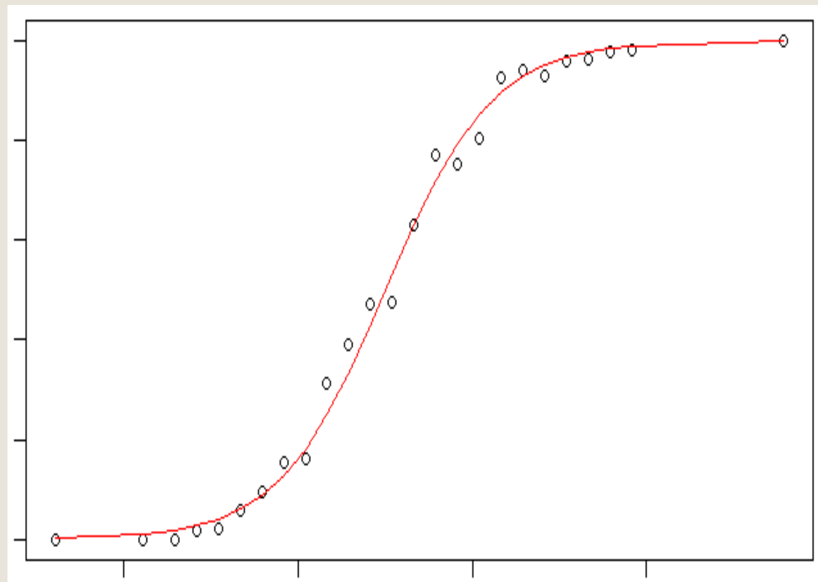
Features

- unigram (“apple”), bigram (“good+day”), and trigram (“I+love+you”) frequencies
  - separate ones for semantic, syntactic, and valence collapses

	unigram	bigram	trigram
normal	apple	good+day	I+love+you
semantic	ediblefruit	good+timeunit	I+love+you
syntactic	NN	JJ+NN	I+VBP+you
valence	positive	positive+positive	I+positive+you

Classifier

- Krishnapuram (2005)
  - Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds
- Sparse Multinomial Logistic Regression (SMLR)
- upweights the useful features that do the work
- zeroes the less useful features



Recall and Precision

- Recall P (labeled deceptive | it is deceptive)
  - Probability that someone guesses that a message is deceptive given that the message is actually deceptive.

	Decep	Polit	Ruden	Embar	Confi	Disbe	Forma	Persu
Deception	31	11	6	15	23	7	5	27

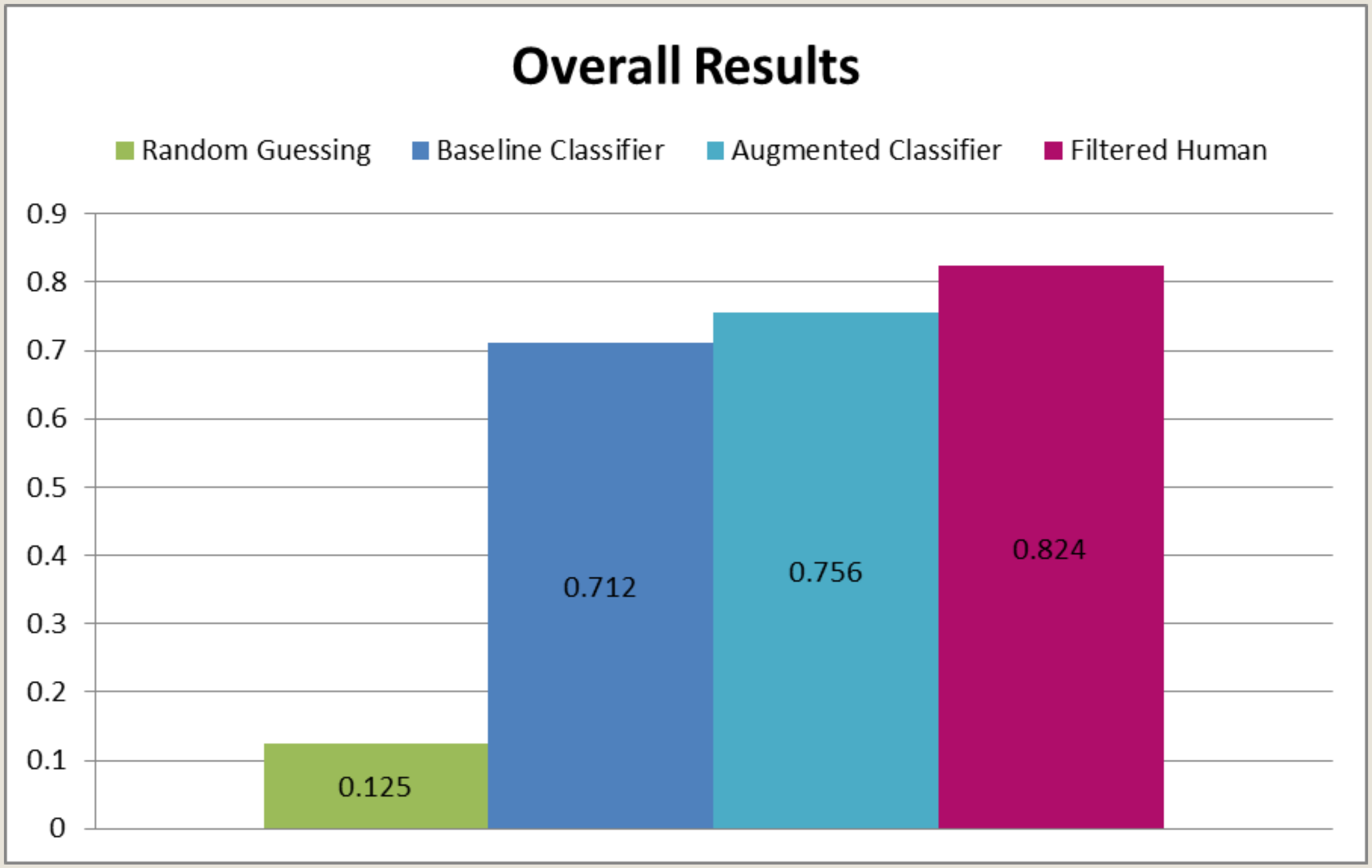
- Precision P (it is deceptive | labeled deceptive)
  - Probability that a message is actually deceptive given that someone guesses that the message is deceptive.

	Deception
Deception	31
Politeness	1
Rudeness	0
Embarrassment	0
Confidence	2
Disbelief	0
Formality	1
Persuasion	0

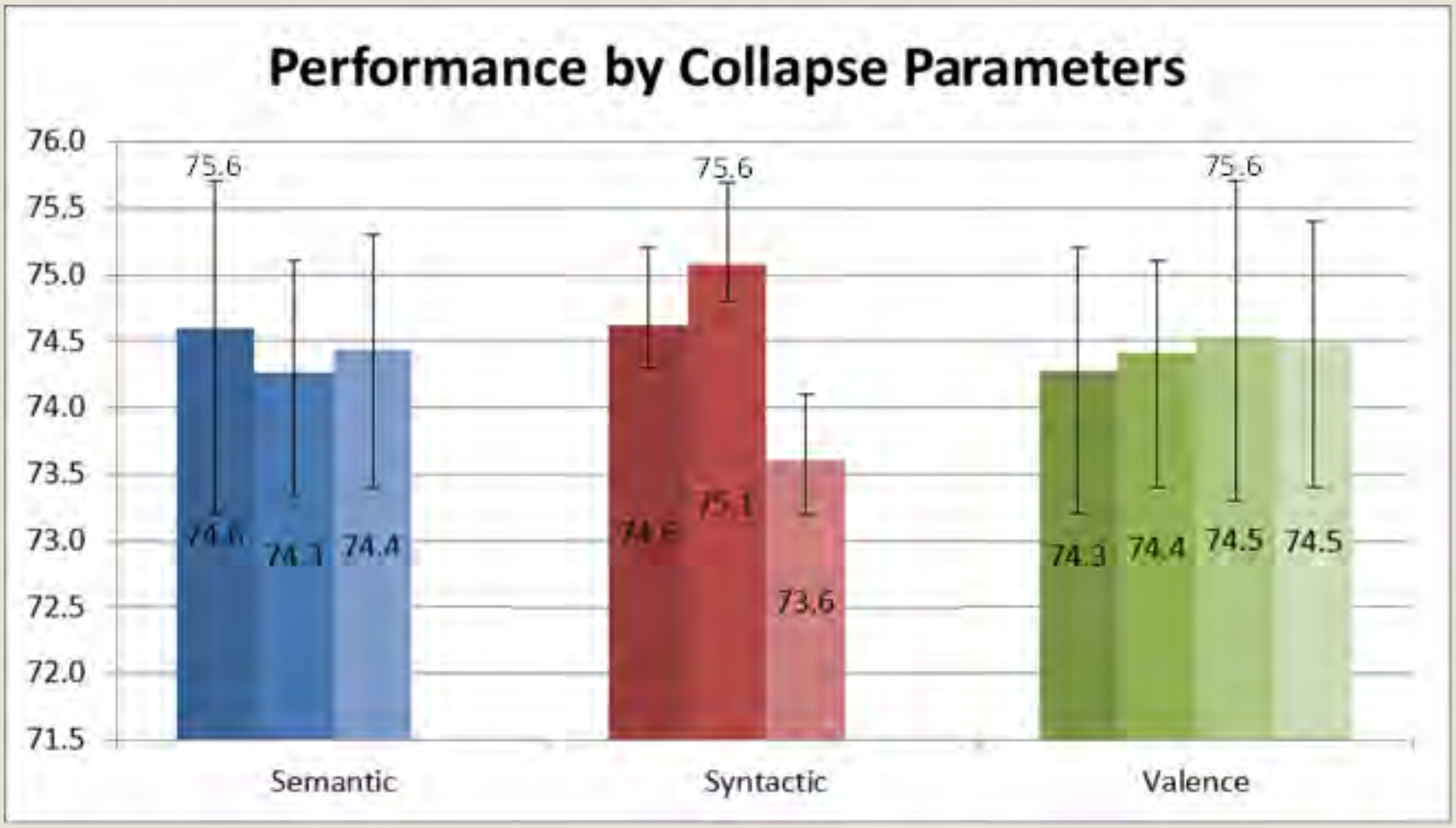
F-Score

- F-Score
  - Combines both scores to give just one number that can easily be compared.
  - $$F1 = 2 * \frac{(precision * recall)}{precision + recall}$$

Results

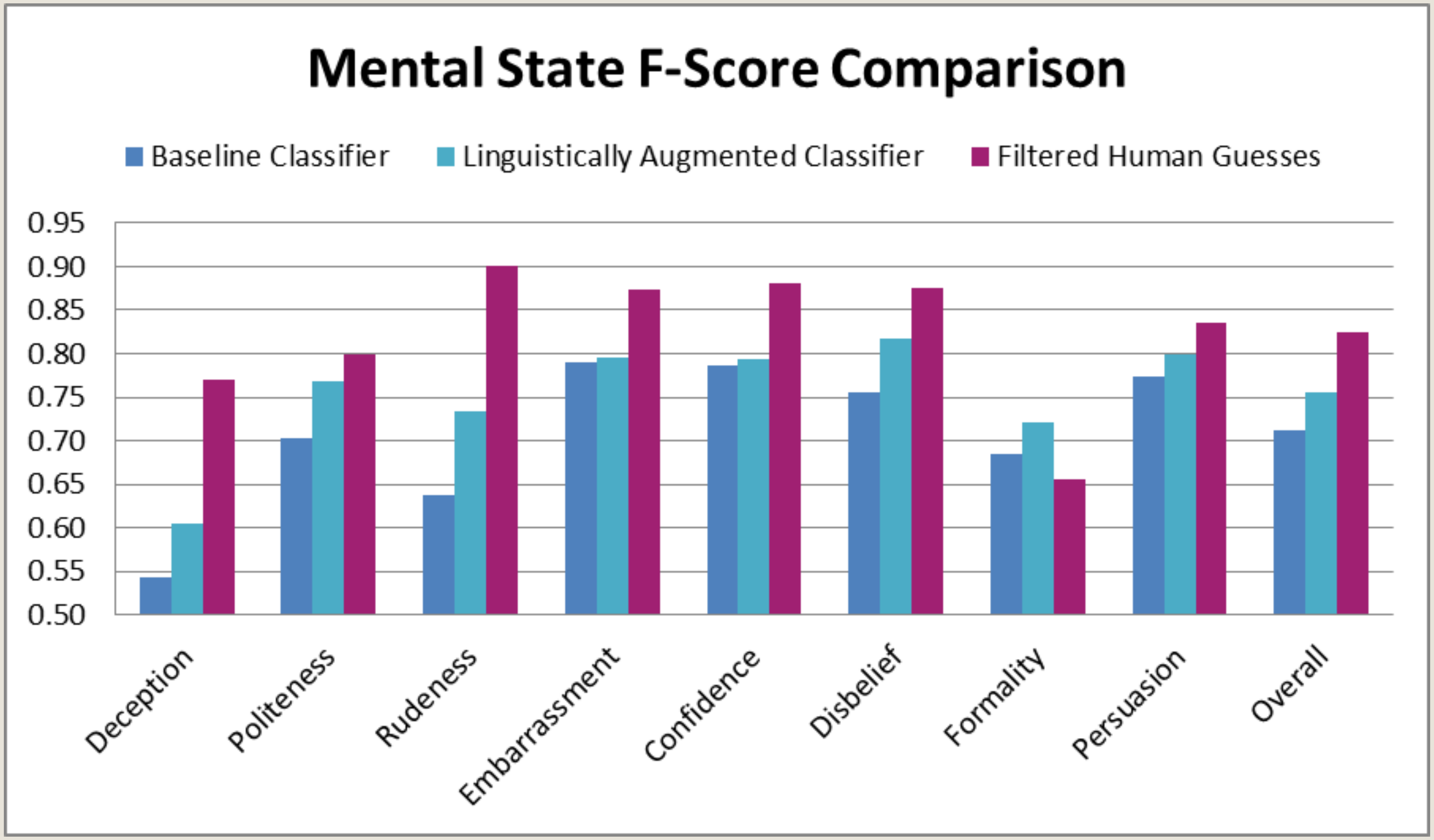

Overall Results
Random Guessing
Baseline Classifier
Augmented Classifier
Filtered Human
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.125
0.712
0.756
0.824

Results



Performance by Collapse Parameters
76.0
75.5
75.0
74.5
74.0
73.5
73.0
72.5
72.0
71.5
75.6
74.6
74.3
74.4
75.6
74.6
75.1
73.6
75.6
74.3
74.4
74.5
74.5
Semantic
Syntactic
Valence

Results



Mental State F-Score Comparison
Baseline Classifier
Linguistically Augmented Classifier
Filtered Human Guesses
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Deception
Politeness
Rudeness
Embarrassment
Confidence
Disbelief
Formality
Persuasion
Overall

Classifier Performance

Formality

- better than humans
- accentuate the formal features
- “Good morning, Mr. Smith.”
  - good+morning
  - mr

Deception

- worse than humans
- what does a deceptive feature look like
- “I’m at the store right now.”
- “I am an apple.”
- not about the content

Results

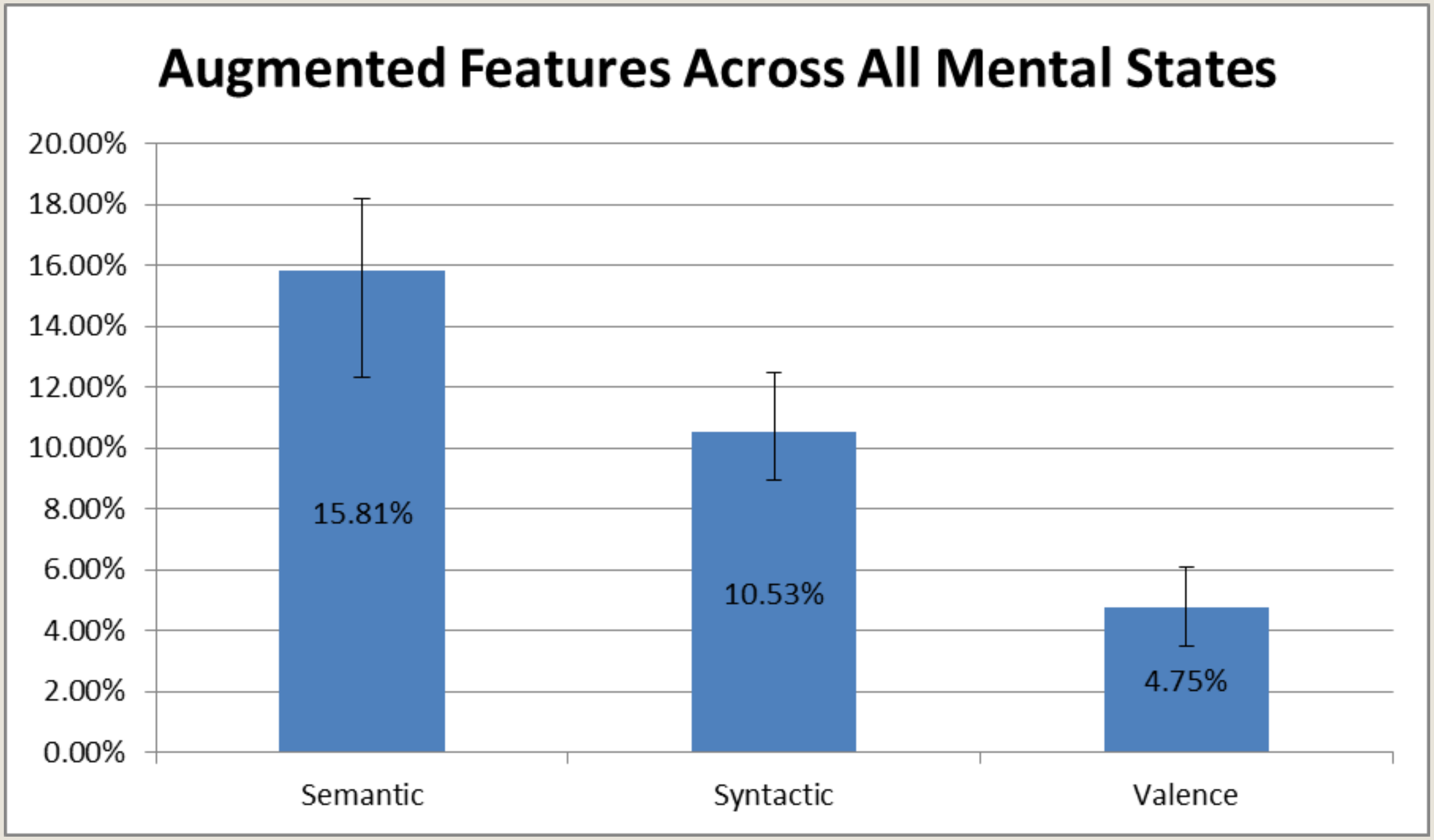
Human

Recall	Dec	Pol	Rud	Emb	Con	Dis	For	Per
Deception	**0.76**	0.02	0.04	0.02	0.03	0.03	0.02	0.08
Politeness	0.01	**0.81**	0.01	0.01	0.02	0.01	0.09	0.03
Rudeness	0.01	0.01	**0.90**	0.02	0.01	0.02	0.01	0.02
Embarrassment	0.02	0.02	0.02	**0.85**	0.01	0.06	0.01	0.01
Confidence	0.02	0.02	0.01	0.01	**0.88**	0.01	0.01	0.05
Disbelief	0.02	0.02	0.04	0.02	0.02	**0.87**	0.01	0.01
Formality	0.01	0.19	0.01	0.01	0.02	0.01	**0.71**	0.04
Persuasion	0.04	0.03	0.02	0.00	0.04	0.01	0.02	**0.84**

Augmented Classifier

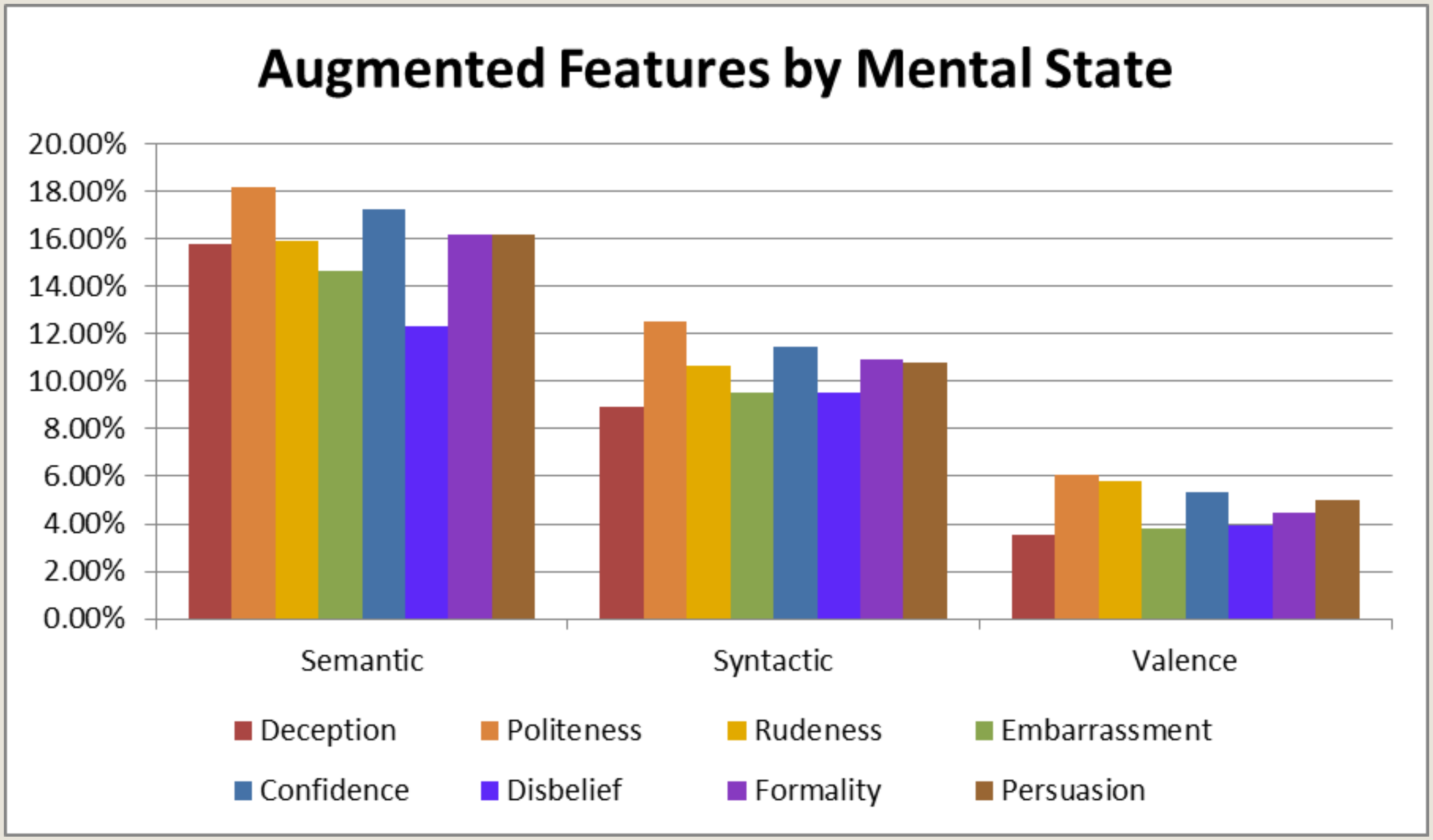
Recall	Dec	Pol	Rud	Emb	Con	Dis	For	Per
Deception	**0.54**	0.01	0.06	0.08	0.12	0.07	0.04	0.09
Politeness	0.01	**0.81**	0.05	0.02	0.02	0.02	0.05	0.03
Rudeness	0.03	0.07	**0.71**	0.04	0.05	0.05	0.01	0.03
Embarrassment	0.07	0.01	0.04	**0.81**	0.02	0.04	0.01	0.01
Confidence	0.05	0.03	0.01	0.03	**0.84**	0.01	0.01	0.03
Disbelief	0.03	0.03	0.01	0.03	0.01	**0.86**	0.01	0.01
Formality	0.04	0.11	0.02	0.03	0.04	0.04	**0.67**	0.05
Persuasion	0.02	0.05	0.03	0.01	0.03	0.01	0.04	**0.82**

Results



Augmented Features Across All Mental States
20.00%
18.00%
16.00%
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
15.81%
10.53%
4.75%
Semantic
Syntactic
Valence

Results



Augmented Features by Mental State
20.00%
18.00%
16.00%
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
Semantic
Syntactic
Valence
Deception
Politeness
Rudeness
Embarrassment
Confidence
Disbelief
Formality
Persuasion

Conclusion

Summary

- adding semantic, syntactic, and valence features helped
- some of these features were more helpful than others
- performance now much closer to humans

Future Work

- domain-specific knowledge of the mental states
- finding classes of words
- branching out to a different data set
  - Live Journal