Manual and Automatic Subjectivity and Sentiment Analysis

Jan Wiebe

University of Pittsburgh

- This tutorial covers topics in manual and automatic subjectivity and sentiment analysis
- Work of many groups
- But I want to start with acknowledgments to colleagues and students in our group

CERATOPS

Center for Extraction and Summarization of Events and Opinions in Text

Jan Wiebe, U. Pittsburgh

Claire Cardie, Cornell U.

Ellen Riloff, U. Utah

Word Sense and Subjectivity

Learning Multi-Lingual Subjective Language

Rada Mihalcea

Jan Wiebe

Our Student Co-Authors in Subjectivity and Sentiment Analysis

• Carmen Banea	North Texas
• Eric Breck	Cornell
• Yejin Choi	Cornell
• Paul Hoffman	Pittsburgh
• Wei-Hao Lin	CMU
• Sidd Patwardhan	Utah
• Bill Phillips	Utah
• Swapna Somasundaran	Pittsburgh
• Ves Stoyanov	Cornell
• Theresa Wilson	Pittsburgh

Preliminaries

- What do we mean by subjectivity?
- The linguistic expression of somebody's emotions, sentiments, evaluations, opinions, beliefs, speculations, etc.
  - **Wow**, this is my 4th Olympus camera.
  - Staley **declared** it to be "**one hell of** a collection".
  - Most voters **believe** that he's not going to raise their taxes

One Motivation

- Automatic question answering…

Fact-Based Question Answering

- Q: When is the first day of spring in 2007?

- Q: Does the us have a tax treaty with cuba?

Fact-Based Question Answering

- Q: When is the first day of spring in 2007?
- A: **March 21**
- Q: Does the US have a tax treaty with Cuba?
- A: **[Thus,] the U.S. has no tax treaties with nations like Iraq and Cuba**.

Opinion Question Answering

Q: What is the international reaction to the reelection of Robert Mugabe as President of Zimbabwe?

A: African observers generally approved of his victory while Western Governments denounced it.

More motivations

- **Product review mining:** What features of the ThinkPad T43 do customers like and which do they dislike?
- **Review classification:** Is a review positive or negative toward the movie?
- **Tracking sentiments toward topics over time:** Is anger ratcheting up or cooling down?
- **Etc.**

Foci of this Talk

- Lower-level linguistic expressions rather than whole sentences or documents

- Developing an understanding of the problem rather than trying to implement a particular solution

Outline

- Corpus Annotation
- Pure NLP
  - Lexicon development
  - Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis
- Applications
  - Product review mining
- Citations

Corpus Annotation

Wiebe, Wilson, Cardie 2005

Annotating Expressions of Opinions and Emotions in Language

Overview

- Fine-grained: **expression-level** rather than sentence or document level
  - The photo quality was the best that I have seen in a camera.
  - The photo quality was the best that I have seen in a camera.

Overview

- Fine-grained: **expression-level** rather than sentence or document level
  - The photo quality was the best that I have seen in a camera.
  - The photo quality was the best that I have seen in a camera.
- Annotate
  - expressions of **opinions**, **evaluations**, **emotions, beliefs**
  - material attributed to a source, but presented objectively

Overview

- Opinions, evaluations, emotions, speculations are private states.
- They are expressed in language by subjective expressions.

Private state: state that is not open to objective observation or verification.

Quirk, Greenbaum, Leech, Svartvik (1985). *A Comprehensive Grammar of the English Language*.

Overview

- Focus on three ways private states are expressed in language
  - Direct subjective expressions
  - Expressive subjective elements
  - Objective speech events

Direct Subjective Expressions

- Direct mentions of private states

  The United States **fears** a spill-over from the anti-terrorist campaign.

- Private states expressed in speech events

  “We foresaw electoral fraud but not daylight robbery,” Tsvangirai **said**.

Expressive Subjective Elements

[Banfield 1982]

- "We foresaw electoral fraud but not **daylight robbery**," Tsvangirai said
- The part of the US human rights report about China is **full of absurdities and fabrications**

Objective Speech Events

- Material attributed to a source, but presented as objective fact

The government, it **added**, has amended the Pakistan Citizenship Act 10 of 1951 to enable women of Pakistani descent to claim Pakistani nationality for their children born to foreign husbands.


INFORMED SOURCE

Nested Sources

“The report is full of absurdities,’’ Xirao-Nima said the next day.

Nested Sources

(Writer)

“The report is full of absurdities,’’ Xirao-Nima said the next day.

Nested Sources

(Writer, Xirao-Nima)

“The report is full of absurdities,’’ Xirao-Nima said the next day.

Nested Sources

(Writer Xirao-Nima) (Writer Xirao-Nima)

"The report is full of absurdities,'' Xirao-Nima said the next day.

Nested Sources

(Writer)

(Writer Xirao-Nima)

(Writer Xirao-Nima)

"The report is full of absurdities," Xirao-Nima said the next day.

"The report is full of absurdities," Xirao-Nima said the next day.

Objective speech event
anchor: *the entire sentence*
source: <writer>
implicit: true

Direct subjective
anchor: *said*
source: <writer, Xirao-Nima>
intensity: high
expression intensity: neutral
attitude type: negative
target: report

Expressive subjective element
anchor: *full of absurdities*
source: <writer, Xirao-Nima>
intensity: high
attitude type: negative

“The report is full of absurdities,” Xirao-Nima said the next day.

Objective speech event
anchor: *the entire sentence*
source: <writer>
implicit: true

Direct subjective
anchor: *said*
source: <writer, Xirao-Nima>
intensity: high
expression intensity: neutral
attitude type: negative
target: report

Expressive subjective element
anchor: *full of absurdities*
source: <writer, Xirao-Nima>
intensity: high
attitude type: negative

"The report is full of absurdities," Xirao-Nima said the next day.

Objective speech event
anchor: *the entire sentence*
source: <writer>
implicit: true

Direct subjective
anchor: *said*
source: <writer, Xirao-Nima>
intensity: high
expression intensity: neutral
attitude type: negative
target: report

Expressive subjective element
anchor: *full of absurdities*
source: <writer, Xirao-Nima>
intensity: high
attitude type: negative

"The report is full of absurdities," Xirao-Nima said the next day.

Objective speech event
anchor: *the entire sentence*
source: <writer>
implicit: true

Direct subjective
anchor: *said*
source: <writer, Xirao-Nima>
intensity: high
expression intensity: neutral
attitude type: negative
target: report

Expressive subjective element
anchor: *full of absurdities*
source: <writer, Xirao-Nima>
intensity: high
attitude type: negative

"The report is full of absurdities," Xirao-Nima said the next day.

Objective speech event
anchor: *the entire sentence*
source: <writer>
implicit: true

Direct subjective
anchor: *said*
source: <writer, Xirao-Nima>
intensity: high
expression intensity: neutral
attitude type: negative
target: report

Expressive subjective element
anchor: *full of absurdities*
source: <writer, Xirao-Nima>
intensity: high
attitude type: negative

"The report is full of absurdities," Xirao-Nima said the next day.

Objective speech event
anchor: *the entire sentence*
source: <writer>
implicit: true

Direct subjective
anchor: *said*
source: <writer, Xirao-Nima>
intensity: high
expression intensity: neutral
attitude type: negative
target: report

Expressive subjective element
anchor: *full of absurdities*
source: <writer, Xirao-Nima>
intensity: high
attitude type: negative

“The US fears a spill-over’’, said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.

(Writer)

"The US fears a spill-over'', said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.

(writer, Xirao-Nima)

“The US fears a spill-over’’, said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.

(writer, Xirao-Nima, US)

"The US fears a spill-over'', said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.

(Writer)

(writer, Xirao-Nima, US)

(writer, Xirao-Nima)

"The US fears a spill-over'', said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.

"The US fears a spill-over'', said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.

Objective speech event
anchor: *the entire sentence*
source: <writer>
implicit: true

Objective speech event
anchor: *said*
source: <writer, Xirao-Nima>

Direct subjective
anchor: *fears*
source: <writer, Xirao-Nima, US>
intensity: medium
expression intensity: medium
attitude type: negative
target: spill-over

The report has been strongly criticized and condemned by many countries.

The report has been strongly criticized and condemned by many countries.

Objective speech event
anchor: *the entire sentence*
source: <writer>
implicit: true

Direct subjective
anchor: *strongly criticized and condemned*
source: <writer, many-countries>
intensity: high
expression intensity: high
attitude type: negative
target: report

As usual, the US state Department published its annual report on human rights practices in world countries last Monday.

And as usual, the portion about China contains little truth and many absurdities, exaggerations and fabrications.

As usual, the US state Department published its annual report on human rights practices in world countries last Monday.

And as usual, the portion about China contains little truth and many absurdities, exaggerations and fabrications.

Objective speech event
anchor : *the entire 1st sentence*
source : <writer>
implicit : true

Direct subjective
anchor : *the entire 2nd sentence*
source : <writer>
implicit : true
intensity : high
expression intensity : medium
attitude type : negative
target : report

Expressive subjective element
anchor : *And as usual*
source : <writer>
intensity : low
attitude type : negative

Expressive subjective element
anchor : *little truth*
source : <writer>
intensity : medium
attitude type : negative

Expressive subjective element
anchor : *many absurdities, exaggerations, and fabrications*
source : <writer>
intensity : medium
attitude type : negative

Corpus

- www.cs.pitt.edu/mqpa/databaserelease (version 2)
- English language versions of articles from the world press (*187 news sources)*
- Also includes contextual polarity annotations (later)
- Themes of the instructions:
  - No rules about how particular words should be annotated.
  - Don't take expressions out of context and think about what they *could* mean, but judge them as they are used in that sentence.

Agreement

- Inter-annotator agreement studies performed on various aspects of the scheme

Agreement

Annotator 1

Two council street wardens who helped lift a 14-ton bus off an injured schoolboy are to be especially commended for their heroic actions.

Nathan Thomson and Neville Sharpe will receive citations from the mayor of Croydon later this month.

Annotator 2

Two council street wardens who helped lift a 14-ton bus off an injured schoolboy are to be especially commended for their heroic actions.

Nathan Thomson and Neville Sharpe will receive citations from the mayor of Croydon later this month.

Agreement

- Inter-annotator agreement studies performed on various aspects of the scheme
- Kappa is a measure of the degree of nonrandom agreement between observers and/or measurements of a specific categorical variable
- Kappa values range between .70 and .80

Extensions

Wilson 2007

Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states

Table 6.2: Measures of intensity for different attitude types.

Attitude Type	Measure of Intensity	Example
Positive Sentiment	degree of positiveness	*like* < *love*
Negative Sentiment	degree of negativeness	*criticize* < *excoriate*
Positive Agreement	degree of agreement	*mostly agree* < *agree*
Negative Agreement	degree of disagreement	*mostly disagree* < *completely disagree*
Positive Arguing	degree of certainty/strength of belief	*critical* < *absolutely critical*
Negative Arguing	degree of certainty/strength of belief	*should not* < *really should not*
Positive Intention	degree of determination	*promise* < *promise with all my heart*
Negative intention	degree of determination	*no intention* < *absolutely no intention*
Speculation	degree of likelihood	*might win* < *really might win*

Extensions

Wilson 2007

I think people are happy because Chavez has fallen.

direct subjective
span: *think*
source: <writer, I>
attitude:

attitude
span: *think*
type: **positive arguing**
intensity: medium
target:

target
span: *people are happy because Chavez has fallen*

direct subjective
span: *are happy*
source: <writer, I, People>
attitude:

attitude
span: *are happy*
type: pos sentiment
intensity: medium
target:

target
span: *Chavez has fallen*

inferred attitude
span: *are happy because Chavez has fallen*
type: neg sentiment
intensity: medium
target:

target
span: *Chavez*

Outline

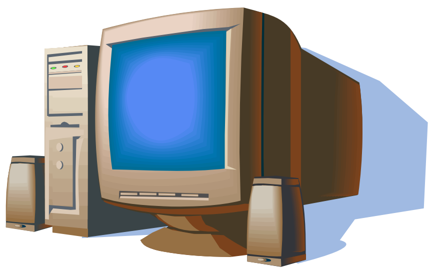
- Corpus Annotation
- Pure NLP
  - Lexicon development
  - Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis
- Applications
  - Product review mining

Who does lexicon development ?

- Humans

- Semi-automatic

- Fully automatic



What?

- Find relevant words, phrases, patterns that can be used to express subjectivity
- Determine the polarity of subjective expressions

JMW1

Words

- Adjectives (e.g. Hatzivassiloglou & McKeown 1997, Wiebe 2000, Kamps & Marx 2002, Andreevskaia & Bergler 2006)
  – positive: honest important mature large patient
    - Ron Paul is the only honest man in Washington.
    - Kitchell's writing is unbelievably mature and is only likely to get better.
    - To humour me my patient father agrees yet again to my choice of film

JMW1 Many people have worked on finding adjectives

Janyce M. Wiebe, 19/07/2007

Words

- Adjectives (e.g. Hatzivassiloglou & McKeown 1997, Wiebe 2000, Kamps & Marx 2002, Andreevskaia & Bergler 2006)
  - positive
  - negative: harmful hypocritical inefficient insecure
    - It was a macabre and hypocritical circus.
    - Why are they being so inefficient ?

Words

- Adjectives (e.g. Hatzivassiloglou & McKeown 1997, Wiebe 2000, Kamps & Marx 2002, Andreevskaia & Bergler 2006)
  - positive
  - negative
  - Subjective (but not positive or negative sentiment): curious, peculiar, odd, likely, probable
    - He spoke of Sue as his probable successor.
    - The two species are likely to flower at different times.

- Other parts of speech (e.g. Turney & Littman 2003, Riloff, Wiebe & Wilson 2003, Esuli & Sebastiani 2006)
  - Verbs
    - positive: **praise, love**
    - negative: **blame, criticize**
    - subjective: **predict**
  - Nouns
    - positive: **pleasure, enjoyment**
    - negative: **pain, criticism**
    - subjective: **prediction, feeling**

Phrases

- Phrases containing adjectives and adverbs (e.g. Turney 2002, Takamura, Inui & Okumura 2007)
  - positive: high intelligence, low cost
  - negative: little variation, many troubles

Patterns

- Lexico-syntactic patterns (Riloff & Wiebe 2003)
- way with <np>: … to ever let China use force to have its way with …
- expense of <np>: at the expense of the world's security and stability
- underlined <dobj>: Jiang's subdued tone … underlined his desire to avoid disputes …

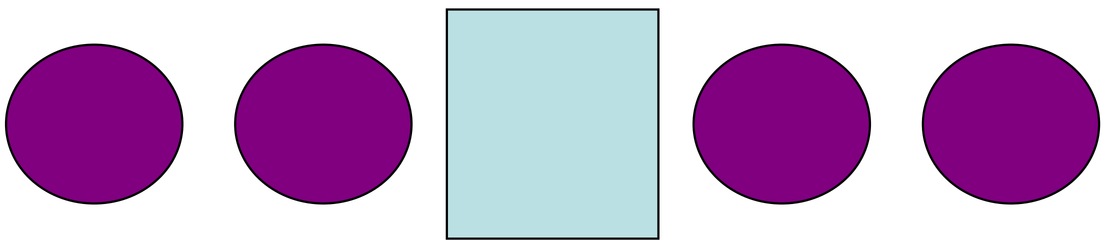
How?

- How do we identify subjective items?

How?

- How do we identify subjective items?

- Assume that contexts are coherent



Conjunction

Web Results **1** - **10** of about **762,000** for **"was very nice and"**.

The Homestay Experience - Cultural Kaleidoscope 2006
My host's home **was very nice and** comfortable. I got to try all types of food; Malaysian, Chinese, Indonesian and I loved it all. My host's parents were very **...**
www.gardenschool.edu.my/studentportal/aec/Kaleidoscope06/experience.asp - 10k - Cached - Similar pages - Note this

PriceGrabber User Rating for Watch Your Budget - PriceGrabber.com
Reviews, Camera I purchased **was very nice and** a bargain. There was a problem with shipping, but was resolved quickly. Buy with confidence from this vendor. **...**
www.pricegrabber.com/rating_getreview.php/retid=5821 - Similar pages - Note this

Testimonials
"Everybody **was very nice and** service was as fast as they possibly could. **...** "Staff member who helped me **was very nice and** easy to talk to." **...**
www.sa.psu.edu/uhs/news/testimonials.cfm - 22k - Cached - Similar pages - Note this

Naxos Villages - Naxos Town or Chora Reviews: Very nice and very **...**
-Did you enjoy the trip to Naxos Town: Yes it **was very nice and** very scenic -In order to get to the village were there enough signs in order to find it: It **...**

Statistical association

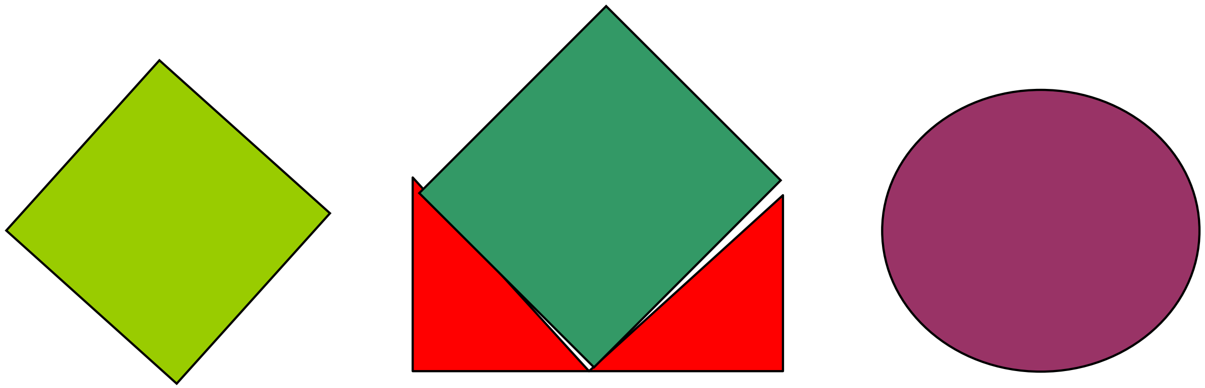
- If words of the same orientation like to co-occur together, then the presence of one makes the other more probable

- Use statistical measures of association to capture this interdependence
  - E.g., Mutual Information (Church & Hanks 1989)

How?

- How do we identify subjective items?

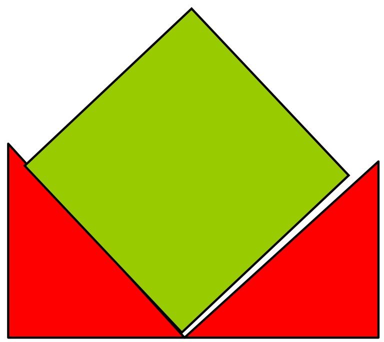
- Assume that contexts are coherent
- Assume that alternatives are similarly subjective

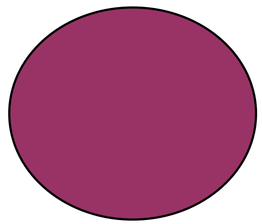


How?

- How do we identify subjective items?

- Assume that contexts are coherent
- Assume that alternatives are similarly subjective





WordNet

- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*

WordNet

- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*

WordNet relations

- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*

WordNet relations

- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*

WordNet relations

- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*

WordNet glosses

- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*

WordNet examples

- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*

How? Summary

- How do we identify subjective items?

- Assume that contexts are coherent
- Assume that alternatives are similarly subjective
- Take advantage of word meanings

Home | Concordance | Word List | Word Sketch | Thesaurus | Sketch-Diff

cause British National Corpus freq = 20207

object	15651	5.8
damage	938	10.09
harm	276	8.91
injury	295	8.38
problem	1014	8.37
trouble	249	8.32
death	383	7.96
delay	146	7.87
confusion	137	7.8
difficulty	223	7.74
disruption	111	7.71
distress	101	7.52
concern	190	7.35
pain	126	7.24
chaos	82	7.22
accident	126	7.22
loss	190	7.14
controversy	81	7.12
pollution	88	7.04
havoc	64	7.01
cancer	93	7.01
stir	62	7.0
suffering	70	6.99
disease	141	6.95
explosion	72	6.93
embarrassment	63	6.87

subject	9100	6.2
negligence	55	7.34
virus	53	7.14
smoking	27	6.36
defect	29	6.32
bacterium	26	6.23
infection	32	6.09
factor	76	6.07
assault	28	6.05
pollution	31	6.04
recession	28	5.99
stress	28	5.88
accident	36	5.8
bomb	26	5.78
disease	50	5.74
fire	45	5.63
lack	37	5.6
organism	18	5.58
deficiency	16	5.57
fault	21	5.57
delay	20	5.55
damage	30	5.51
weather	22	5.39
explosion	15	5.27
drought	12	5.27
parasite	12	5.26

modifier	1971	1.4
reasonable	26	8.72
indirectly	16	7.77
possibly	27	7.67
thereby	26	7.66
mainly	32	7.51
inevitably	20	7.48
partly	22	7.47
probably	51	7.1
thus	36	7.01
recklessly	8	6.97
in part	10	6.94
undoubtedly	11	6.91
deliberately	14	6.81
certainly	26	6.77
intentionally	7	6.77
sometimes	31	6.69
often	72	6.67
directly	24	6.66
reportedly	9	6.64
usually	37	6.61
either	28	6.49
primarily	10	6.45
largely	18	6.39
allegedly	7	6.39
also	196	6.39

and/or	130	0.0
permit	18	6.09
contribute	5	4.21
use	6	0.22

part_intrans	10	0.0
by	9	6.71

unary rels		
np_VPto	2407	25.0
prep_Sing	158	11.3
np_pp	4585	5.1

pp_by-p	3374	16.4
negligence	46	8.14
defect	21	6.81
bacterium	17	6.62
virus	17	6.4
smoking	13	6.4
fault	19	6.15
lack	37	5.98
deficiency	10	5.95
shortage	12	5.75
blockage	6	5.71
breach	14	5.66
parasite	7	5.65
default	7	5.59
error	18	5.58
abnormality	7	5.58
theft	9	5.58
pollution	14	5.57
enteritis	5	5.53
build-up	6	5.51
fall	17	5.5
exposure	11	5.5
warming	6	5.46
drought	6	5.45
failure	24	5.38
recession	11	5.36

Specific papers using these ideas

Just a Sampling...

Hatzivassiloglou & McKeown 1997
Predicting the semantic orientation of adjectives

1. Build training set: label all adjectives with frequency > 20

   Test agreement with human annotators

Hatzivassiloglou & McKeown 1997

1. Build training set: label all adj. with frequency > 20; test agreement with human annotators
2. Extract all conjoined adjectives

Web Results **1** - **10** of about **762,000** for **"was very nice and"**.

The Homestay Experience - Cultural Kaleidoscope 2006
My host's home **was very nice and** comfortable. I got to try all types of food; Malaysian, Chinese, Indonesian and I loved it all. My host's parents were very **...**
www.gardenschool.edu.my/studentportal/aec/Kaleidoscope06/experience.asp - 10k -
Cached - Similar pages - Note this

PriceGrabber User Rating for Watch Your Budget - PriceGrabber.com
Reviews, Camera I purchased **was very nice and** a bargain. There was a problem with shipping, but was resolved quickly. Buy with confidence from this vendor. **...**
www.pricegrabber.com/rating_getreview.php/retid=5821 - Similar pages - Note this

Testimonials
"Everybody **was very nice and** service was as fast as they possibly could. **...** "Staff member who helped me **was very nice and** easy to talk to." **...**
www.sa.psu.edu/uhs/news/testimonials.cfm - 22k - Cached - Similar pages - Note this

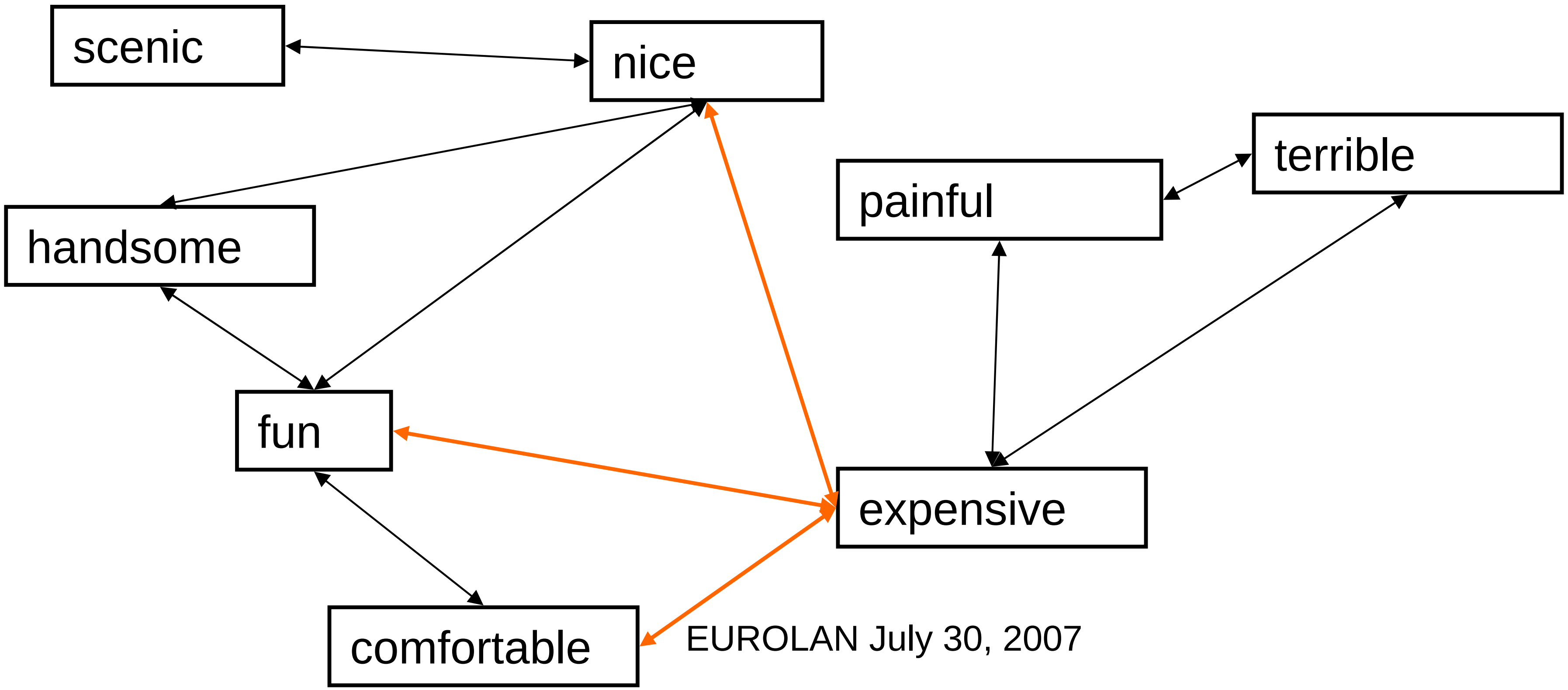
Naxos Villages - Naxos Town or Chora Reviews: Very nice and very ...
-Did you enjoy the trip to Naxos Town: Yes it **was very nice and** very scenic. -In order to get to the village were there enough signs in order to find it: It **...**

nice and comfortable

nice and scenic

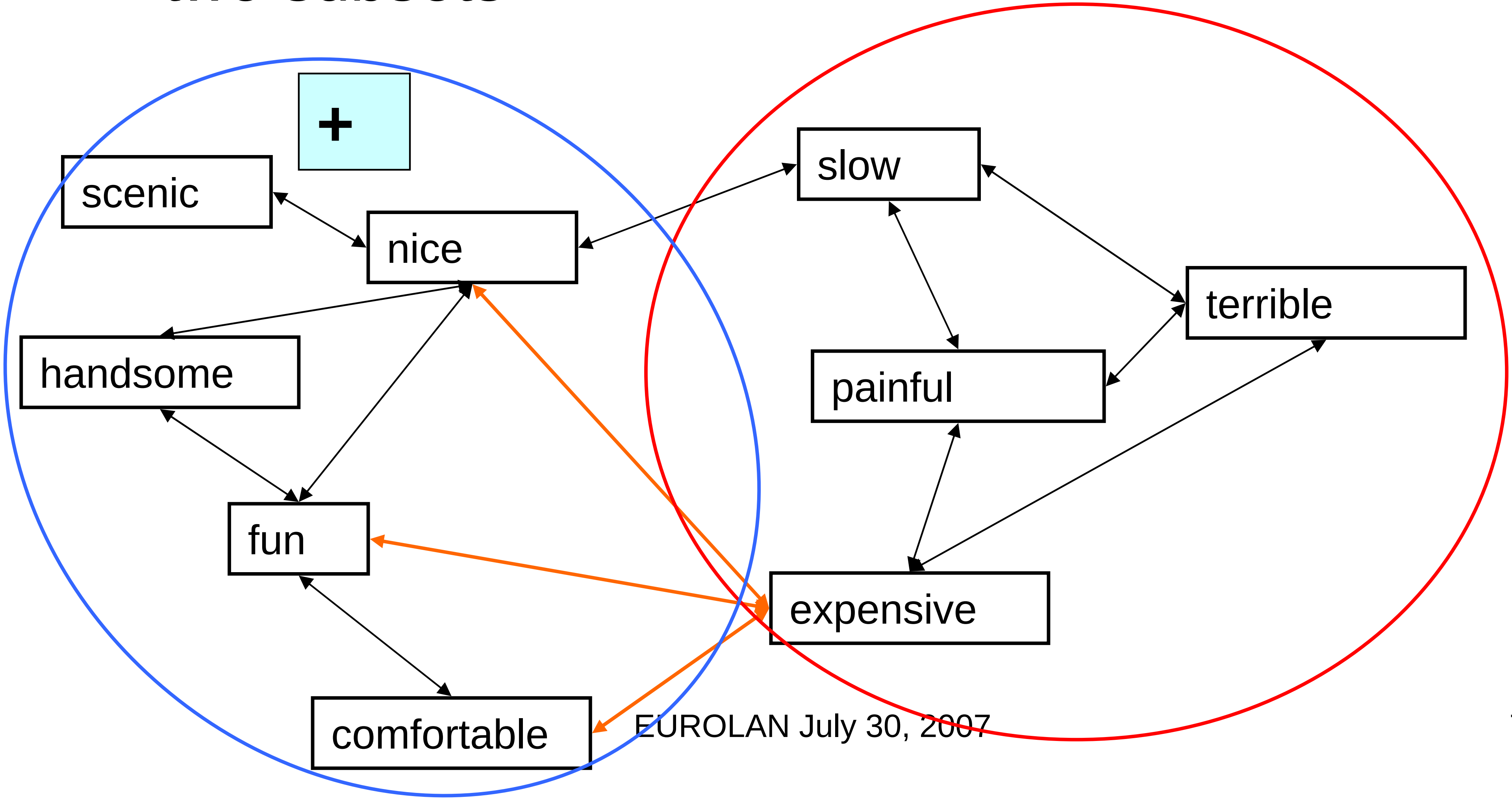
Hatzivassiloglou & McKeown 1997

3. A supervised learning algorithm builds a graph of adjectives linked by the same or different semantic orientation



Hatzivassiloglou & McKeown 1997

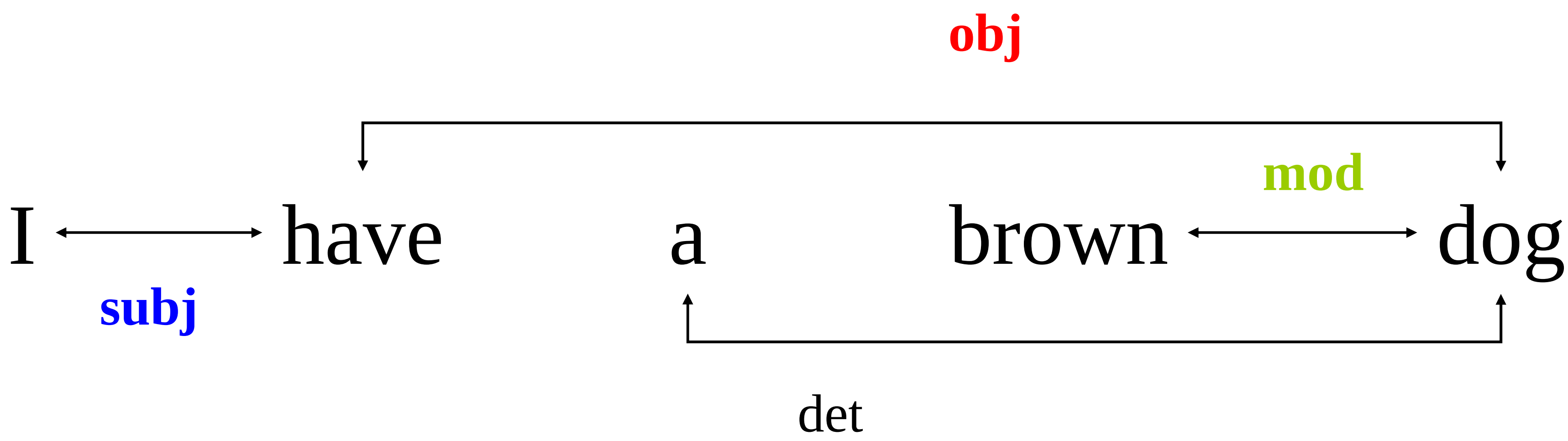
4. A clustering algorithm partitions the adjectives into two subsets



Wiebe 2000
Learning Subjective Adjectives From Corpora

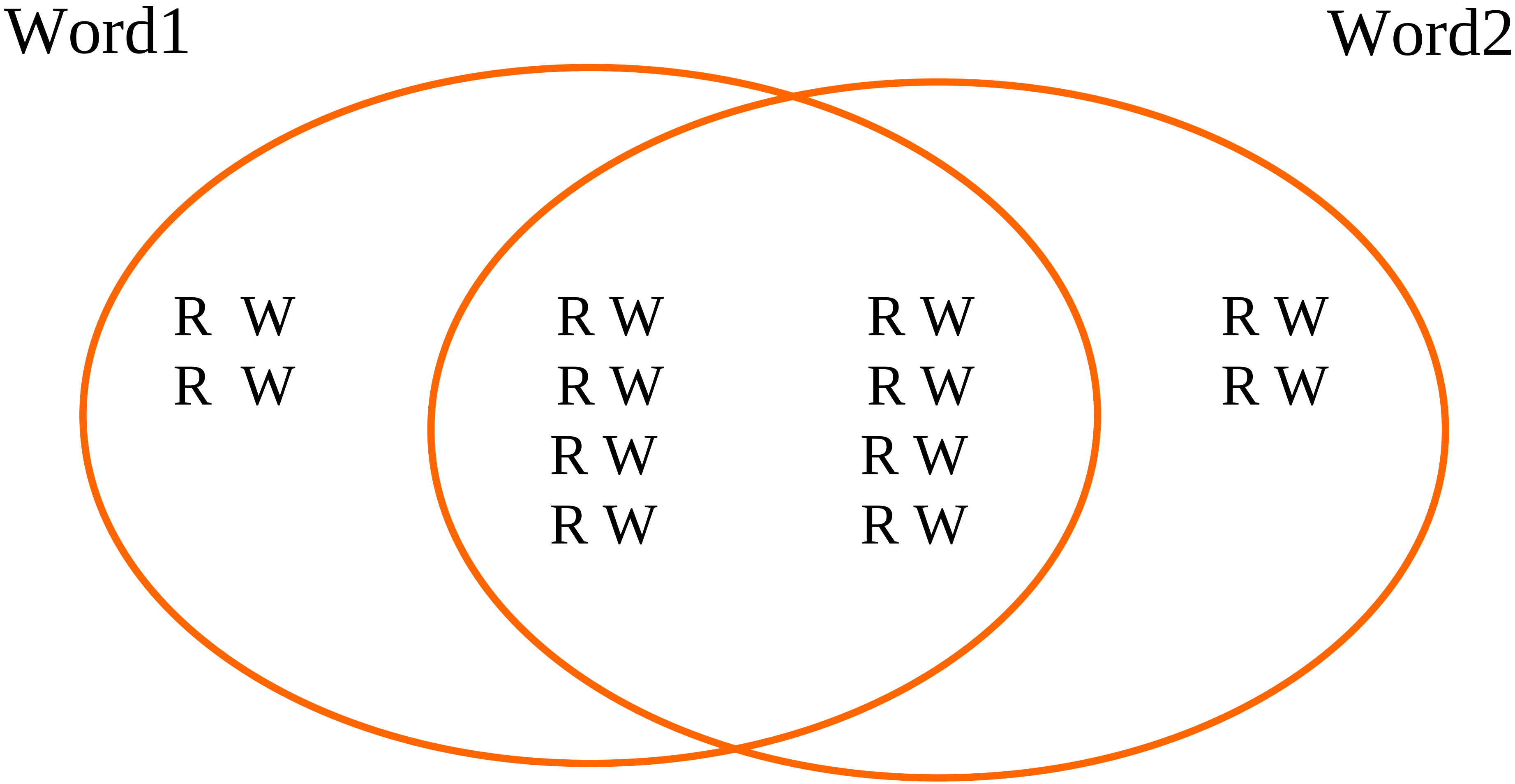
- Learning evaluation and opinion clues
  - Distributional similarity process
  - Small amount of annotated data, large amount of unannotated data
  - Refinement with lexical features
  - Improved results from both

Lin's (1998) Distributional Similarity



Word	R	W
I	subj	have
have	obj	dog
brown	mod	dog
• • •		

Lin's Distributional Similarity



Bizarre

strange similar scary unusual fascinating interesting curious tragic different contradictory peculiar silly sad absurd poignant crazy funny comic compelling odd

Bizarre

strange similar scary unusual fascinating interesting curious tragic different contradictory peculiar silly sad absurd poignant crazy funny comic compelling odd

Bizarre

strange similar scary unusual fascinating interesting curious tragic different contradictory peculiar silly sad absurd poignant crazy funny comic compelling odd

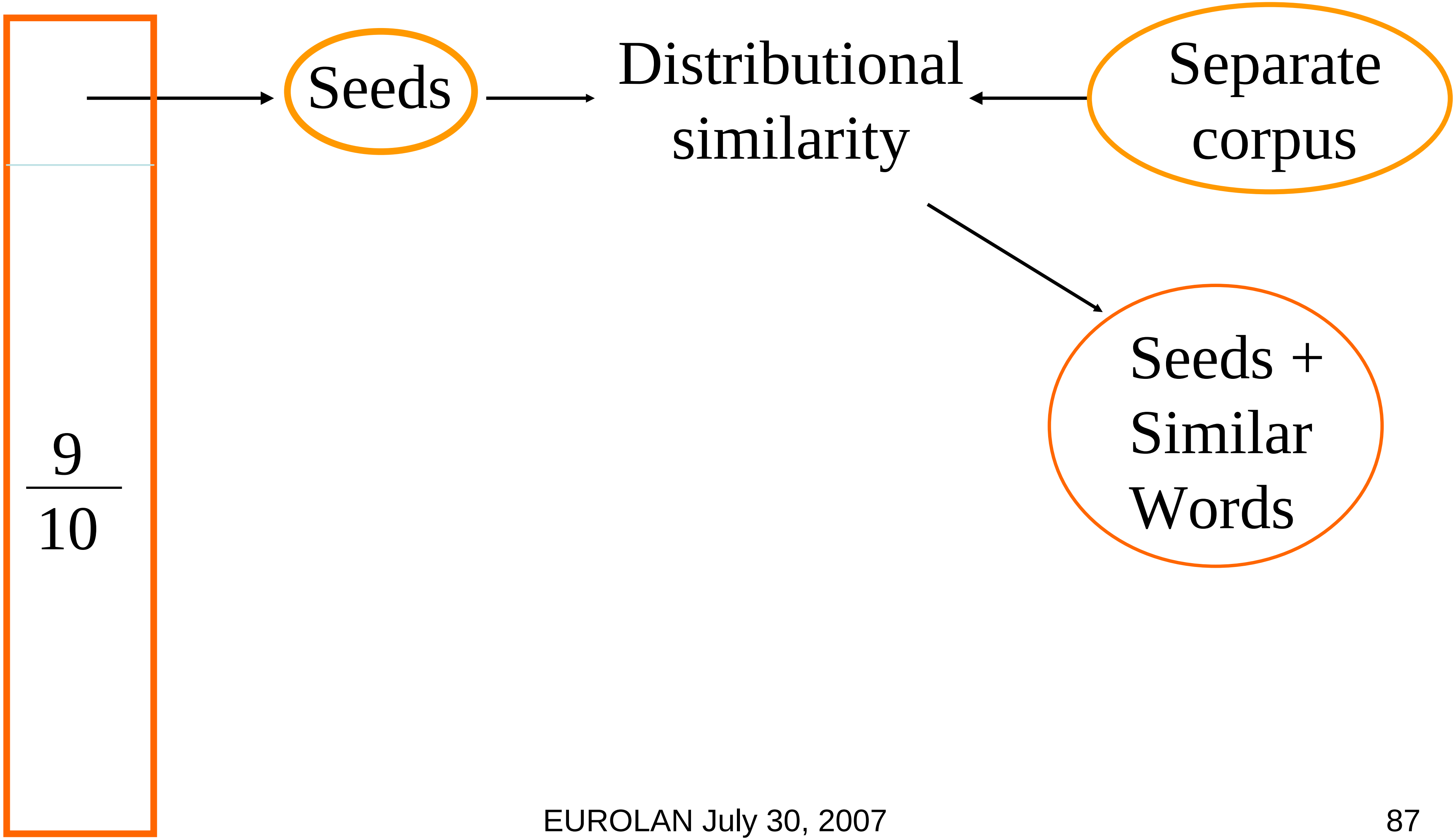
Bizarre

strange similar scary unusual fascinating interesting curious tragic different contradictory peculiar silly sad absurd poignant crazy funny comic compelling odd

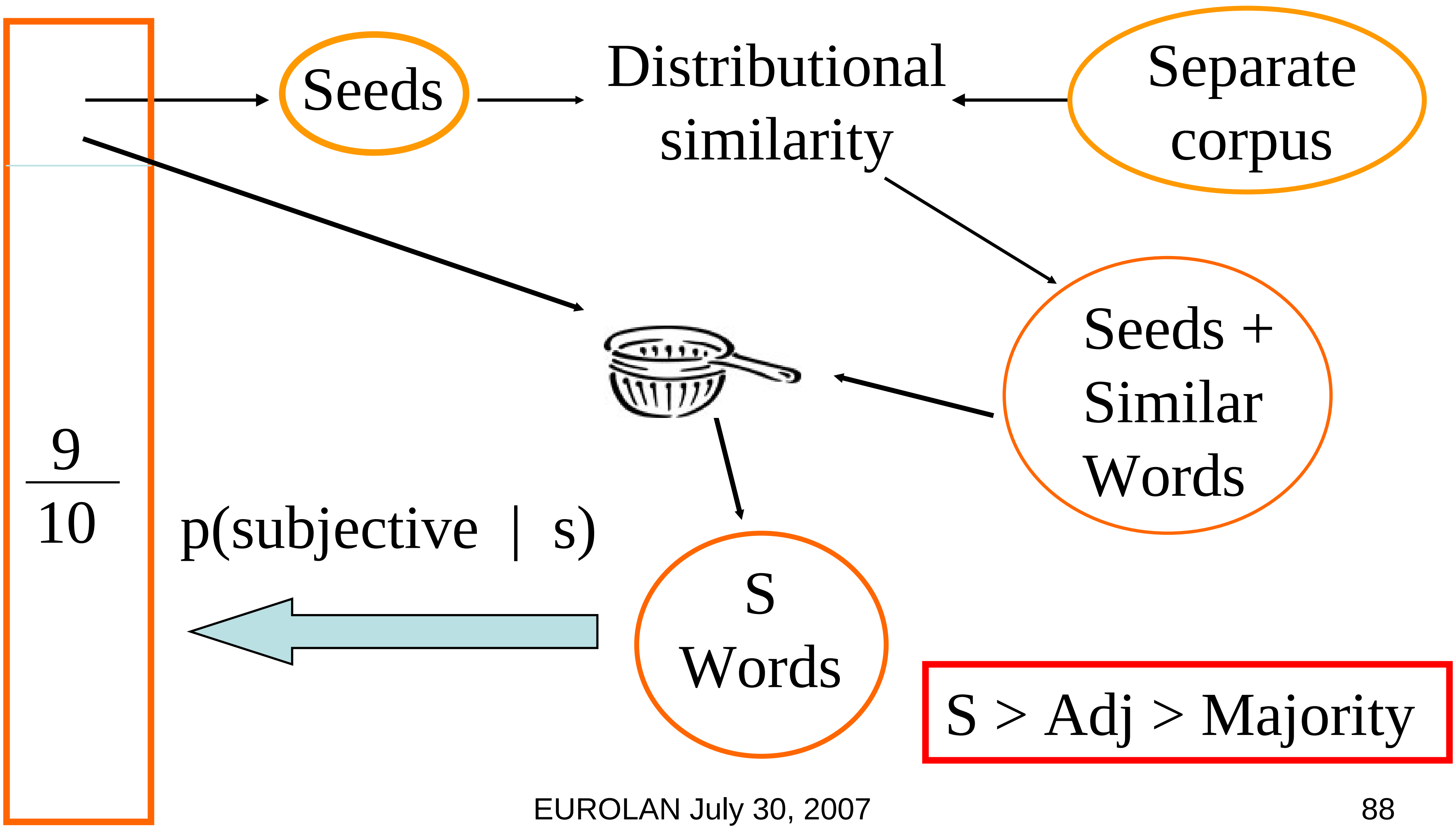
Experiments

$$\frac{9}{10}$$

Experiments



Experiments



Turney 2002; Turney & Littman 2003

Thumbs up or Thumbs down?
Unsupervised learning of semantic orientation from a hundred-billion-word corpus

- Determine the semantic orientation of each extracted phrase based on their association with seven positive and seven negative words

$$PMI(word_1, word_2) = \log_2 \left[\frac{p(word_1 \,\&\, word_2)}{p(word_1)\, p(word_2)} \right]$$

$$SO-PMI-IR(word) = \log_2 \left[\frac{hits(word\ NEAR\ p\_query)\, hits(n\_query)}{hits(word\ NEAR\ n\_query)\, hits(p\_query)} \right]$$

Turney 2002; Turney & Littman 2003

- Determine the semantic orientation of each extracted phrase based on their association with seven positive and seven negative words

$$PMI(word_1, word_2) = \log_2 \left[\frac{p(word_1 \,\&\, word_2)}{p(word_1)\, p(word_2)} \right]$$

$$SO-PMI-IR(word) = \log_2 \left[\frac{hits(word\ NEAR\ p\_query)hits(n\_query)}{hits(word\ NEAR\ n\_query)hits(p\_query)} \right]$$

Pang, Lee, Vaithyanathan 2002

- Movie review classification using Naïve Bayes, Maximum Entropy, SVM
  - Results do not reach levels achieved in topic categorization
- Various feature combinations (unigram, bigram, POS, text position)
  - Unigram presence works best
- Challenge:discourse structure

Riloff & Wiebe 2003
Learning extraction patterns for subjective expressions

- Observation: subjectivity comes in many (low-frequency) forms → better to have more data
- Boot-strapping produces cheap data

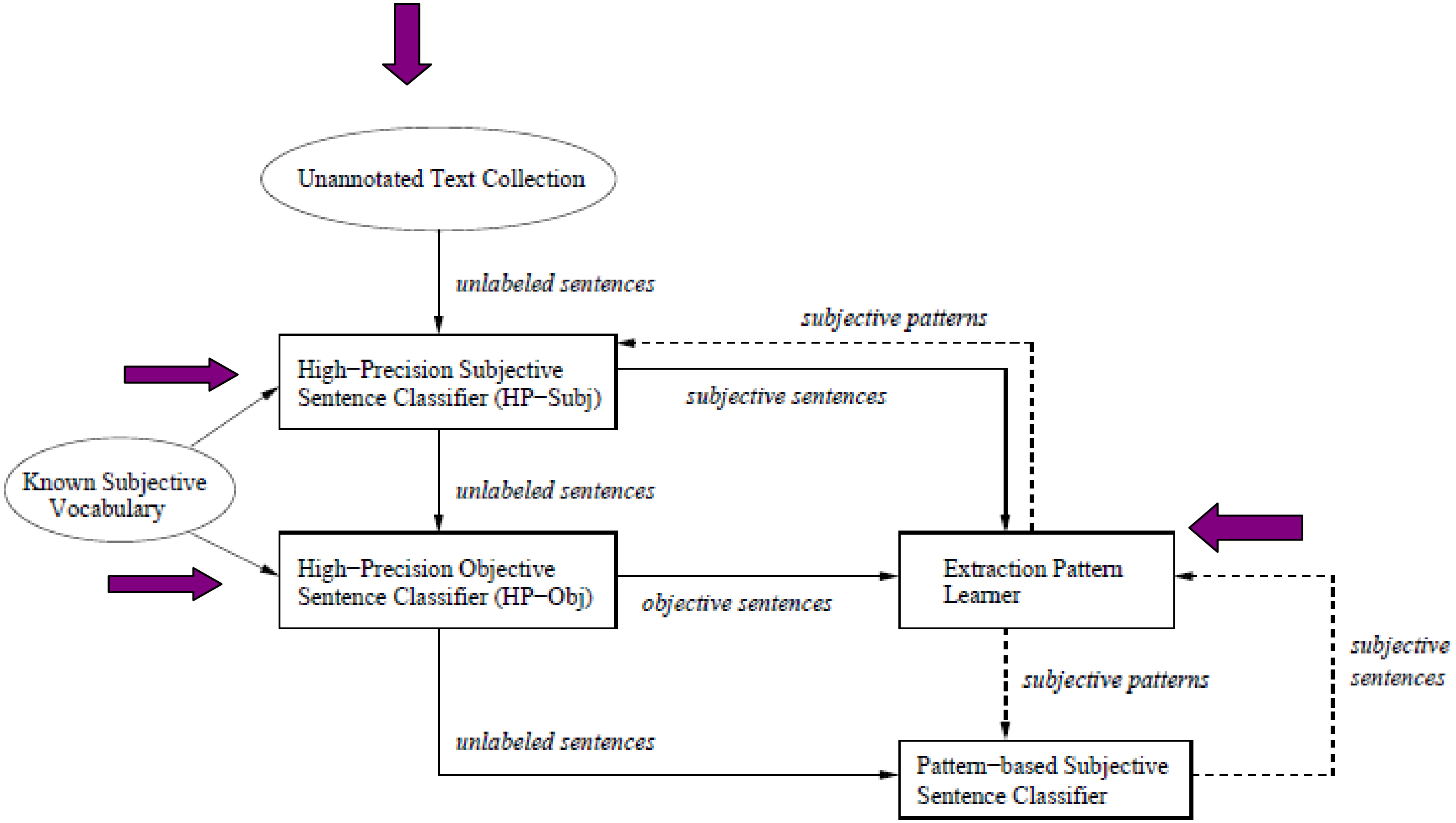


Figure 1: Bootstrapping Process

Riloff & Wiebe 2003

- Observation: subjectivity comes in many (low-frequency) forms → better to have more data
- Boot-strapping produces cheap data
- High-precision classifiers look for sentences that can be labeled subjective/objective with confidence
- Extraction pattern learner gathers **patterns biased towards subjective texts**
- Learned patterns are fed back into high precision classifiers

Subjective Expressions as IE Patterns

PATTERN	*FREQ*	*P(Subj | Pattern)*
<subj> asked	128	0.63
<subj> was asked	11	1.00
<subj> was expected	45	0.42
was expected from <np>	5	1.00
<subj> put	187	0.67
<subj> put end	10	0.90
<subj> talk	28	0.71
talk of <np>	10	0.90
<subj> is talk	5	1.00
<subj> is fact	38	1.00
fact is <dobj>	12	1.00

Yu & Hatzivassiloglou 2003

Toward answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences

- Classifying documents: naïve bayes, words as features
- Finding opinion sentences:
  - 2 similarity approaches
  - Naïve bayes (n-grams, POS, counts of polar words, counts of polar sequences, average orientation)
  - Multiple naïve bayes

Yu & Hatzivassiloglou 2003

- Tagging words and sentences:
  - modified log-likelihood ratio of collocation with pos, neg adjectives in seed sets
  - Adjectives, adverbs, and verbs provide best combination for tagging polarity of sentences

Yu & Hatzivassiloglou 2003

$$L(W_i, POS_j) = \log\left(\frac{\dfrac{Freq(W_i, POS_j, ADJ_p) + \varepsilon}{Freq(W_{all}, POS_j, ADJ_p)}}{\dfrac{Freq(W_i, POS_j, ADJ_n) + \varepsilon}{Freq(W_{all}, POS_j, ADJ_p)}}\right.$$

Kim & Hovy 2005
Automatic Detection of Opinion Bearing Words and Sentences

- WordNet-based method for collecting opinion-bearing adjectives and verbs
  - manually constructed strong seed set
  - manually labeled reference sets (opinion-bearing or not)
  - for synonyms/antonyms of seed set, calculate an opinion strength relative to reference sets
  - expand further with naïve bayes classifier

$$\arg\max P(c \mid w) = \arg\max P(c)P(w \mid c)$$
$$= \arg\max P(c)P(syn_1\ syn_2\ ..syn_n \mid c)$$
$$= \arg\max P(c)\prod P(f_k \mid c)^{count(f_k, synset(w))}$$

Kim & Hovy 2005

- Corpus-based method (WSJ)
- Calculate bias of words for particular text genre (Editorials and Letter to editor)

$$\mathrm{EditorialProb(w)} = \frac{\#\, w \text{ in editorial documents}}{\text{total words in editorial documents}}$$

$$Score(w) = \frac{Editorial\ \Pr ob(w)}{Noneditorial\ \Pr ob(w)}$$

Esuli & Sebastiani 2005

Determining the semantic orientation of terms through gloss classification

- use seed sets (positive and negative)
- use lexical relations like synonymy and antonymy to extend the seed sets
  - brilliant->brainy->intelligent->smart->…
  - brilliant->unintelligent->stupid, brainless->…
- extend sets iteratively

Esuli & Sebastiani 2005

- use final sets as gold standard to train a classifier, which uses all or part of the glosses in some format as features
- the trained classifier can then be used to label any term that has a gloss with sentiment

Adjective

- S: (adj) awful, dire, direful, dread, dreaded, dreadful, fearful, fearsome, frightening, **horrendous**, horrific, terrible (causing fear or dread or terror) *"the awful war"; "an awful risk"; "dire news"; "a career or vengeance so direful that London was shocked"; "the dread presence of the headmaster"; "polio is no longer the dreaded disease it once was"; "a dreadful storm"; "a fearful howling"; "horrendous explosions shook the city"; "a terrible curse"*

w(awful)	**…**	**w(dire)**	**w(direful)**	**…**	**w(dread)**	**W(dreaded)**	**…**	**…**

Esuli & Sebastiani 2006

Determining Term Subjectivity and Term Orientation for Opinion Mining

- Uses best system of 2005 paper
- Additional goal of distinguishing neutral from positive/negative
- Multiple variations on learning approach, learner, training set, feature selection
- The new problem is harder! Their best accuracy is 66% (83% in 2005 paper)

Suzuki et al. 2006

Application of semi-supervised learning to evaluative expression classification

- Automatically extract and filter "evaluative expressions": The storage capacity of this HDD is high.
- Classifies these as pos, neg, or neutral
- Use bootstrapping to be able to train an evaluative expression classifier based on a larger collection of unlabeled data.
- Learn contexts that contain evaluative expressions
  - **I am really happy because** [the storage capacity is high]
  - **Unfortunately**, [the laptop was too expensive].

Suzuki et al. 2006

Evaluation

Attribute

- Automatically extract and filter "evaluative expressions": The storage capacity of this HDD is high.

Subject

- Classifies these as pos, neg, or neutral
- Use bootstrapping to be able to train an evaluative expression classifier based on a larger collection of unlabeled data.
- Learn contexts that contain evaluative expressions
  - I am really happy because [the storage capacity is high]
  - Unfortunately, [the laptop was too expensive].

Suzuki et al. 2006

- Comparison of semi-supervised methods:
  - Nigam et al.'s (2000) Naive Baiyes + EM method
  - Naive Bayes + EM + SVM (SVM combined with Naive Bayes + EM using Fisher kernel)
- And supervised methods:
  - Naive Bayes
  - SVM

Suzuki et al. 2006

- Features:
  'Phew, [the noise] of [this HDD] is **annoyingly high** :-('.

  – Candidate evaluative expression
  – "Exclamation words" detected by POS tagger
  – Emoticons and their emotional categories
  – Words modifying words in the candidate evaluation expression
  – Words modified by words in the candidate evaluative word

Suzuki et al. 2006

- Both Naive Bayes + EM, and Naive Bayes + EM + SVM work better than Naive Bayes and SVM.
- Results show that Naive Bayes + EM **boosted accuracy** regardless of size of labeled data
- Using more unlabeled data appeared to give better results.
- **Qualitative analysis** of the impact of the semi-supervised approaches by looking at the top 100 features that had the highest probability P(feature—positive) before and after EM:
  - more contextual features like exclamations, the happy emoticons, a negation + ’but’, ‘therefore’ + ’interesting’, and ‘therefore‘ + ’comfortable.’

Surely

… we've thought of everything by now?

Word senses

Senses

Adjective

- (11) S: (adj) **brilliant**, superb (of surpassing excellence) *"a brilliant performance"; "a superb actor"*
- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*
- (2) S: (adj) **brilliant**, glorious, magnificent, splendid (characterized by grandeur) *"the brilliant court life at Versailles"; "a glorious work of art"; "magnificent cathedrals"; "the splendid coronation ceremony"*
- (2) S: (adj) bright, **brilliant**, vivid (having striking color) *"bright dress"; "brilliant tapestries"; "a bird with vivid plumage"*
- (2) S: (adj) **brilliant** (full of light; shining intensely) *"a brilliant star"; "brilliant chandeliers"*
- (1) S: (adj) bright, **brilliant** (clear and sharp and ringing) *"the bright sound of the trumpet section"; "the brilliant sound of the trumpets"*

Senses

Adjective

- (11) S: (adj) **brilliant**, superb (of surpassing excellence) *"a brilliant performance"; "a superb actor"*
- (7) S: (adj) brainy, **brilliant**, smart as a whip (having or marked by unusual and impressive intelligence) *"some men dislike brainy women"; "a brilliant mind"; "a brilliant solution to the problem"*
  - *similar to*
    - S: (adj) intelligent (having the capacity for thought and reason especially to a high degree) *"is there intelligent life in the universe?"; "an intelligent question"*
  - *derivationally related form*
    - W: (n) brilliancy [Related to: brilliant] (a quality that outshines the usual)
    - W: (n) brilliance [Related to: brilliant] (unusual mental ability)
  - ***antonym***
    - W: (adj) unintelligent [Indirect via intelligent] (lacking intelligence) *"a dull job with lazy and unintelligent co-workers"*
- (2) S: (adj) **brilliant**, glorious, magnificent, splendid (characterized by grandeur) *"the brilliant court life at Versailles"; "a glorious work of art"; "magnificent cathedrals"; "the splendid coronation ceremony"*
- (2) S: (adj) bright, **brilliant**, vivid (having striking color) *"bright dress"; "brilliant tapestries"; "a bird with vivid plumage"*
- (2) S: (adj) **brilliant** (full of light; shining intensely) *"a brilliant star"; "brilliant chandeliers"*
- (1) S: (adj) bright, **brilliant** (clear and sharp and ringing) *"the bright sound of the trumpet section"; "the brilliant sound of the trumpets"*

Non-subjective senses of brilliant

1. Method for identifying **brilliant material** in paint - US Patent 7035464
2. Halley shines in a **brilliant light**.
3. In a classic *pasodoble*, an opening section in the minor mode features a **brilliant trumpet melody**, while the second section in the relative major begins with the violins.

Andreevskaia and Bergler 2006
Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses

- Using wordnet relations (synonymy, antonymy and hyponymy) and glosses
- Classify as positive, negative, or neutral
- Step algorithm with known seeds:
  - First expand with relations
  - Next expand via glosses
  - Filter out wrong POS and multiply assigned
- Evaluate against General inquirer (which contains words, not word senses)

Andreevskaia and Bergler 2006

- Partitioned the entire Hatzivassiloglou & McKeown list into 58 non-intersecting seed lists of adjectives
- Performance of the system exhibits substantial variability depending on the composition of the seed list, with accuracy ranging from 47.6% to 87.5% percent (Mean = 71.2%, Standard Deviation (St.Dev) = 11.0%).
- The 58 runs were then collapsed into a single set of unique words.
- Adjectives identified by STEP in multiple runs were counted as one entry in the combined list. the collapsing procedure resulted in lower-accuracy (66.5% - when GI-H4 neutrals were included) but a much larger list of adjectives marked as positive (n = 3,908) or negative (n = 3,905).
- The 22, 141 WordNet adjectives not found in any STEP run were deemed **neutral** (n = 14, 328).
- System's **66.5% accuracy** on the collapsed runs is comparable to the accuracy reported in the literature for other systems run on large corpora (Turney and Littman, 2002; Hatzivassilglou and McKeown 1997).

Andreevskaia and Bergler 2006

- Disagreements between human labelers as a sign of fuzzy category structure
  - HM and General Inquirer have 78.7% tag agreement for shared adjectives
- Find way to measure the degree of centrality of words to the category of sentiment
- Net overlap scores correlate with human agreement

Outline

- Corpus Annotation
- Pure NLP
  - Lexicon development
  - **Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis**
- Applications
  - Product review mining

Wilson, Wiebe, Hoffmann 2005

Recognizing Contextual Polarity in Phrase-level Sentiment Analysis

Prior Polarity versus Contextual Polarity

- Most approaches use a lexicon of positive and negative words

  Prior polarity: out of context, positive or negative

  beautiful → positive

  horrid → negative

- A word may appear in a phrase that expresses a different polarity in context

"Cheers to Timothy Whitfield for the **wonderfully horrid** visuals."

Contextual polarity

Example

Philip Clap, President of the National Environment Trust, sums up well the general thrust of the reaction of environmental movements: there is no reason at all to believe that the polluters are suddenly going to become reasonable.

Example

Philip Clap, President of the National Environment **Trust**, sums up **well** the general thrust of the reaction of environmental movements: there is no **reason** at all to believe that the **polluters** are suddenly going to become **reasonable**.

Example

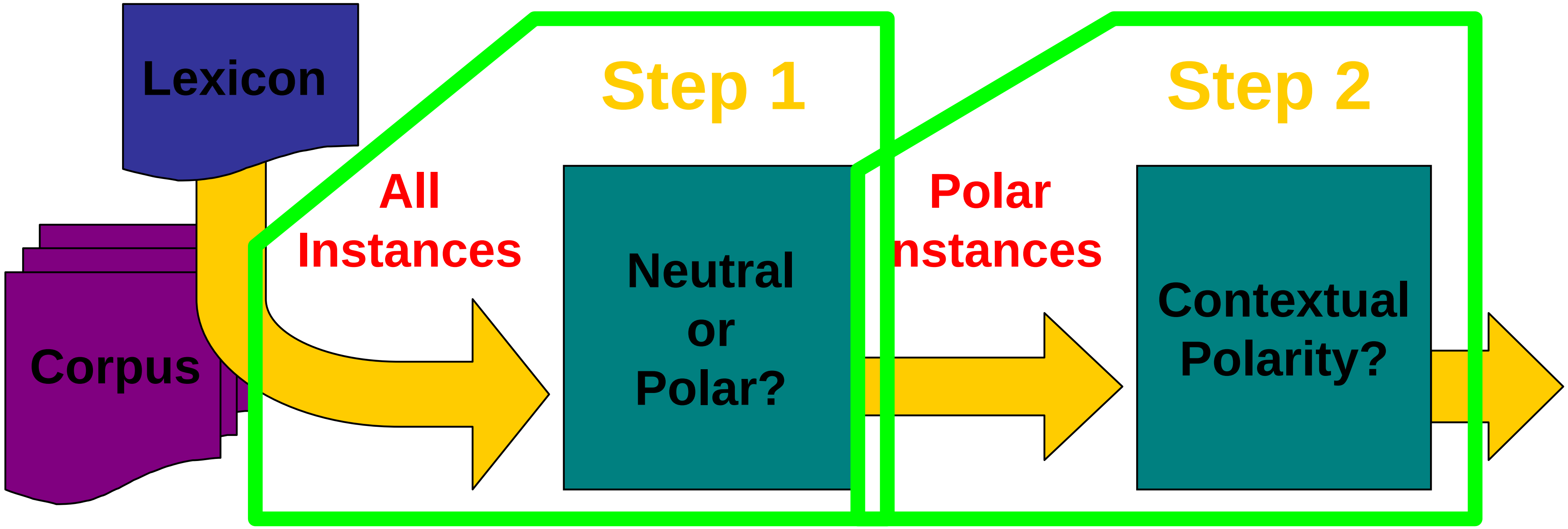
Philip Clap, President of the National Environment ~~**Trust**~~, sums up **well** the general thrust of the reaction of environmental movements: there is no **reason** at all to believe that the ~~**polluters**~~ are suddenly going to become **reasonable**.

prior polarity ← **Contextual polarity**

Goal of This Work

- Automatically distinguish prior and contextual polarity

Approach



- Use machine learning and variety of features
- Achieve significant results for a large subset of sentiment expressions

Manual Annotations

Subjective expressions of the MPQA corpus annotated with contextual polarity

Annotation Scheme

- Mark polarity of subjective expressions as **positive**, negative, both, or **neutral**

positive

African observers **generally approved** of his victory while Western governments **denounced** it.

negative

Besides, politicians refer to **good and evil** …

both

Jerome says the hospital **feels** no different than a hospital in the states.

neutral

Annotation Scheme

- Judge the contextual polarity of sentiment ***ultimately*** being conveyed

They have not succeeded, and will never succeed, in breaking the will of this valiant people.

Annotation Scheme

- Judge the contextual polarity of sentiment ***ultimately*** being conveyed

They have not **succeeded**, and will never succeed, in breaking the will of this valiant people.

Annotation Scheme

- Judge the contextual polarity of sentiment ***ultimately*** being conveyed

They have **not succeeded**, and will never succeed, in breaking the will of this valiant people.

Annotation Scheme

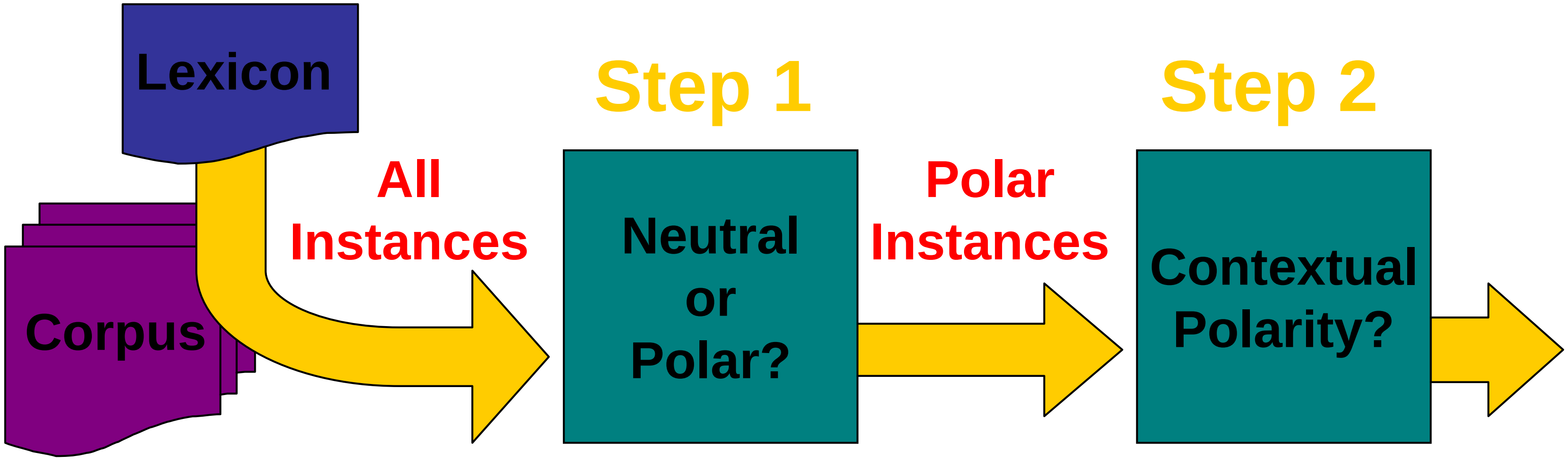
- Judge the contextual polarity of sentiment ***ultimately*** being conveyed

They **have not succeeded, and will never succeed**, in breaking the will of this valiant people.

Prior-Polarity Subjectivity Lexicon

- Over 8,000 words from a variety of sources
  - Both manually and automatically identified
  - Positive/negative words from General Inquirer and Hatzivassiloglou and McKeown (1997)
- All words in lexicon tagged with:
  - **Prior polarity**: positive, negative, **both**, **neutral**
  - **Reliability**: strongly subjective (strongsubj),
weakly subjective (weaksubj)

Experiments



Both Steps:

- BoosTexter AdaBoost.HM 5000 rounds boosting
- 10-fold cross validation

➢ Give each instance its own label

Definition of Gold Standard

Given an instance ***inst*** from the lexicon:

- if ***inst*** not in a subjective expression:
  - goldclass(***inst***) = neutral
- else if ***inst*** in at least one positive and one negative subjective expression:
  - goldclass(***inst***) = both
- else if ***inst*** in a mixture of negative and neutral:
  - goldclass(***inst***) = negative
- else if ***inst*** in a mixture of positive and neutral:
  - goldclass(***inst***) = positive
- else: goldclass(***inst***) = contextual polarity of subjective expression

Features

- Many inspired by Polanyi & Zaenen (2004): *Contextual Valence Shifters*

  Example: **little** threat

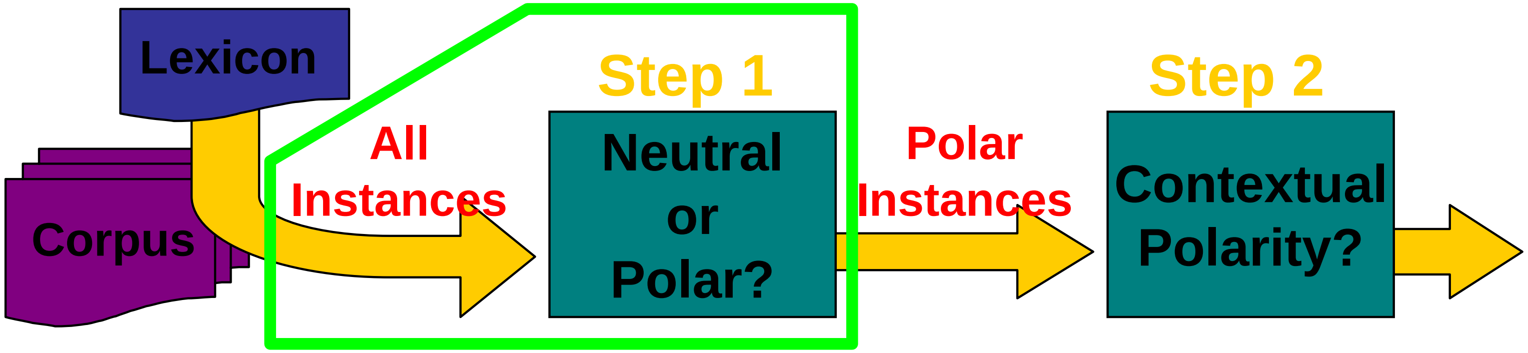
  little truth

- Others capture dependency relationships between words

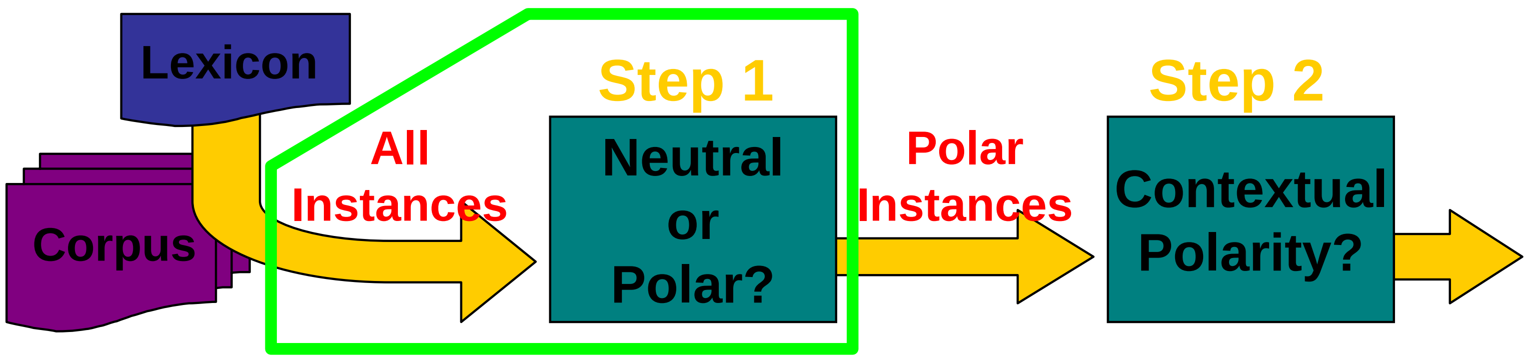
  Example: **pos**

  wonderfully horrid

  mod

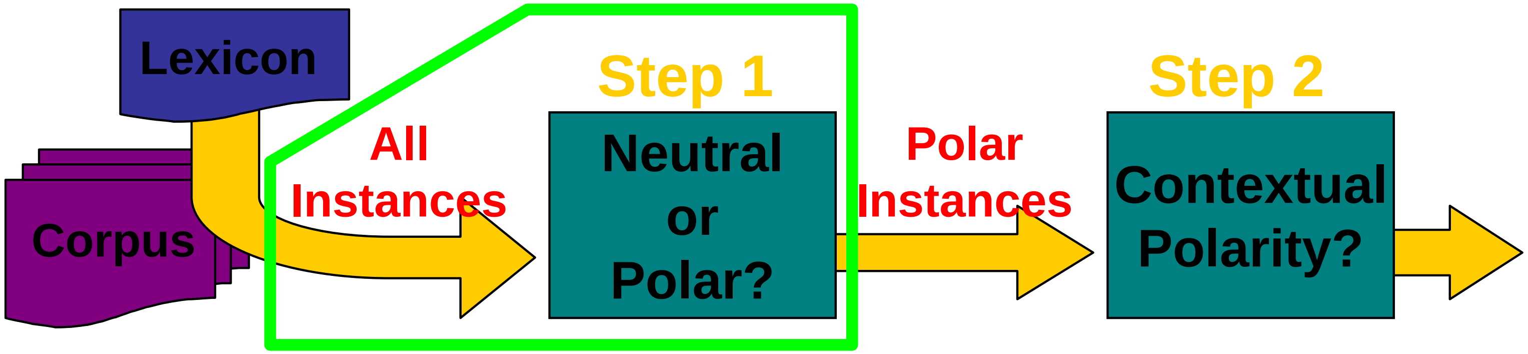


1. Word features
2. Modification features
3. Structure features
4. Sentence features
5. Document feature

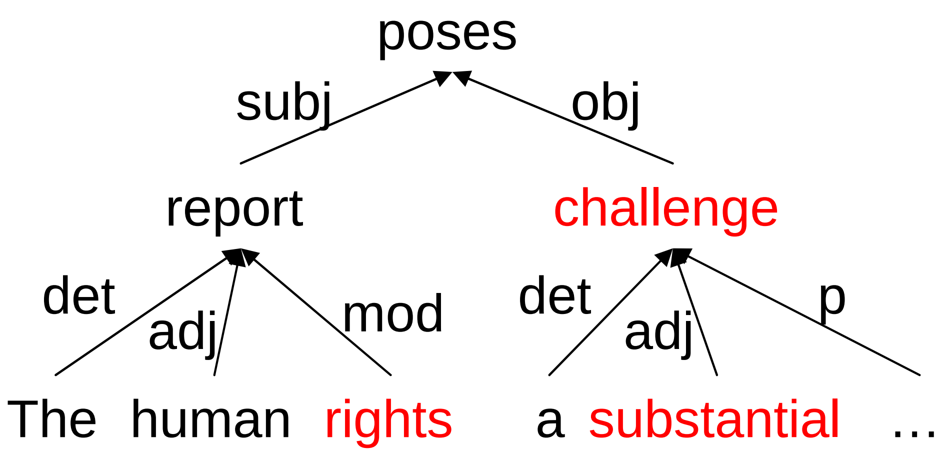


1. **Word features**
2. Modification features
3. Structure features
4. Sentence features
5. Document feature

- Word token
  terrifies
- Word part-of-speech
  VB
- Context
  that terrifies me
- Prior Polarity
  negative
- Reliability
  strongsubj



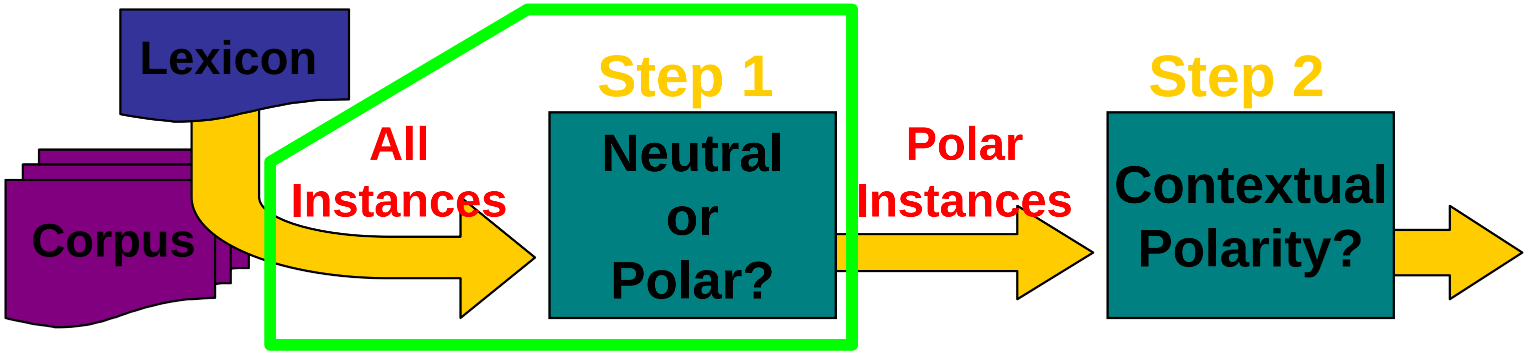
1. Word features
2. **Modification features**
3. Structure features
4. Sentence features
5. Document feature



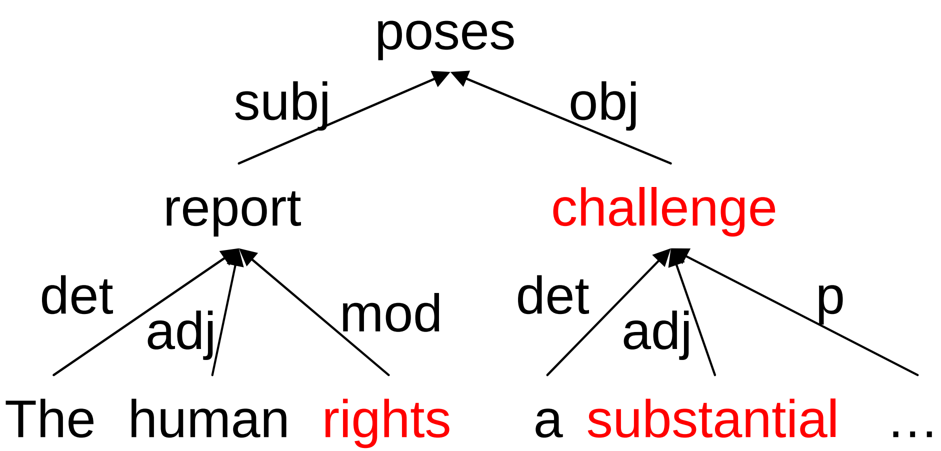
Binary features:

- Preceded by
  - adjective
  - adverb (other than not)
  - intensifier
- Self intensifier
- Modifies
  - strongsubj clue
  - weaksubj clue
- Modified by
  - strongsubj clue
  - weaksubj clue

Dependency Parse Tree



1. Word features
2. Modification features
3. **Structure features**
4. Sentence features
5. Document feature



Binary features:

- In subject

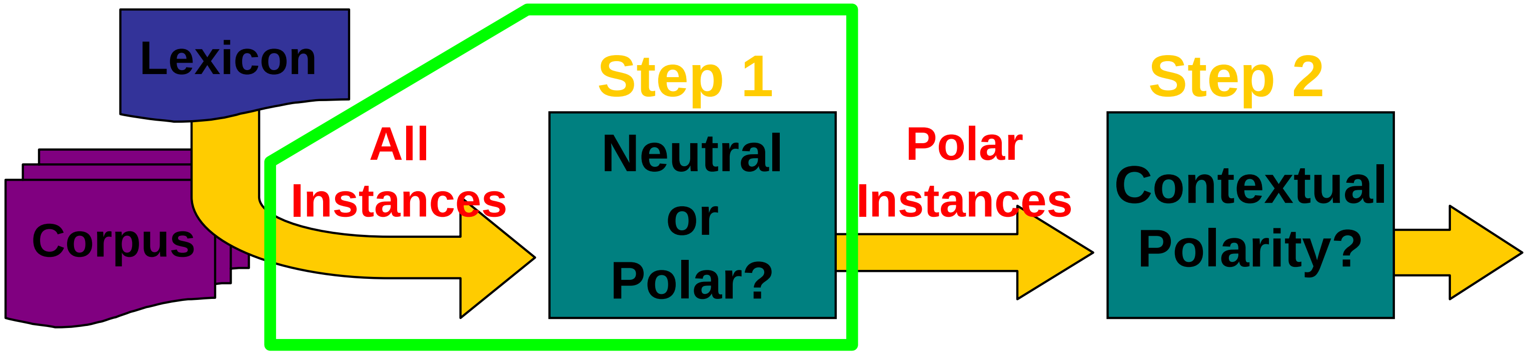
*[The human **rights** report] poses*

- In copular

  *I am **confident***

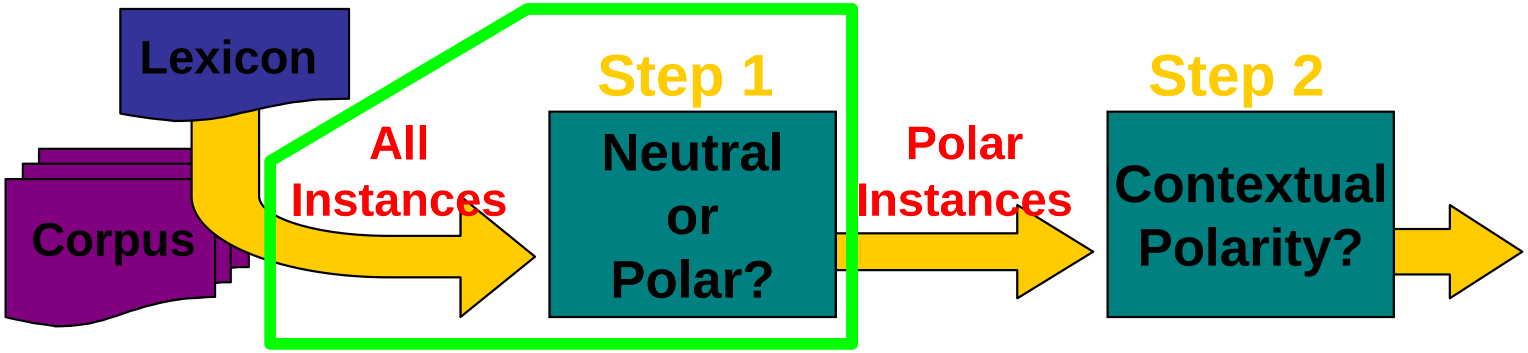
- In passive voice

  *must be **regarded***



1. Word features
2. Modification features
3. Structure features
4. **Sentence features**
5. Document feature

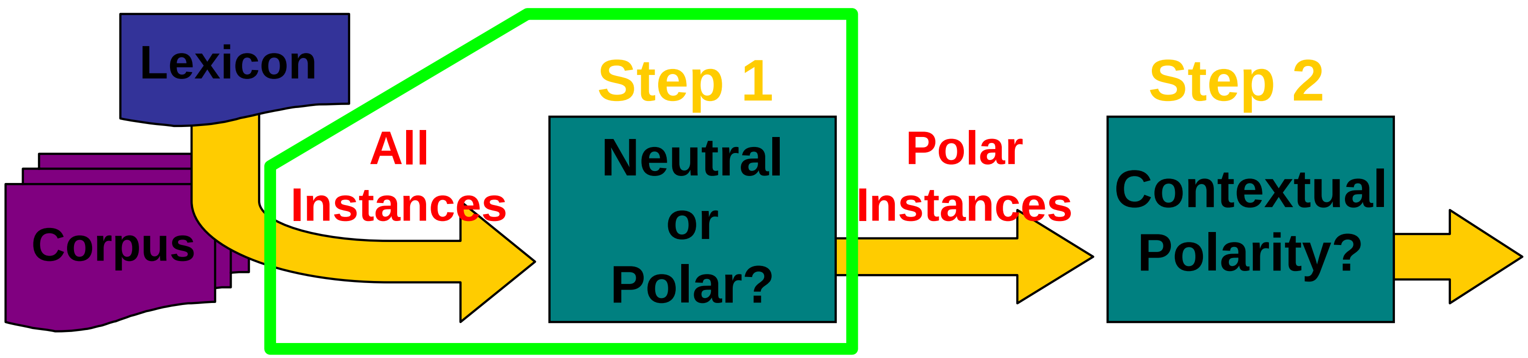
- **Count of strongsubj clues in** previous, current, next sentence
- **Count of weaksubj clues in** previous, current, next sentence
- **Counts of various parts of speech**



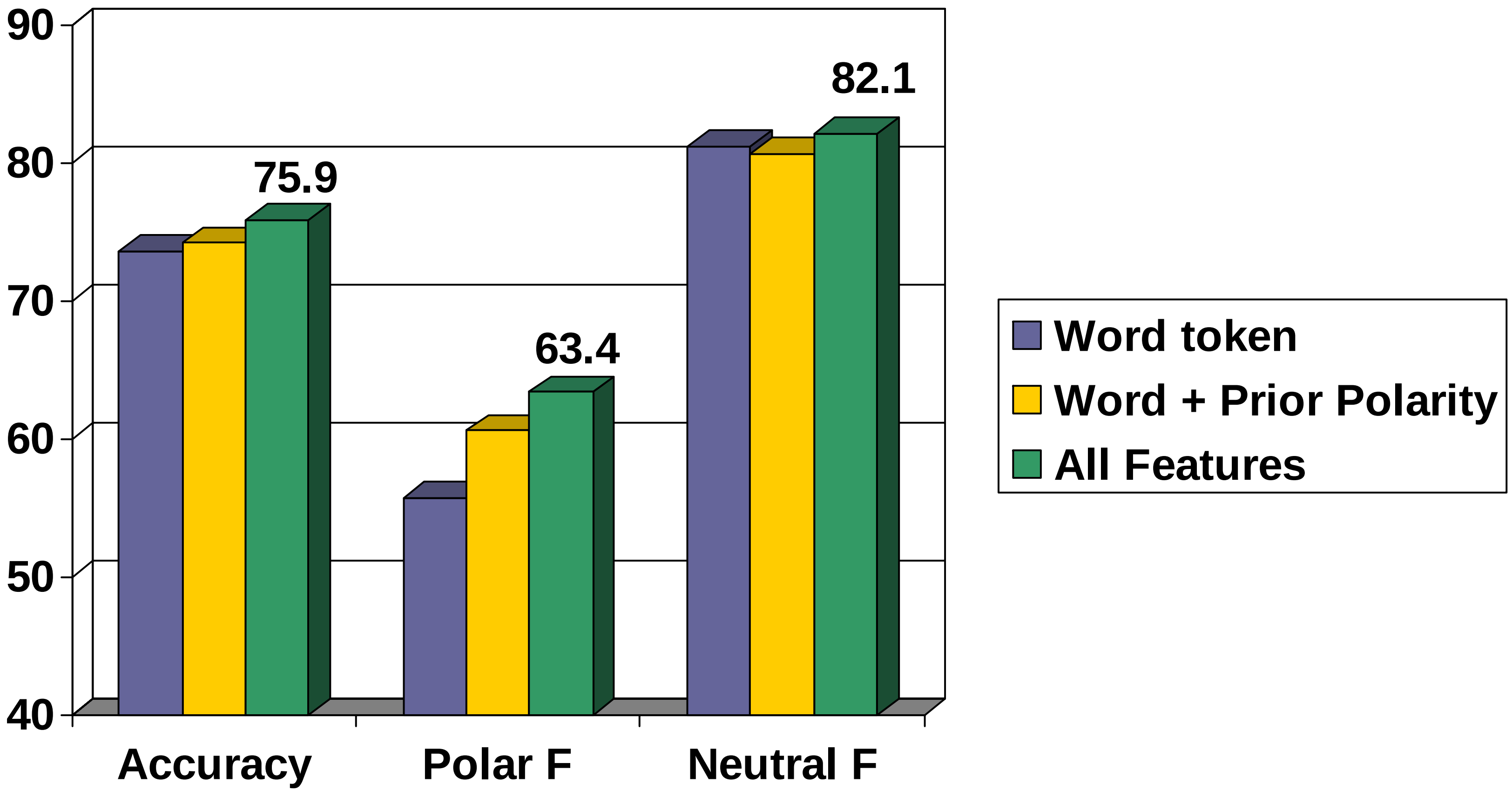
1. Word features
2. Modification features
3. Structure features
4. Sentence features
5. **Document feature**

- **Document topic (15)**
  – economics
  – health
  ⋮
  – Kyoto protocol
  – presidential election in Zimbabwe

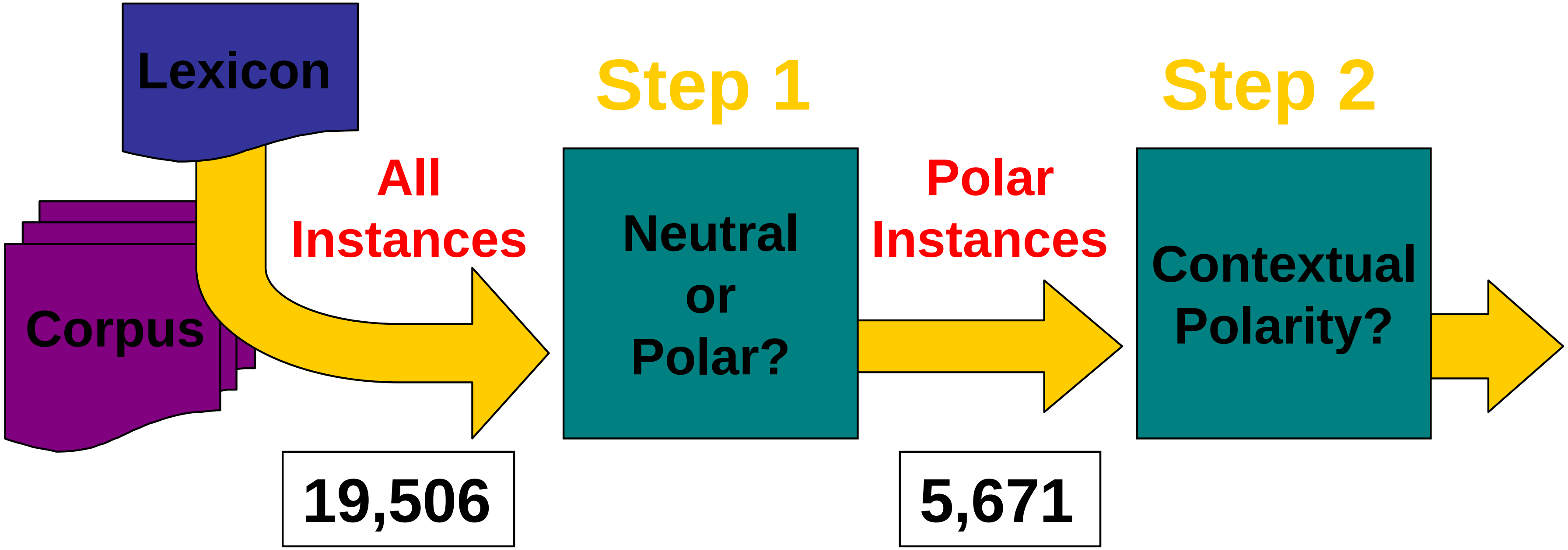
Example: The disease can be contracted if a person is bitten by a **certain** tick or if a person comes into contact with the blood of a congo **fever** **sufferer**.


Lexicon
Corpus
All Instances
Step 1
Neutral or Polar?
Polar Instances
Step 2
Contextual Polarity?

Results 1a

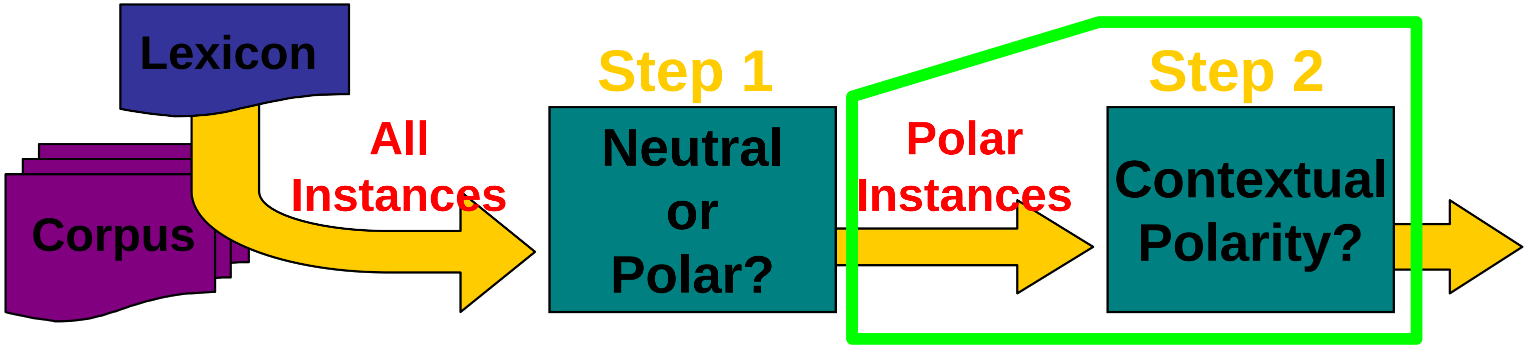

90
80
70
60
50
40
75.9
63.4
82.1
Accuracy
Polar F
Neutral F
Word token
Word + Prior Polarity
All Features

Step 2: Polarity Classification

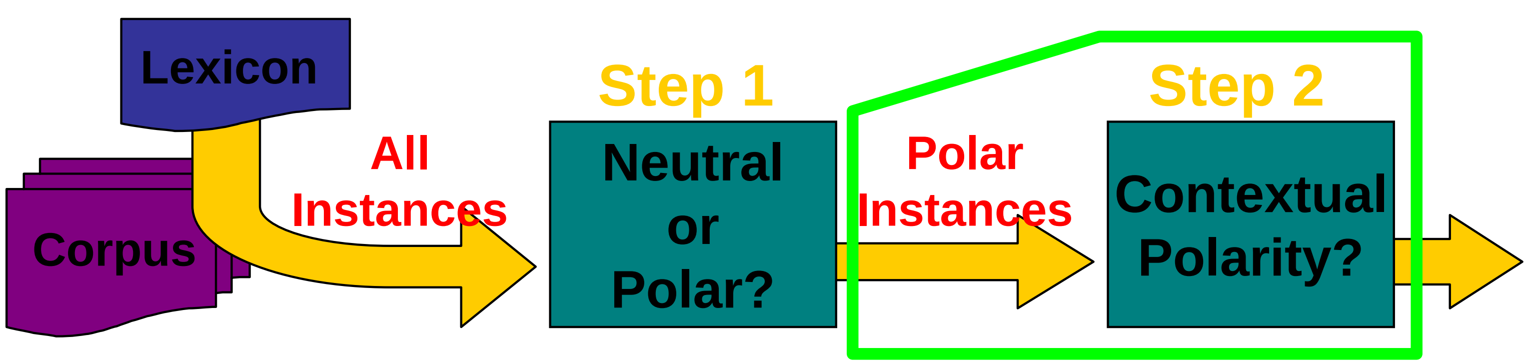


Classes

– positive, negative, both, neutral



- Word token
- Word prior polarity
- Negated
- Negated subject
- Modifies polarity
- Modified by polarity
- Conjunction polarity
- General polarity shifter
- Negative polarity shifter
- Positive polarity shifter



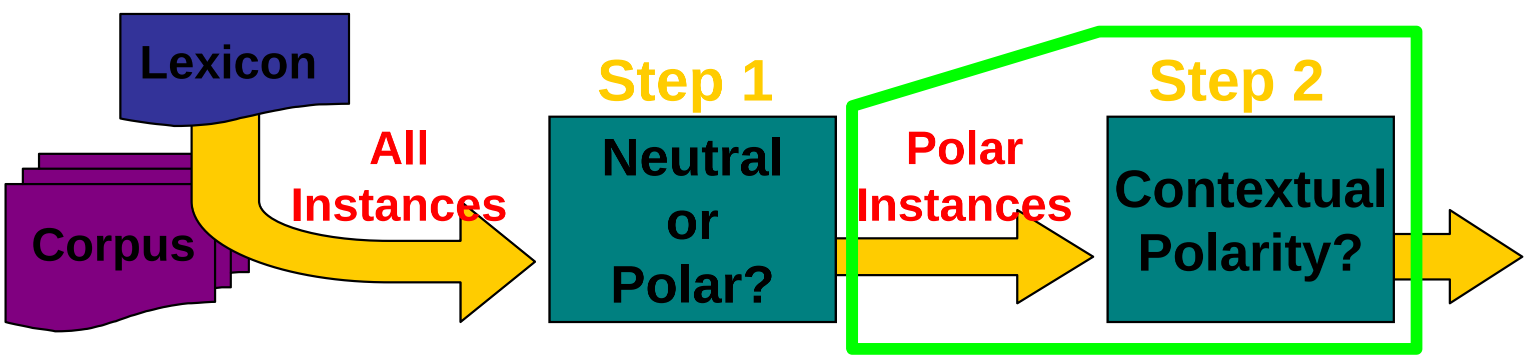
- **Word token**
- **Word prior polarity**
- Negated
- Negated subject
- Modifies polarity
- Modified by polarity
- Conjunction polarity
- General polarity shifter
- Negative polarity shifter
- Positive polarity shifter

Word token

terrifies

Word prior polarity

negative



- Word token
- Word prior polarity
- **Negated**
- **Negated subject**
- Modifies polarity
- Modified by polarity
- Conjunction polarity
- General polarity shifter
- Negative polarity shifter
- Positive polarity shifter

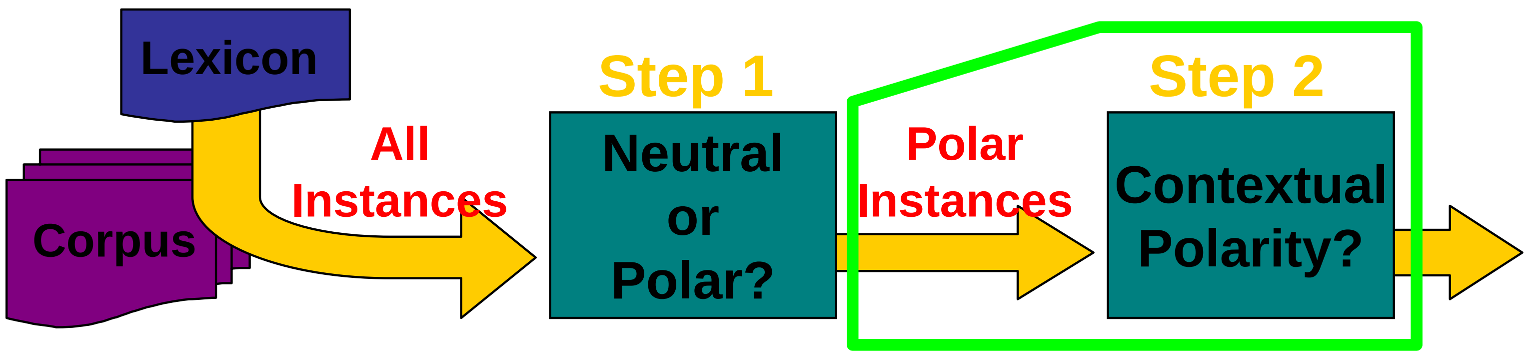
Binary features:

- **Negated**

  For example:
  - ***not*** *good*
  - ***does not*** *look very good*
  - ❖ ***not only*** *good but amazing*

- **Negated subject**

  No politically prudent Israeli *could support either of them.*



- Word token
- Word prior polarity
- Negated
- Negated subject
- **Modifies polarity**
- **Modified by polarity**
- Conjunction polarity
- General polarity shifter
- Negative polarity shifter
- Positive polarity shifter

- **Modifies polarity**

  5 values: positive, negative, neutral, both, not mod

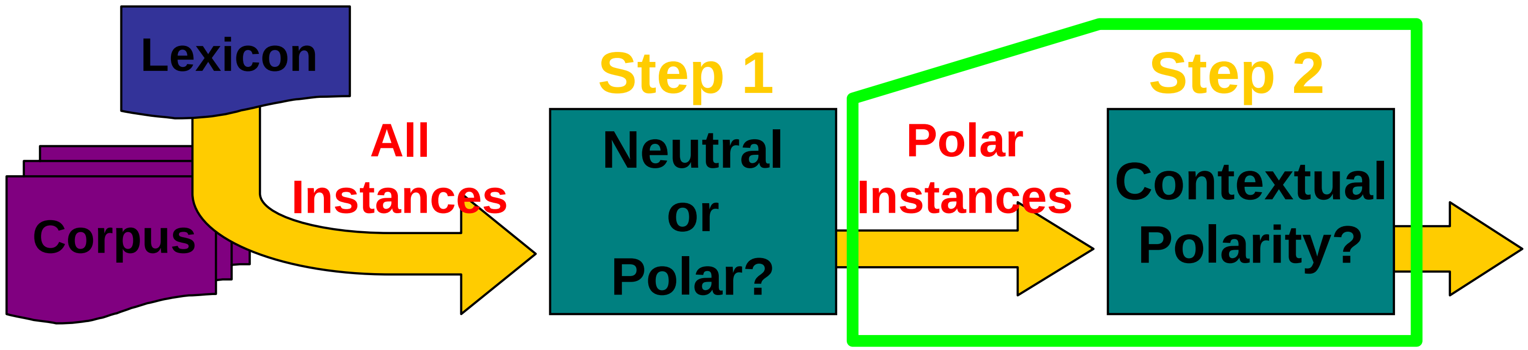
  substantial: negative

- **Modified by polarity**

  5 values: positive, negative, neutral, both, not mod

  ***challenge*: positive**

substantial (pos) ***challenge*** (neg)



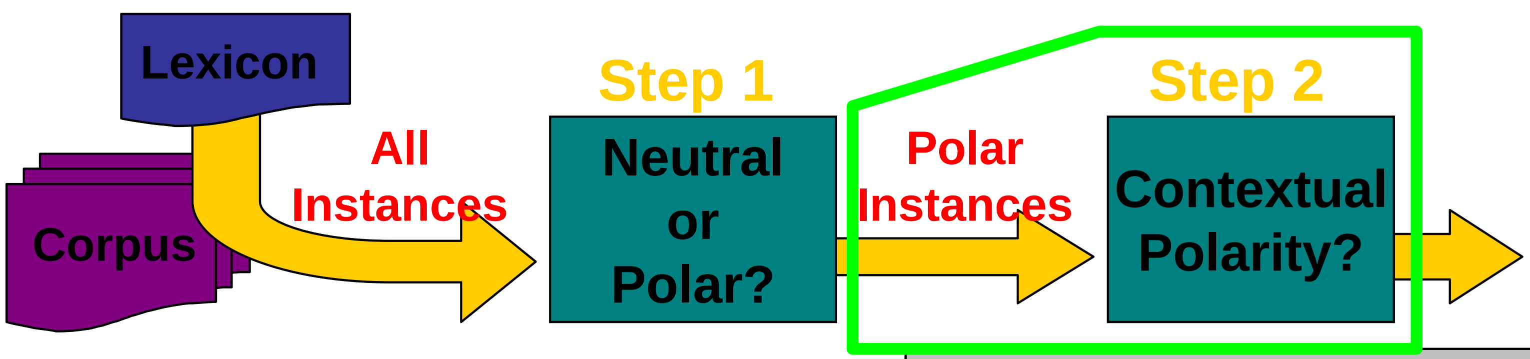
- Word token
- Word prior polarity
- Negated
- Negated subject
- Modifies polarity
- Modified by polarity
- **Conjunction polarity**
- General polarity shifter
- Negative polarity shifter
- Positive polarity shifter

- **Conjunction polarity**

5 values: positive, negative, neutral, both, not mod

good: ***negative***

good (pos) *and* ***evil*** (neg)



- Word token
- Word prior polarity
- Negated
- Negated subject
- Modifies polarity
- Modified by polarity
- Conjunction polarity
- **General polarity shifter**
- **Negative polarity shifter**
- **Positive polarity shifter**

- General polarity shifter

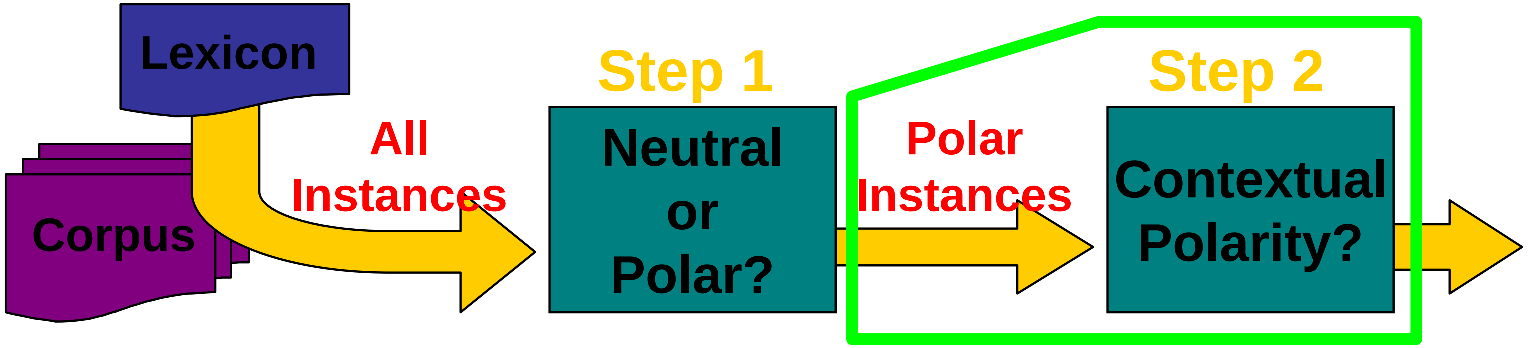
  have ***few*** ***risks*****/rewards**

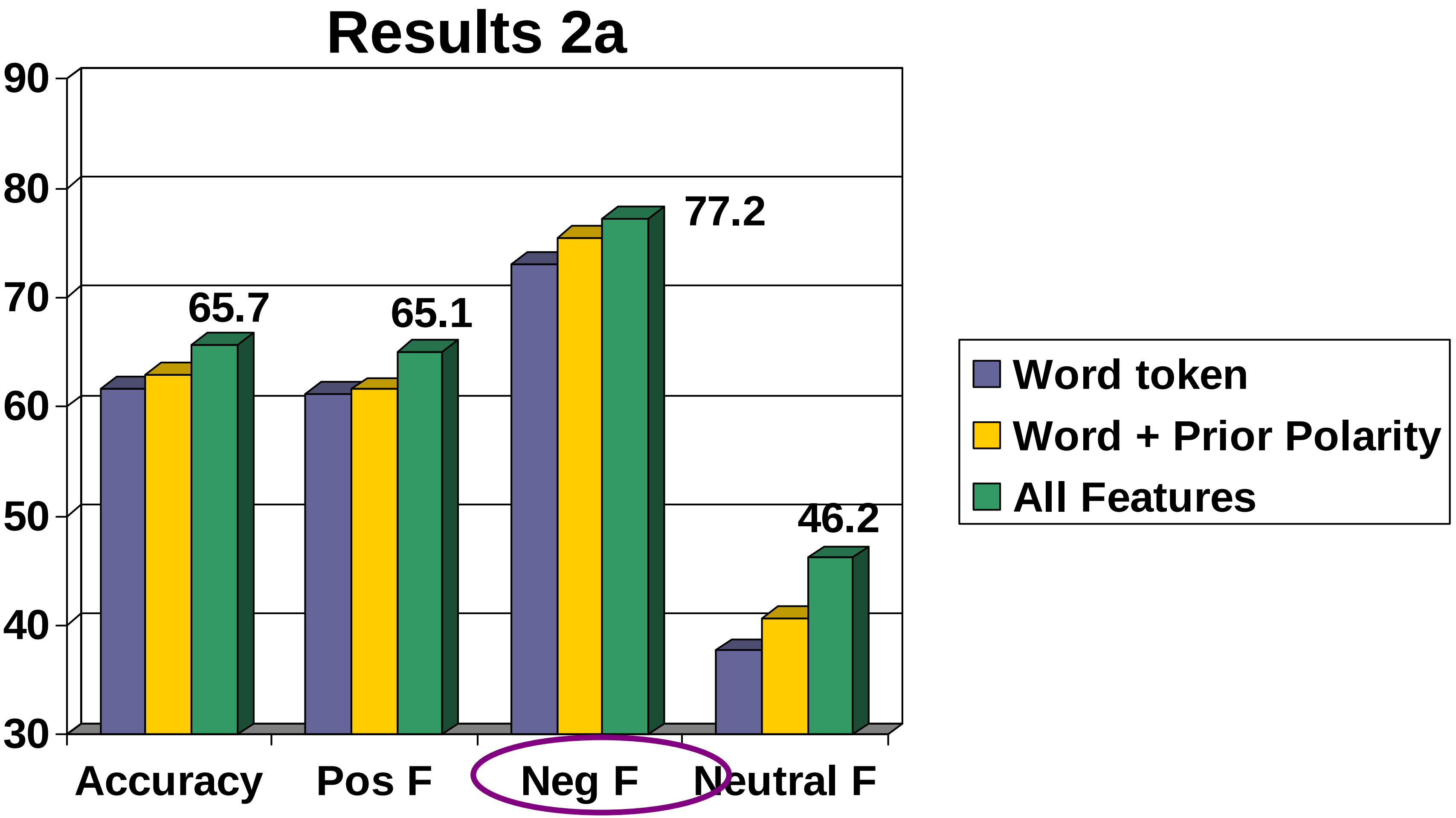
- Negative polarity shifter

  lack *of* **understanding**

- Positive polarity shifter

  abate *the* ***damage***


Lexicon
Corpus
All Instances
Step 1
Neutral or Polar?
Polar Instances
Step 2
Contextual Polarity?


Results 2a
90
80
70
60
50
40
30
65.7
65.1
77.2
46.2
Accuracy
Pos F
Neg F
Neutral F
Word token
Word + Prior Polarity
All Features

Outline

- Corpus Annotation
- Pure NLP
  - Lexicon development
  - Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis
- Applications
  - **Product review mining**

Product review mining

"First, they do an on-line search."

Product review mining

- Goal: summarize a **set** of reviews
- **Targeted** opinion mining: topic is given
- Two levels:
  - Product
  - Product and features
- Typically done for pre-identified reviews but review identification may be necessary

Laptop review 1

- **A Keeper**
  - Reviewed By: N.N. on 5/12/2007
  - **Tech Level**: average - **Ownership**: 1 week to 1 month
  - **Pros:** Price/Value. XP OS NOT VISTA! Screen good even in bright daylignt. Easy to access USB, lightweight.
  - **Cons:** A bit slow - since we purchased this for vacation travel (email & photos) speed is not a problem.
  - **Other Thoughts:** Would like to have card slots for camera/PDA cards. Wish we could afford two so we can have a "spare".

Laptop review 1

- **A Keeper**
  - Reviewed By: N.N. on 5/12/2007
  - **Tech Level**: average - **Ownership**: 1 week to 1 month
  - **Pros:** Price/Value. XP OS NOT VISTA! Screen good even in bright **daylignt**. Easy to access USB, lightweight.
  - **Cons:** A bit slow - since we purchased this for vacation travel (email & photos) speed is not a problem.
  - **Other Thoughts:** Would like to have card slots for camera/PDA cards. Wish we could afford two so we can have a "spare".

Laptop review 2

- By **N.N.** (New York - USA) - See all my reviews
  I was looking for a laptop for long time, doing search, comparing brands, technology, cost/benefits etc.... I should say that I am a normal user and this laptop satisfied all my expectations, the screen size is perfect, its very light, powerful, bright, lighter, elegant, delicate... But the only think that I regret is the Battery life, barely 2 hours... some times less... it is too short... this laptop for a flight trip is not good companion...
  Even the short battery life I can say that I am very happy with my Laptop VAIO and I consider that I did the best decision. I am sure that I did the best decision buying the SONY VAIO

Laptop review 2

- By **N.N.** (New York - USA) - See all my reviews
  I was looking for a laptop for long time, doing search, comparing brands, technology, cost/benefits etc.... I should say that **I am a normal user** and this laptop satisfied all my expectations, the screen size is perfect, **its** very light, powerful, bright, lighter, elegant, delicate... But the only **think** that I regret is the Battery life, barely 2 hours**...** some times less**...** it is too short... this laptop for a flight trip is not good companion...
  Even the short battery life I can say that I am very happy with my Laptop VAIO and I consider that I did the best decision. I am sure that I did the best decision buying the SONY VAIO

Laptop review 3

- **LOVE IT....Beats my old HP Pavillion hands down**, May 16, 2007
- By **N.N.** (Chattanooga, TN USA) - See all my reviews
 I'd been a PC person all my adult life. However I bought my wife a 20" iMac for Christmas this year and was so impressed with it that I bought the 13" MacBook a week later. It's faster and extremely more reliable than any PC I've ever used. Plus nobody can design a gorgeous product like Apple. The only down side is that Apple ships alot of trial software with their products. For the premium price you pay for an Apple you should get a full software suite. Still I'll never own another PC. I love my Mac!

Laptop review 3

- **LOVE IT....Beats my old HP Pavillion hands down**, May 16, 2007
- By **N.N.** (Chattanooga, TN USA) - See all my reviews
  I'd been a PC person all my adult life. However I bought my wife a 20" iMac for Christmas this year and was so impressed with it that I bought the 13" MacBook a week later. It's faster and extremely more reliable than any PC I've ever used. Plus nobody can design a gorgeous product like Apple. The only down side is that Apple ships alot of trial software with their products. For the premium price you pay for an Apple you should get a full software suite. Still I'll never own another PC. I love my Mac!

Some challenges

- Available NLP tools have harder time with review data (misspellings, incomplete sentences)
- Level of user experience (novice, …, prosumer)
- Various types and formats of reviews
- Additional buyer/owner narrative
- What rating to assume for unmentioned features?
- How to aggregate positive and negative evaluations?
- How to present results?

Core tasks of review mining

- Finding product features
- Recognizing opinions

Feature finding

- Wide variety of linguistic expressions can evoke a product feature
  - … you **can't see** the LCD very well in sunlight.
  - … it is very **difficult to see** the LCD.
  - … in the sun, the LCD screen is **invisible**
  - It is very **difficult** to take pictures outside in the sun **with only the LCD screen**.

Opinions v. Polar facts

- Some statements **invite** emotional appraisal but do not explicitly denote appraisal.
- While such polar facts may in a particular context seem to have an obvious value, their evaluation may be very different in another one.

- **A Keeper**
  - Reviewed By: N.N. on 5/12/2007
  - Tech Level: average - Ownership: 1 week to 1 month
  - **Pros:** Price/Value. **XP OS NOT VISTA**! Screen good even in bright daylignt. Easy to access USB, lightweight.
  - **Cons:** A bit slow - since we purchased this for vacation travel (email & photos) speed is not a problem.
  - **Other Thoughts:** Would like to have card slots for camera/PDA cards. Wish we could afford two so we can have a "spare".

Use coherence to resolve orientation of polar facts

- Is a sentence framed by two positive sentences likely to also be positive?
- Can context help settle the interpretation of inherently non-evaluative attributes (e.g. *hot room* v. *hot water* in a hotel context; Popescu & Etzioni 2005) ?

Specific papers using these ideas

Just a Sampling...

Dave, Lawrence, Pennock 2003
Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews

- Product-level review-classification
- Train Naïve Bayes classifier using a corpus of self-tagged reviews available from major web sites (C|net, amazon)
- Refine the classifier using the same corpus before evaluating it on sentences mined from broad web searches

Dave, Lawrence, Pennock 2003

- Feature selection
  - Substitution (statistical, linguistic)
    - I called Kodak
    - I called Nikon
    - I called Fuji

I called
COMPANY

Dave, Lawrence, Pennock 2003

- Feature selection
  - Substitution (statistical, linguistic)
  - Backing off to wordnet synsets
    - brilliant -> {brainy, brilliant, smart as a whip}

Dave, Lawrence, Pennock 2003

- Feature selection
  - Substitution (statistical, linguistic)
  - Backing off to wordnet synsets
  - Stemming
    - bought them
    - buying them
    - buy them

buy them

Dave, Lawrence, Pennock 2003

- Feature selection
  - Substitution (statistical, linguistic) Backing off to wordnet synsets
  - Stemming
  - N-grams
    - last long enough
    - too hard to

Dave, Lawrence, Pennock 2003

- Feature selection
  - Substitution (statistical, linguistic)
    Backing off to wordnet synsets
  - Stemming
  - N-grams
  - arbitrary-length substrings

Dave, Lawrence, Pennock 2003

- Laplace (add-one) smoothing was found to be best
- 2 types of test (1 balanced, 1 unbalanced)
  - SVM did better on Test 2 (balanced data) but not Test 1
- Experiments with weighting features did not give better results

Hu & Liu 2004
Mining Opinion Features in Customer Reviews

- Here: explicit product features only, expressed as nouns or compound nouns
- Use association rule mining technique rather than symbolic or statistical approach to terminology
- Extract associated items (item-sets) based on support (>1%)

Hu & Liu 2004

- Feature **pruning**
  - compactness
    - *"I had searched for a digital camera for 3 months."*
    - *"This is the best digital camera on the market"*
    - *"The camera does not have a digital zoom"*
  - Redundancy
    - manual ; manual mode; manual setting

Hu & Liu 2004

- For sentences with frequent feature, extract nearby **adjective** as opinion
- Based on opinion words, gather infrequent features (N, NP nearest to an opinion adjective)
  - *The salesman was easy going and let me try all the models on display.*

Yi & Niblack 2005

Sentiment mining in WebFountain

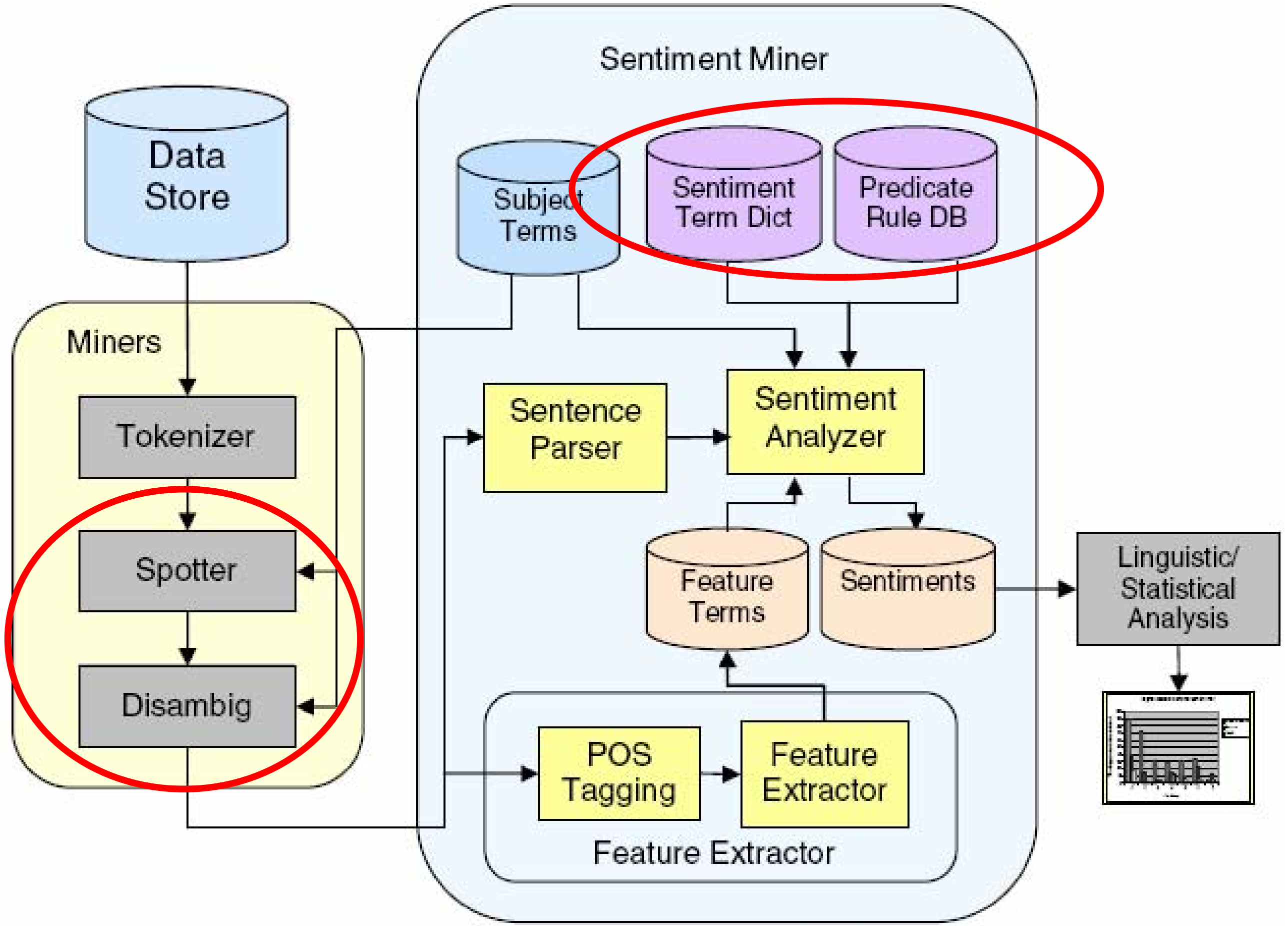


Figure 2. The Sentiment Mining Process with a Predefined Set of Subjects

Yi & Niblack 2005

- Product feature terms are extracted heuristically, with high precision
  - For all definite base noun phrases,
    - the NN
    - the JJ NN
    - the NN NN NN
    - …
  - calculate a statistic based on likelihood ratio test

$$-2log\lambda = \begin{cases} -2 * lr & \text{if } r_2 < r_1 \\ 0 & \text{if } r_2 \geq r_1 \end{cases} \quad (1)$$

$$\begin{aligned} lr &= (C_{11} + C_{21}) \cdot log(r) \\ &\quad +(C_{12} + C_{22}) \cdot log(1 - r) \\ &\quad -C_{11}log(r_1) - C_{12}log(1 - r_1) \\ &\quad -C_{21}log(r_2) - C_{22}log(1 - r_2) \end{aligned}$$

$$\begin{aligned} r_1 &= \frac{C_{11}}{C_{11}+C_{12}} \\ r_2 &= \frac{C_{21}}{C_{21}+C_{22}} \\ r &= \frac{C_{11}+C_{21}}{C_{11}+C_{12}+C_{21}+C_{22}} \end{aligned}$$

Yi & Niblack 2005

- Manually constructed
  - Sentiment lexicon: *excellent JJ +*
  - Pattern database: *impress + PP(by; with)*
- Sentiment miner identifies the best fitting pattern for a sentence based on the parse

Yi & Niblack 2005

- Manually constructed
  - Sentiment lexicon: *excellent JJ +*
  - Pattern database: *impress + PP(by; with)*
- Sentiment miner identifies the best fitting pattern for a sentence based on the parse
- Sentiment is assigned to opinion target

Yi & Niblack 2005

- Discussion of hard cases:
  - Sentences that are ambiguous out of context
  - Cases that did not express a sentiment at all
  - Sentences that were not about the product:
    - → Need to associate opinion and target

Summary

- Subjectivity is common in language

Summary

- Subjectivity is common in language
- Recognizing it is useful in many NLP tasks

Summary

- Subjectivity is common in language
- Recognizing it is useful in many NLP tasks
- It comes in many forms and often is context-dependent

Summary

- Subjectivity is common in language
- Recognizing it is useful in many NLP tasks
- It comes in many forms and often is context-dependent
- Contextual coherence and distributional similarity are important linguistic notions in lexicon building
- A wide variety of features seem to be necessary for opinion and polarity recognition

Summary

- Subjectivity is common in language
- Recognizing it is useful in many NLP tasks
- It comes in many forms and often is context-dependent
- Contextual coherence and distributional similarity are important linguistic notions in lexicon building
- A wide variety of features seem to be necessary for opinion and polarity recognition

Additional material

Some Early Work on Point of View

- Jame Carbonell 1979. Subjective Understanding: Computer Models of Belief Systems. *PhD Thesis.*
- Yorick Wilks and Janusz Bien 1983. Beliefs, Points of View, and Multiple Environments. *Cognitive Science (7).*
- Eduard Hovy 1987. Generating Natural Language under Pragmatic Constraints. *PhD Thesis.*

Our Early Work on Point of View

- Jan Wiebe & William Rapaport 1988. A Computational Theory of Perspective and Reference in Narrative. *ACL.*
- Jan Wiebe 1990. Recognizing Subjective Sentences: A Computational Investigation of Narrative Text. *PhD Thesis.*
- Jan Wiebe 1994. Tracking Point of View in Narrative. *Computational Linguistics 20 (2).*

Work on the intensity of private states

- Theresa Wilson, Janyce Wiebe and Rebecca Hwa 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence, 22 (2), pp. 73-99*.
- Theresa Wilson 2007. *Ph.D. Thesis.* Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of private states.

- James R. Martin and Peter R.R. White. 2005. The Language of Evaluation: The Appraisal Framework.
  - An approach to evaluation that comes from within the theory of systemic-functional grammar.
- Website on this theory maintained by P.R. White:
  - http://www.grammatics.com/appraisal/index.html

- Kenneth Bloom, Navendu Garg, and Shlomo Argamon 2007. Extracting Appraisal Expressions. *NAACL.*
- Casey Whitelas, Navendu Garg, and Shlomo Argamon 2005. Using appraisal groups for sentiment analysis. *CIKM*.

More work related to lexicon building

- Alina Andreevskaia and Sabine Bergler. 2006. Sentiment Tag Extraction from WordNet Glosses. *LREC.*
- Nancy Ide 2006. Making senses: bootstrapping sense-tagged lists of semantically-related words. *CICling.*
- Jan Wiebe and Rada Mihalcea 2006. Word Sense and Subjectivity. *ACL*
- Riloff, Patwardhan, Wiebe 2006. Feature Subsumption for Opinion Analysis. *EMNLP.*

- Alessandro Valitutti, Carol Strapparava, & Oliviero Stock 2004. Developing affective lexical resources. *PsychNology.*
- M. Taboada, C. Anthony, and K. Voll 2006. Methods for creating semantic orientation databases. *LREC.*

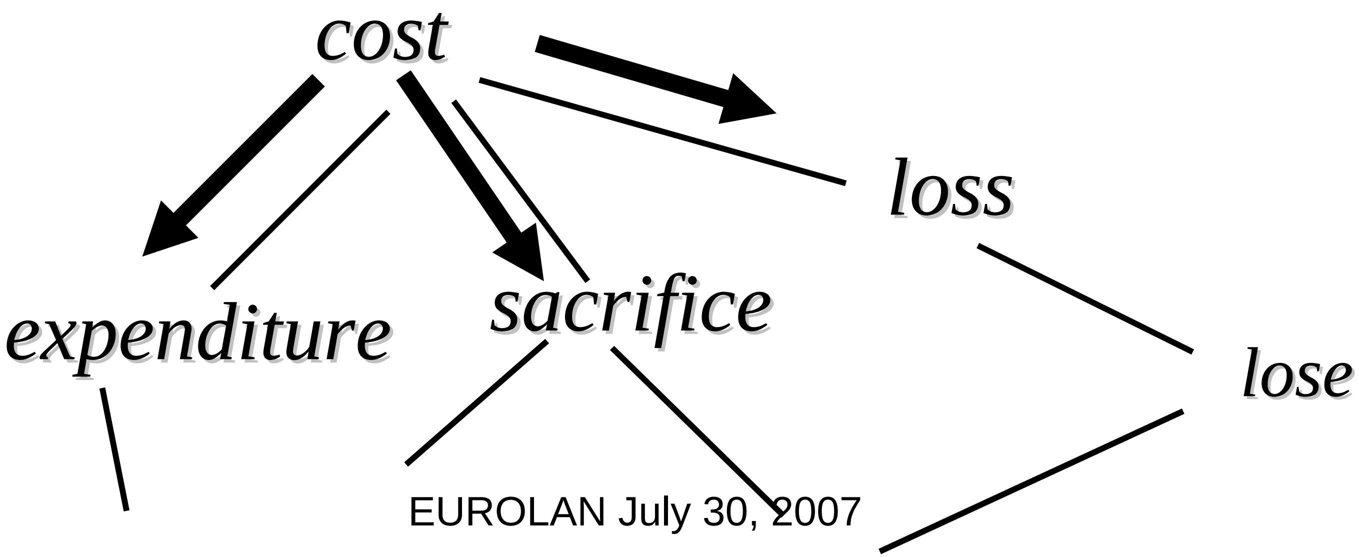
Takamura et al. 2007

Extracting Semantic Orientations of Phrases from Dictionary

- Use a Potts model to categorize Adj+Noun phrases
- Targets ambiguous adjectives like low, high, small, large
- Connect two nouns, if one appears in gloss of other
- Nodes have orientation values (pos, neg, neu) and are connected by same or different orientation links

A Sample Lexical Network

WORD	*GLOSS*
cost	*loss or sacrifice, expenditure*
loss	*something lost*



Takamura et al 2007

Probabilistic Model on the Lexical Network (Potts model)

$$H(\mathbf{c}) = -\beta \sum_{ij} w_{ij}\delta(c_i, c_j) + \alpha \sum_{i \in L} -\delta(c_i, a_i)$$

i and j index for node

L set of seed words

c_i state of node i

a_i class label of seed word i

α β constants

Takamura et al. 2007

$$H(\mathbf{c}) = -\beta \sum_{ij} w_{ij}\delta(c_i, c_j) + \alpha \sum_{i \in L} -\delta(c_i, a_i)$$

- The state of a seed word becomes a_i
- Neighboring nodes tend to have the same label.

“low cost” = “low expenditure”

Takamura et al. 2007

- Manually labeled adj+noun data provide noun seeds of known orientation
- The network assigns orientation to nouns not seen in training data

Further work on review mining

- Morinaga et. al. 2002. Mining Product Reputations on the Web
- Kobayashi et al. 2004. Collecting Evaluative Expressions for Opinion Extraction
- Hu & Liu. 2006. Opinion Feature Extraction Using Class Sequential Rules

Popescu & Etzioni 2005

- Report on a product review mining system that extracts and labels opinion expressions their attributes
- They use the relaxation-labeling technique from computer vision to perform unsupervised classification satisfying local constraints (which they call neighborhood features)
- The system tries to solve several classification problems (e.g. opinion and target finding) at the same time rather than separately.

Applications of Subjectivity and Sentiment analysis not discussed earlier

Question Answering

"I'd like your honest, unbiased and possibly career-ending opinion on something."

Question Answering

- Much work on Subjectivity & Sentiment has been motivated by QA.
  - Yu, H. & Hatzivassiloglou, V. (2003)
  - Kim, S. & Hovy, E. (AAAI-Workshop 2005)
- Some QA work has also indicated that making a subjective/objective distinction would be useful for the seemingly objective task of definitional QA
  - Lita et al. (2005)

Question Answering

- Making subjective/objective distinction has been showed to be useful in answering opinion based questions in news text
  - Stoyanov et al.(2005)
- Making finer grained distinction of subjective types (Sentiment and Arguing) further improves the QA system (world press and online discussion forums)
  - Somasundaran et al. (2007)

Information Extraction

- Information Extraction has been used to learn subjective patterns
  - Wiebe and Riloff (2005)
- Subjectivity has been shown to improve IE
  - Riloff et al. (2005)

Summarization

- Opinion Summaries from documents have been created
  - Stoyanov & Cardie (2006)
    - They combine fine grained opinions from the same source to create a source specific summary of opinion
  - Carenini et al.(IUI-2006)
    - They summarize a large corpora of evaluative text about a single entity (product)
- Different aspects of subjectivity analysis have been used to enhance summarizing systems.
  - Seki et al. (2005)
    - Summarization based on user's needs (benefits, positive/negative factors, commentary, etc).

Blog analysis

- Analysis of sentiments on Blog posts
  - Chesley et al.(2006)
    - Perform subjectivity and polarity classification on blog posts
- Sentiment has been used for blog analysis
  - Balog et al. (2006)
    - Discover irregularities in temporal mood patterns (fear, excitement, etc) appearing in a large corpus of blogs
  - Kale et al. (2007)
    - Use link polarity information to model trust and influence in the blogosphere
- Blog sentiment has been used in applications
  - Mishne and Glance (2006)
    - Analyze Blog sentiments about movies and correlate it with its sales

Human Computer Interaction

- Affect sensing
  - Liu et al. (2003)

- Human Robot Interaction
  - Tokuhisa & Terashima (2006)
    - Correlate enthusiasm levels in dialogs with subjective language for human robot interaction

Visualization

- Visualization of sentiments
  - Gregory et al. (2007)
    - Visualize affect distribution in social media (blogs) for a topic.
  - Gamon et al. (2005)
    - Visualize sentiment and its orientation for a topic from large number of customer feedback texts.

Trends & Buzz

- Stock market
  - Koppel & Shtrimberg(2004)
    - Correlate positive/negative news stories about publicly traded companies and the stock price changes

- Market Intelligence from message boards, forums, blogs.
  - Glance et al. (2005)

Source and Target Finding

"Who is the fairest one of all,
and state your sources!"

Bethard et al. 2004

Automatic Extraction of Opinion Propositions and their Holders

- Find verbs that express opinions in *propositional* form, and their holders
  - Still, Vista officials realize they're relatively fortunate.
- Modify algorithms developed in earlier work on semantic parsing to perform binary classification (opinion or not)
- Use presence of subjectivity clues to identify opinionated uses of verbs

Choi et al.2005

Identifying sources of opinions with conditional random fields and extraction patterns

- Treats **source** finding as a combined sequential tagging and information extraction task
- IE patterns are high precision, lower recall
- Base CRF uses information about noun phrase semantics, morphology, syntax
- IE patterns connect opinion words to sources
- Conditional Random Fields given IE features perform better than CRFs alone

Kim & Hovy 2006

Extracting opinions, opinion holders, and topics expressed in online news media text

- Perform semantic role labeling (FrameNet) for a set of adjectives and verbs (pos, neg)
- Map semantic roles to holder and target
  - E.g. for Desiring frame: Experiencer->Holder
- Train on FN data, test on FN data and on news sentences collected and annotated by authors' associates
- Precision is higher for topics, recall for holders

Choi, Breck, Cardie 2006

Joint extraction of entities and relations for opinion reocgnition

- Find direct expressions of opinions and their sources jointly
- Uses sequence-tagging CRF classifiers for opinion expressions, sources, and potential link relations
- Integer linear programming combines local knowledge and incorporates constraints
- Performance better even on the individual tasks

Further references on Source and Target Finding

- Breck & Cardie 2004. Playing the Telephone Game: Determining the Hierarchical Structure of Perspective and Speech Expressions.

- Bloom et al. 2007. Extracting Appraisal Expressions.

2007 NLP papers NAACL

- N07-1037 [bib]: **Hiroya Takamura; Takashi Inui; Manabu Okumura**
  Extracting Semantic Orientations of Phrases from Dictionary
- N07-1038 [bib]: **Benjamin Snyder; Regina Barzilay**
  Multiple Aspect Ranking Using the Good Grief Algorithm
- N07-1039 [bib]: **Kenneth Bloom; Navendu Garg; Shlomo Argamon**
  Extracting Appraisal Expressions

2007 NLP Papers ACL [1]

- P07-1053 [bib]: **Anindya Ghose; Panagiotis Ipeirotis; Arun Sundararajan**
  Opinion Mining using Econometrics: A Case Study on Reputation Systems
- P07-1054 [bib]: **Andrea Esuli; Fabrizio Sebastiani**
  PageRanking WordNet Synsets: An Application to Opinion Mining
- P07-1055 [bib]: **Ryan McDonald; Kerry Hannan; Tyler Neylon; Mike Wells; Jeff Reynar**
  Structured Models for Fine-to-Coarse Sentiment Analysis
- P07-1056 [bib]: **John Blitzer; Mark Dredze; Fernando Pereira**
  Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification

2007 NLP Papers
ACL [2]

- P07-1123 [bib]: **Rada Mihalcea; Carmen Banea; Janyce Wiebe**
  Learning Multilingual Subjective Language via Cross-Lingual Projections
- P07-1124 [bib]: **Ann Devitt; Khurshid Ahmad**
  Sentiment Polarity Identification in Financial News: A Cohesion-based Approach
- P07-1125 [bib]: **Ben Medlock; Ted Briscoe**
  Weakly Supervised Learning for Hedge Classification in Scientific Literature

2007 NLP Papers EMNLP

- D07-1113 [bib]: **Soo-Min Kim; Eduard Hovy** *Crystal: Analyzing Predictive Opinions on the Web*
- D07-1114 [bib]: **Nozomi Kobayashi; Kentaro Inui; Yuji Matsumoto** *Extracting Aspect-Evaluation and Aspect-Of Relations in Opinion Mining*
- D07-1115 [bib]: **Nobuhiro Kaji; Masaru Kitsuregawa** *Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents*

Bibliographies

- Bibliography of papers in this tutorial:
  - www.cs.pitt.edu/~wiebe/eurolan07.bib
  - www.cs.pitt.edu/~wiebe/eurolan07.html
- Andrea Esuli's extensive "Sentiment Classification" bibliography (not limited to sentiment or classification)
  - http://liinwww.ira.uka.de/bibliography/Misc/Sentiment.html

Yahoo! Group

- SentimentAI
  - http://tech.groups.yahoo.com/group/SentimentAI/