Subjectivity in Language

- Subjective language is the expression of ***private states***: opinions, sentiments, emotions, evaluations, beliefs, speculations, stances.
- A *private state* is not open to objective observation or verification. [Quirk et al., 1985]
- ***Subjectivity analysis*** is the general task of identifying private states mentioned in text.
- ***Subjectivity classification*** determines whether text is subjective or objective.

Types of Subjectivity

Sentiments: positive or negative emotions, evaluations, stances.

Emotions: emotional state of someone
"I am angry/happy/excited/sad."

Evaluations: emotion or judgement toward something
"Great product!", "What an idiot."
"The economy is in serious trouble"
"This movie is action-packed and thrilling"

Stances: a position taken by an entity
"The University of Utah is against the new policy"

Beliefs: a personal belief
"I think that UFOs are real."

Speculations: speculation, uncertainty, allegations
"I suspect that the butler did it."

Applications

- **Classifying Reviews:** positive/negative labeling of reviews for hotels, movies, restaurants, etc.
- **Product Review Mining:** do people like/dislike a product? What aspects of the product do they like/dislike?
- **Corporate Reputation Tracking:** financial market trend analysis, stock predictions
- **Political Analysis:** tracking opinions toward candidates, predicting election outcomes
- **Opinion Summarization:** summarize the opinions of people over a large set of reviews or documents (e.g., summarize the pros and cons of a product).
- **Multi-perspective Question Answering:** produce answers for questions that have multiple perspectives (e.g., *"What do people think about the government shutdown?"*)

Sentiment Analysis

- ***Sentiment Analysis*** (also called ***Opinion Analysis*** or ***Semantic Orientation***) generally focuses on identifying positive and negative sentiments expressed by an entity.
- Classifiers typically assign ***polarity*** (or ***orientation***) labels:
  - *positive, negative, or neutral.*
- Sentiment analyzers can operate at different levels of granularity: *document classification*, *sentence classification*, identifying *opinion expressions*.

  But … documents and sentences often contain multiple sentiments!

Opinion Extraction

Information extraction systems aim to decompose an opinion into its components:

1. Opinion Expression*:* phrase that describes an attitude toward or evaluation of something
2. Opinion Holder (Source)*:* the entity whose opinion is being expressed (usually a person or organization)
3. Opinion Target*:* the entity, object, or concept that the opinion is about

According to UN officials, the human rights record in Syria is horrendous.

Sentiment Lexicons

Many sentiment lexicons and lists have been created, for example:

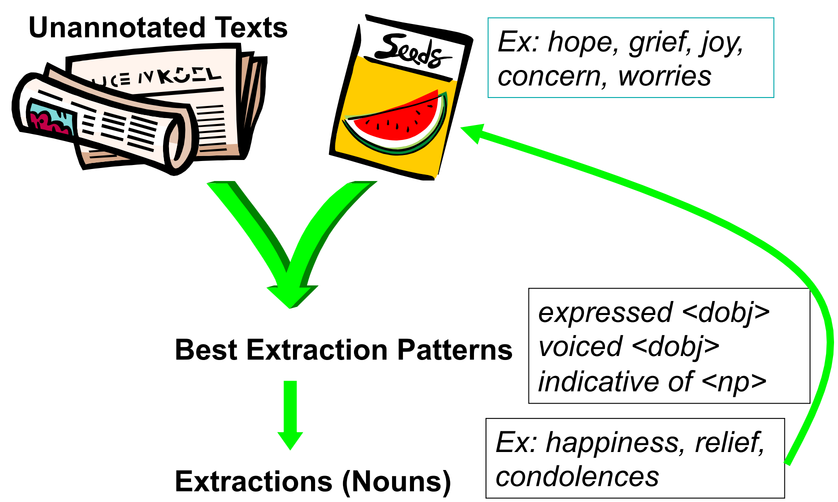
- General (Harvard) Inquirer [Stone et al., 1966]
- Liu et al's opinion lexicon [Liu et al., 2005]
- OpinionFinder lexicon [Wiebe & Riloff, 2005]
- SentiWordNet [Esuli and Sebastiani, 2006]
- Micro-WNOp [Cerini et al. 2007]
- AFINN, designed for microblogs [Nielson, 2011]

Learning Subjective Expressions

[Riloff, Wiebe, Wilson, 2003]

expressed <dobj>	condolences, hope, grief, views, worries
indicative of <np>	compromise, desire, thinking
inject <dobj>	vitality, hatred
reaffirmed <dobj>	resolve, position, commitment
voiced <dobj>	outrage, support, skepticism, opposition, gratitude, indignation
show of <np>	support, strength, goodwill, solidarity
<subj> was shared	anxiety, view, niceties, feeling

Bootstrapped Learning of Subjective Nouns and Expressions



Examples of Strong Subjective Nouns

anguish	exploitation	pariah
antagonism	evil	repudiation
apologist	fallacies	revenge
atrocities	genius	rogue
barbarian	goodwill	sanctimonious
belligerence	humiliation	scum
bully	ill-treatment	smokescreen
condemnation	injustice	sympathy
denunciation	innuendo	tyranny
devil	insinuation	venom
diatribe	liar	
exaggeration	mockery	

Examples of Weak Subjective Nouns

aberration	eyebrows	resistant
allusion	failures	risk
apprehensions	inclination	sincerity
assault	intrigue	slump
beneficiary	liability	spirit
benefit	likelihood	success
blood	peaceful	tolerance
controversy	persistent	trick
credence	plague	trust
distortion	pressure	unity
drama	promise	
eternity	rejection	

Contextual Polarity

- Sentiment lexicons capture the *prior polarity* of words and phrases.
- However, the polarity of a word often depends on context due to polysemy, negation, polarity shifters, scoping, expressions, etc.

Example from [Wilson, Wiebe, & Hoffmann 2005]:

Philip Clapp, president of the National Environment <u>Trust</u>, sums up <u>well</u> the general thrust of the reaction of environmental movements: "There is no <u>reason</u> at all to believe that the <u>polluters</u> are suddenly going to become <u>reasonable</u>.

Why is sentiment analysis so hard?

Subjective language is often among the most colorful and creative! For example:

- Idiosyncratic expressions
  - *"oh well", "good grief", "you are bad", "that's rad"*
- Clausal multi-word expressions
  - *"stepped on [someone's] toes"*
  - *"drove [person] up the wall"*
- Sarcasm
  - *"I'm going to the dentist today, so thrilled."*
  - *"He read about it in the bible of Cat Fancy."*
- World Knowledge
  - *"My new phone has very long battery life."*
  - *"That restaurant always has very long lines."*

Why is sentiment analysis so hard?

- Metaphor
  - *"Parliament attacked ..."*
- Hyperbole
  - *"We wish to see the blood of the opponents..."*
- Rhetorical Argumentation
  - *"The fact is…"*
- Hypotheticals
  - *"If another earthquake hits, further damage to the reactor would be catastrophic."*

Extracting Opinion Propositions and Holders

[Bethard et al., 2004] developed one of the earliest systems to identify propositional opinions and the opinion holders (sources).

- **Opinion**: answer to the question "How does X feel about Y"
- **Propositional Opinion**: an opinion localized in an argument of a verb, generally a sentential complement.
- **Opinion Holder**: the entity who holds the opinion

For example:

- *I believe [you have to use the system to change it].*
- *Still, Vista officials realize [they're relatively fortunate].*
- *["I'd be destroying myself"] replies Mr. Korotich.*

Sentence Classification

- The first step is to classify sentences into 3 categories: NON-OPINION, OPINION-PROPOSITION, or OPINION-SENTENCE.
- An OPINION-SENTENCE contains an opinion that extends beyond the scope of a verb argument.

Examples:

- NON-OPINION: "*I surmise this is because they are unaware of the shape of humans.*" [surmise represents prediction, not a feeling]
- OPINION-PROPOSITION: "*It makes the system more flexible argues a Japanese businessman.*"
- OPINION-SENTENCE: "*It might be imagined by those who are not themselves Anglican that the habit of going to confession is limited only to markedly High churches but that is not necessarily the case.*"

Gold Standard Sentences

Manually annotated sentences as: NON-OPINION, OPINION-PROPOSITION, or OPINION-SENTENCE.

- sentences from FrameNet that have a verbal argument labeled PROPOSITION
- identified verbs in these FrameNet sentences that highly correlated with OPINION sentences.

accuse argue believe castigate chastise comment confirm criticize demonstrate doubt express forget frame know persuade pledge realize reckon reflect reply scream show signal suggest think understand volunteer

- labeled sentences from PropBank that have these verbs

Source	Sentences	NON-OP	OP-PROP	OP-SENT
FrameNet	3,041	1,910	631	573
PropBank	2,098	1,203	618	390

Gold Standard Opinion Holders

- For each OPINION-PROPOSITION sentence, the OPINION-HOLDER was manually labeled.

  "[OPINION-HOLDER *You*] *can argue* [OPINION-PROPOSITION] *these wars are corrective*."

- The authors observed that most opinion holders were the agents of verbs, so all agents were automatically labeled as opinion holders and then mistakes were fixed.
  - Ultimately,10% of the opinion holders were not agents
- For 10% of the sentences, no opinion holder was labeled
  - the opinion holder was the speaker: 6%
  - the opinion holder was not the speaker but implicit: 4%

Opinion Word Features

- Use 1,286 *strong* and 1,687 *weak* subjective nouns learned by Basilisk bootstrapping algorithm [Riloff et al., 2003].
- Acquired new opinion words by computing the ratio of relative frequencies of words in *opinion-heavy* vs. *fact-heavy* articles (mostly WSJ from TREC collections).
  - 2,877 editorials and 1,685 letters to the editor
  - 2,009 business and 3,714 news articles
- Using1,336 manually annotated "*semantically oriented*" adjectives [Hatzivassiloglou & McKeown, 1997], they identified open class words that co-occur with these adjectives using a modified *log-likelihood ratio*. In general:

  log-likelihood ratio = $\log (L(H_1) / L(H_2))$

Opinion Noun Classifier

- They also created a supervised Naïve Bayes classifier to can label any arbitrary noun as FACT or OPINION.
- Manually annotated randomly selected nouns from the TREC corpus and used 500 FACT nouns and 500 OPINION nouns for training.
- The features for a noun are the set of hypernyms in the WordNet hierarchy.
- The classifier was not meant to be sufficient on its own, but is used to further filter opinion noun lists acquired from other methods.

Opinion Word Results

Discovered that different methods worked best for different syntactic classes.

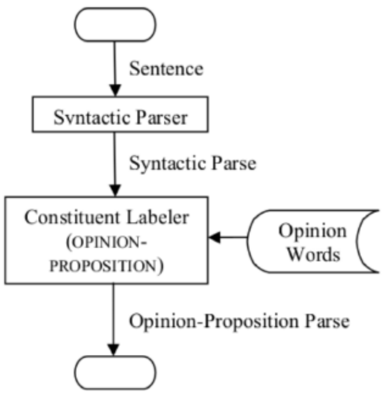
Verbs: fact-heavy vs. opinion-heavy doc freqs worked best.
Nouns & **Adverbs**: adj co-occurrence worked best.
Nouns: WordNet filtering was also applied.
Adjectives: fact-heavy vs. opinion-heavy document freqs was used because it obtained higher recall.

Accuracy using *strong opinion words* as the gold standard:

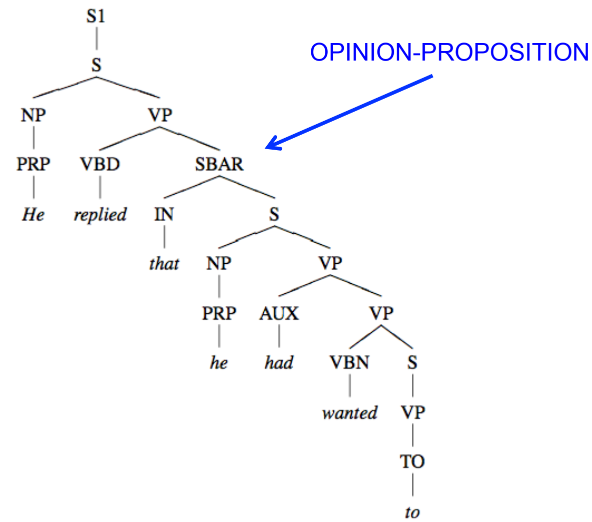
	Subjective	Objective	Precision	Recall
Adj	19,107	14,713	.58	.47
Adv	305	302	.79	.37
Noun	3,188	22,279	.90	.38
Verb	2,329	1,663	.78	.18

One-Tiered Architecture

The first system is an SVM classifier that labels syntactic constituents as either OPINION-PROPOSITION or NULL.



Example



Opinion-Proposition Classifier

They followed the same design as a semantic role labeling classifier by [Pradhan et al., 2003] with 8 syntactic features:

1. the verb
2. verb's cluster
3. subcategorization type of the verb
4. syntactic phrase type of the potential argument
5. head word of the potential argument
6. before/after position of the argument relative to the verb
7. parse tree path between verb and potential argument
8. voice (active/passive) of the verb

This feature set was later augmented with features derived from the acquired opinion words.

Opinion Word Features

Given a constituent to classify, the following features captured opinion word information:

- **Counts**: the number of opinion words in the constituent.
- **Score Sum**: the sum of the opinion scores for each opinion word in the constituent, sometimes with a minimum score threshold.
- **ADJP**: a binary feature indicating whether the constituent contains a complex adjective phrase. (Simple adjectives produce many false hits.) For example:

  excessively affluent
  more bureaucratic

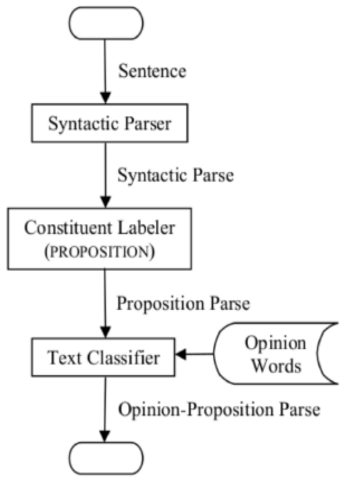
[Note: I've observed "ADV ADJ" to be a useful pattern too.]

Two-Tiered Architecture

The second system performs two steps:

1. A SRL classifier is trained to label constituents only for the PROPOSITION role.

2. A second classifier determines whether the proposition is an OPINION-PROPOSITION, using a sentence-level approach.



Labeling Propositions as Opinions

Three Naïve Bayes classifiers were trained to determine whether a proposition is an OPINION-PROPOSITION.

1. The first model is trained using approximate sentence labels from the fact-heavy vs. opinion-heavy texts.
   - Sentences in editorials and letters to the editor are assumed to contain opinions.
   - Sentences in news and business articles are assumed to be factual.
   - The sentence containing each proposition is classified and the proposition is assigned the label of its sentence.
2. The second model is trained at the sentence-level but predictions are based only on the text of the proposition.
3. For the third model, both training and testing use only the text of the propositions (with the same approximate labeling during training).

Two-Tiered Architecture Training

- All three models used the same set of features:
  - unigrams, bigrams, and trigrams
  - part-of-speech tags
  - presence of opinion and positive/negative words
- The first and second models used:
  - 20,000 random sentences from 2,877 editorials and 3,714 news articles from the WSJ.
- The third model was trained on 5,147 propositions extracted from these documents.
- All three models were evaluated the manually annotated propositions from 5,139 FrameNet and PropBank sentences.

Evaluation Data

The FrameNet and PropBank data were normalized and divided into subsets of 70% for training, 15% for development, and 15% for testing.

Additional models were trained to also identify constituents that correspond to an OPINION-HOLDER, for a 3-way classification task.

The distribution of gold standard constituent labels is:

Dataset	PROPOSITIONAL-OPINION	OPINION-HOLDER	NULL
Training	912	769	89,960
Development	178	149	19,098
Testing	183	162	18,869

Table 2: Distribution of Constituents as Opinion Propositions, Opinion Holders, or Null.

Results for One-Tiered Architecture for 2-Way Classification Task

Features	Precision	Recall
No opinion words	50.97%	43.17%
Counts (external, strong)	50.65%	42.62%
Counts (external, strong+weak)	50.00%	43.72%
Counts (Score $\geq$ 2.0)	52.76%	46.99%
Counts (Score $\geq$ 2.5)	54.66%	48.09%
Counts (Score $\geq$ 3.0)	54.27%	48.63%
Score Sum (Score $\geq$ 0.0)	51.97%	43.17%
Score Sum (Score $\geq$ 2.0)	52.12%	46.99%
Score Sum (Score $\geq$ 2.5)	55.35%	48.09%
Score Sum (Score $\geq$ 3.0)	54.84%	46.45%
ADJP	56.05%	48.09%
ADJP, Score Sum (Score $\geq$ 2.5)	58.02%	51.37%

Table 3: One-Tiered Approach Results for Opinion Propositions

Results for One-Tiered Architecture for 3-Way Classification Task

Features	Precision	Recall
No opinion words	53.43%	42.90%
Counts (external, strong)	51.81%	41.45%
Counts (external, strong+weak)	51.04%	42.61%
Counts (Score $\geq$ 2.0)	54.09%	44.06%
Counts (Score $\geq$ 2.5)	53.90%	44.06%
Counts (Score $\geq$ 3.0)	54.93%	45.22%
Score Sum (Score $\geq$ 0.0)	52.46%	43.19%
Score Sum (Score $\geq$ 2.0)	54.36%	45.22%
Score Sum (Score $\geq$ 2.5)	54.74%	45.22%
Score Sum (Score $\geq$ 3.0)	54.48%	44.06%
ADJP	55.71%	45.22%
ADJP, Score Sum (Score $\geq$ 2.5)	56.75%	47.54%

Table 4: One-Tiered Approach Results for Opinion Propositions and Opinion Holders

Results for Two-Tiered Architecture

The first component that labels PROPOSITION constituents achieved 62% recall with 82% precision. (This was a 10% precision gain over the more general semantic role classifier.)

The results for the 3 models to determine which PROPOSITION constituents are opinions are shown below:

Train on	Predict on	Measure	Features				
			Words	Bigrams	Trigrams	POS	Orientation
Sentence	Sentence	Recall	33.38%	29.69%	30.09%	30.05%	43.72%
		Precision	67.84%	63.13%	62.50%	65.55%	67.97%
Sentence	Proposition	Recall	37.48%	37.32%	37.79%	36.03%	28.81%
		Precision	53.95%	59.00%	59.83%	55.00%	68.41%
Proposition	Proposition	Recall	42.77%	38.07%	37.84%	35.01%	25.75%
		Precision	59.56%	61.63%	60.43%	58.77%	61.66%

Table 5: Two-tiered Approach Results for Opinion Propositions.

Summary

- This work focused on one type of opinion recognition, propositional opinions, and identified the opinion holders.
- This approach is very syntactically-oriented, requiring an alignment between the propositions/holders and syntactic constituents.
  - This approach cannot identify cases where a proposition spans multiple sentences, or the holder is in a different sentence than the proposition.
- The two architectures exhibited a recall/precision trade-off:

  51% R with 58% P for 1 Tiered

  43% R with 68% P for 2 Tiered.
- The automatically learned opinion words improved performance and complex ADJPs proved to be useful.