

# Thumbs Up or Thumbs Down? Detecting Sentiment and Opinion Automatically

Maite Taboada\*  
Department of Linguistics  
mtaboada@sfu.ca

Defining Cognitive Science  
November 7, 2007

\*and Caroline Anthony, Julian Brooke, Katia Dilkina, Mary Ann Gillies,  
Jack Grieve, Montana Hay, Patrick Larrivee-Woods, Graham Lyons,  
Paul McFetridge, Robert Outtrim, Kimberly Voll

# A movie review



Ok let me get this out of the way: I did NOT like this movie. Borderline hated it. Now if you are only after positive reviews, then you might skip this, but if you are interested in why I thought it sucked, read on.

To begin with, I only mildly like Will Ferrell. I don't think he's funny at all on SNL, but I was willing to give him a chance. I had seen previews for this a few weeks before it came out and it looked interesting, so I figured I would give it a shot.

I'm very thankful that a local theater gives military discounts, I only paid \$5.50 to see it, and after watching the movie, I feel like I COMPLETELY WASTED my money.

It's a 100% kids movie. There is ZERO bad language in it. The special effects and props are EXTREMELY low budget and so fake its not even funny. You can clearly tell what parts were filmed on a sound stage, as the backgrounds are blatantly fake, no effort was put into them.

[...]

# Sentiment detection

- Evaluation, subjectivity, opinion, sentiment, stance
- Sentiment
  - Opinion expressed in discourse
  - About anything: a book, a movie, a new car, a politician, a group of people
  - Anywhere: a personal conversation, an e-mail message, an editorial, a newspaper article, a conference presentation, a business report

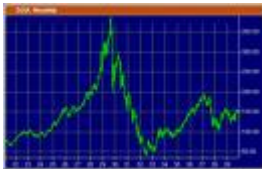
# Applications



- Search for a movie
  - Results are returned in categories: positive and negative reviews



- Consumer reviews
  - Products are ranked according to their reviews



- Market intelligence
  - What are people saying about a new product?  
What do they think about a company?



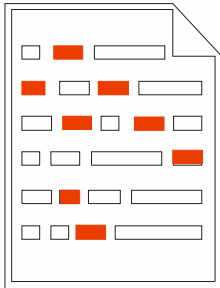
- Politics
  - Opinions about a candidate, a policy, a new piece of legislation (maybe also over time)

# My goal

- Classification of texts based on subjective content (=sentiment)
  - Positive
  - Negative
- Input a text, and produce a numeric value that expresses its subjective content → the text's sentiment
  - -3: quite negative
  - -5: very negative
  - 2: somewhat positive
  - ...



# Approaches



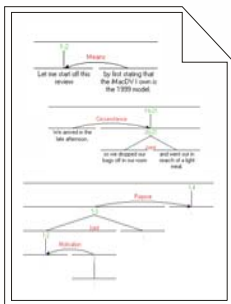
## 1. Key words

- Adjectives (the vast majority of the current research)
- Intensifiers, negation
- Other words (nouns, verbs)



## 2. Topic sentences

- Extract sentiment from topic sentences only



## 3. Discourse structure

- Aggregate sentiment based on the structure of the text



# 1. Key words



Ok let me get this out of the way: I **did NOT** like this movie. Borderline **hated** it. Now if you are only after **positive reviews**, then you might skip this, but if you are interested in why I thought it **sucked**, read on.

To begin with, I only **mildly like** Will Ferrell. I **don't think he's funny** at all on SNL, but I was willing to give him a chance. I had seen previews for this a few weeks before it came out and it looked **interesting**, so I figured I would give it a shot.

I'm **very thankful** that a local theater gives military discounts, I only paid \$5.50 to see it, and after watching the movie, I feel like I **COMPLETELY WASTED** my money.

It's a 100% kids movie. There is **ZERO bad language** in it. The special effects and props are **EXTREMELY low budget** and so **fake its not even funny**. You can clearly tell what parts were filmed on a sound stage, as the backgrounds are **blatently fake**, no effort was put into them.

# Key words: Extracting semantic orientation

- Words have semantic content or orientation (SO)
- SO for adjectives extracted automatically (Hatzivassiloglou and McKeown 1997)
- SO for adjectives or other words can be extracted using different methods:
  - Machine learning (Pang et al. 2002)
  - Pointwise Mutual Information (Turney 2002)
- Basic idea for automatic SO of words: you know the SO of a word by the company that it keeps
  - Excellent and X
  - Depressing but X

# Adjectives

- Need to build a dictionary of adjectives with their semantic orientation
- Used Google and pointwise mutual information
- Results okay, but hit counts vary on different days
- Currently, we're using a hand-built dictionary

Adverb	Highest value	Lowest value	Average value	Standard deviation
away	-1.7719	-6.1984	-3.5277	1.5241
equally	-2.9020	-5.7973	-3.6547	1.0045
eventually	-0.9755	-8.3008	-4.0837	2.1861
hastily	-1.9759	-8.6323	-4.5750	2.1765
heavily	-0.9695	-9.2568	-3.8807	2.7425
madly	-6.8169	-11.5520	-8.3208	1.8720
masterfully	-2.4512	-7.8460	-3.3973	1.9736

# Performance: adjectives only

- Compared to what reviewer's say (recommended or not recommended)

Dictionary	Percent correct
Google	56%
Hand-ranked	63%

- Taboada, Anthony and Voll (2006), Voll and Taboada (2007)

# Intensifiers, downplayers and negation

- Valence shifters (Polanyi and Zaenen 2004)
  - very interesting
  - too bad
  - not bad at all
  - It is sufficient vs. It is **barely** sufficient
  - He got into Harvard vs. He **even** got into Harvard

	not	the	least	bit	interesting	
base value					3	
negation					-3	
downplayer			↓-2	↓-.5	-3	
neg. downpl.			2.5		-3	
final value						-7.5

# Performance: intensifiers and negation

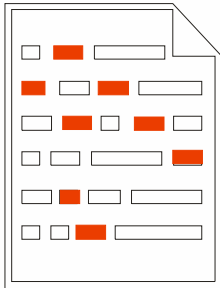
---

Words used	Percent correct
Adjectives only	62.2%
Adjectives, negation and intensifiers	64%
Adjectives, nouns, verbs (with negation and intensifiers)	68%

---

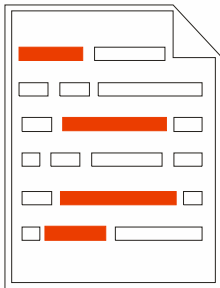
Work by Julian Brooke

# Approaches



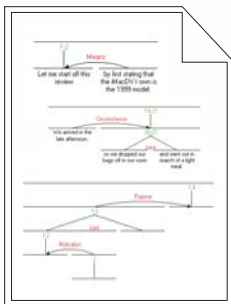
## 1. Key words

- Adjectives (the vast majority of the current research)
- Intensifiers, negation
- Other words (nouns, verbs)



## 2. Topic sentences

- Extract sentiment from topic sentences only



## 3. Discourse structure

- Aggregate sentiment based on the structure of the text



## 2. Topic sentences

- Subject or topic being discussed (“The popcorn was **great**. The movie **sucked**.”)

In California's San Joaquin Delta region, peat dirt islands which help to make up the Riparian feature known as 1,000 Miles Of Waterways, alternately flood and serve to provide the ideal growing medium for one of local agribusinesses major vegetable crops, the asparagus.

For many consumers, this King of Spears is only known as a canned product, to be served chilled, in aspics, (feh!), or salads. Others, encountering the raw variety with its woody base, scaled stalk and feathered head, may be unsure just what to do with it. Which leads me to this niche tool, someone's clever idea, when to my mind a simple 4" paring knife or slightly larger utility knife would do. [Cooking.com Stainless Steel Asparagus Peeler and Tongs \[...\]](#)



# However, in the right context...

Flat out, this popcorn popper makes some of the best popcorn I've had. It comes out light and fluffy, and very tasty. Using this popper I can actually make popcorn in shorter time than using a microwave. It is cheaper, better tasting and faster...what are you waiting for!! Go buy one right now.

Wabash Valley Farms Whirley  
Pop Popcorn Popper, 6-Qt



# Extracting topic sentences

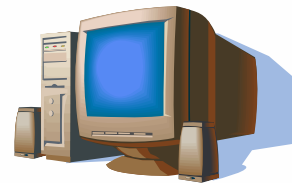
- Topic classifier, using decision trees (WEKA)
- Eight types of reviews, 50 reviews in each:



Books



Cars



Computers



Cookware



Hotels



Movies



Music



Phones

- Train using one as on-topic, all other seven as off-topic
- The classifier is then used to determine which sentences are on-topic, within a document

# Performance: topic classifier

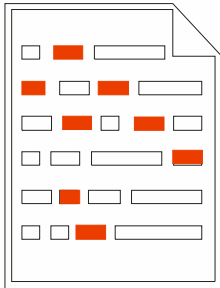
- Tested against a method that uses only adjectives
- Normalization factor: bias towards positive results
  - Negative reviews do not contain as many negative words
- Voll and Taboada (2007)

---

Method	Percent correct
All adjectives	63%
Topic sentences only	69%
Topic sentences, normalized	73%

---

# Approaches



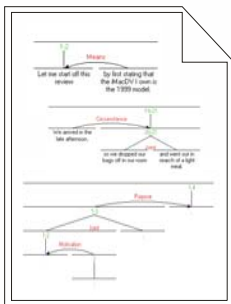
## 1. Key words

- Adjectives (the vast majority of the current research)
- Intensifiers, negation
- Other words (nouns, verbs)



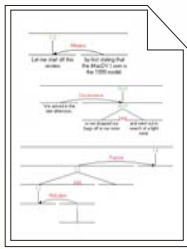
## 2. Topic sentences

- Extract sentiment from topic sentences only



## 3. Discourse structure

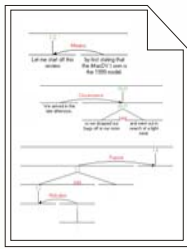
- Aggregate sentiment based on the structure of the text



### 3. Discourse structure



It could have been a great movie. It could have been excellent, and to all the people who have forgotten about the older, greater movies before it, will think that as well. It does have beautiful scenery, some of the best since Lord of the Rings. The acting is well done, and I really liked the son of the leader of the Samurai. He was a likeable chap, and I hated to see him die... But, other than all that, this movie is nothing more than hidden rip-offs.



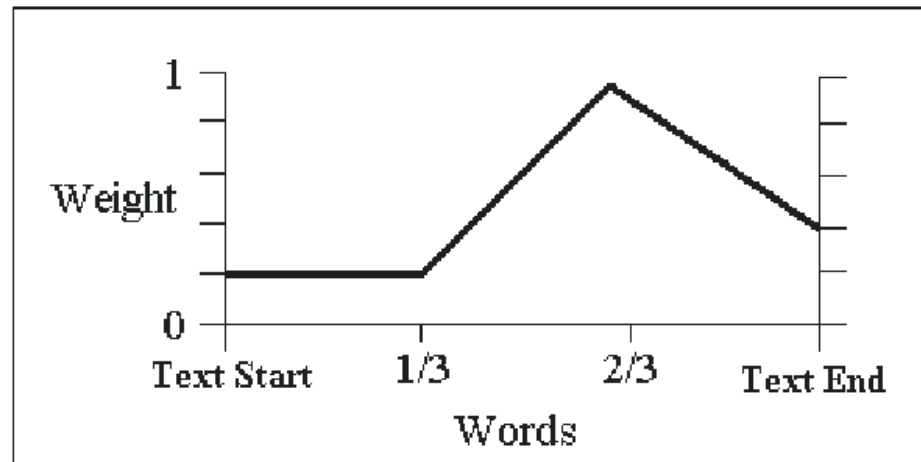
### 3. Discourse structure: keywords are not enough



It could have been a **great** movie. It could have been **excellent**, and to all the people who have forgotten about the older, **greater** movies before it, will think that as well. It does have **beautiful** scenery, **some of the best** since Lord of the Rings. The acting is **well done**, and I **really liked** the son of the leader of the Samurai. He was a **likeable** chap, and I **hated** to see him die... But, other than all that, this movie is nothing more than **hidden rip-offs**.

# Semantic orientation and text structure

- SO based on the number of positive and negative adjectives
- But also weighed according to the adjective's position in the text (naïve approach)

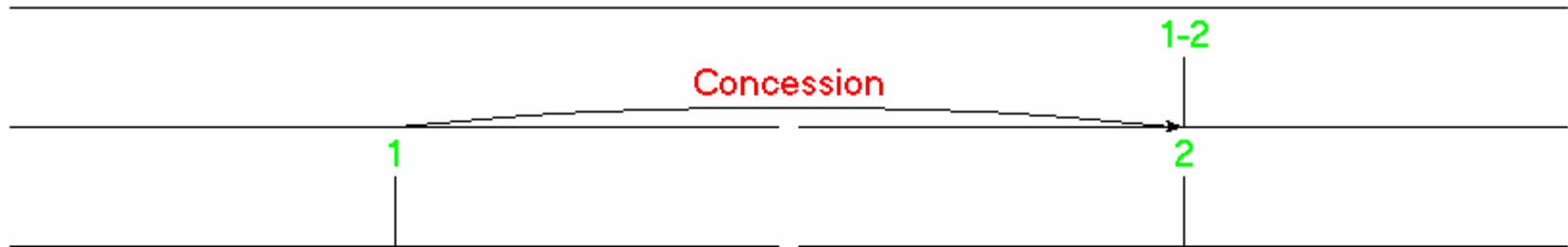


- Evaluation results show accuracy of 65%, when compared with what humans say, better than a baseline with averaged adjectives (Taboada and Grieve 2004)

# More complex structure

- Rhetorical structure of the text
  - Using Rhetorical Structure Theory (Mann and Thompson 1988; Taboada and Mann 2006)
  - Opinions usually summarized at the end
  - Heavy use of concessive relations
  - Elaborations can sometimes be ignored

# Concession



It could have been a great movie. It could have been excellent, and to all the people who have forgotten about the older, greater movies before it, will think that as well. It does have beautiful scenery, some of the best since Lord of the Rings. The acting is well done, and I really liked the son of the leader of the Samurai. He was a likeable chap, and I hated to see him die.

But, other than all that, this movie is nothing more than hidden rip-offs.

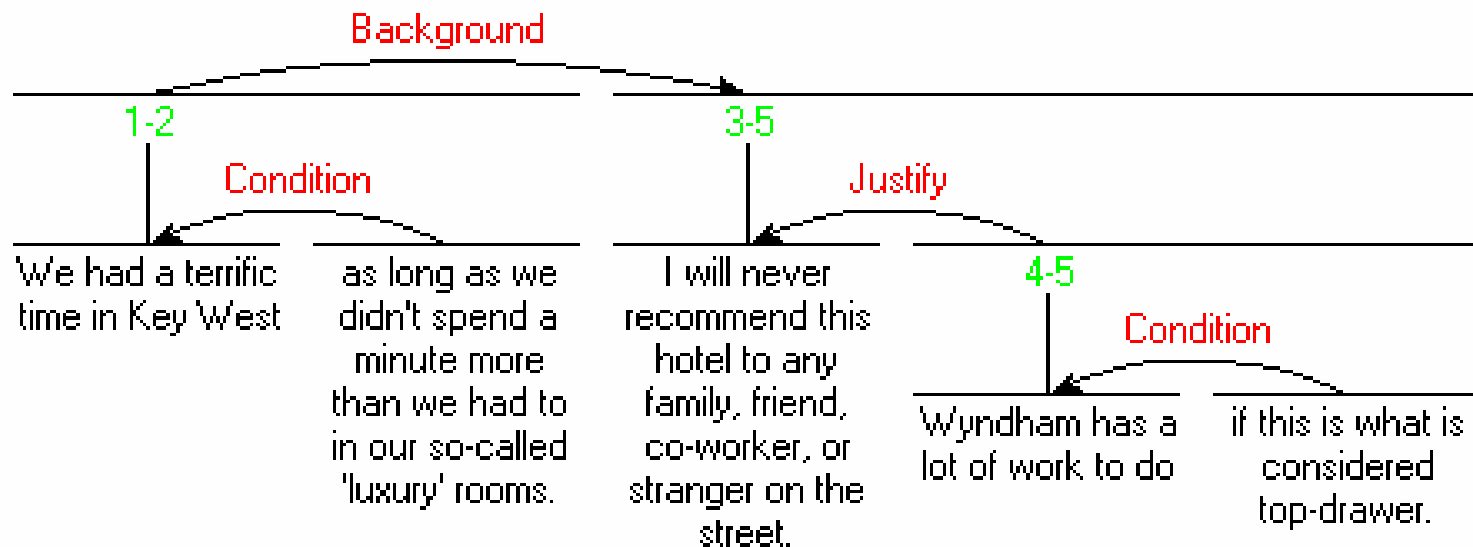
# Rhetorical relations

- Building blocks of texts
- Help explain coherence
- As observed by an analyst, trying to analyze what the author meant
- Psychological validity
  - Do people really have in mind relations such as Cause, Concession, Result, or Elaboration?

# Condition, nuclearity



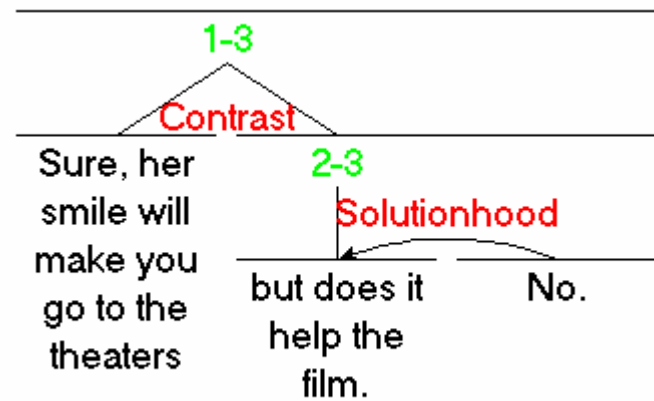
We had a terrific time in Key West as long as we didn't spend a minute more than we had to in our so-called 'luxury' rooms. I will never recommend this hotel to any family, friend, co-worker, or stranger on the street. Wyndham has a lot of work to do if this is what is considered top-drawer.



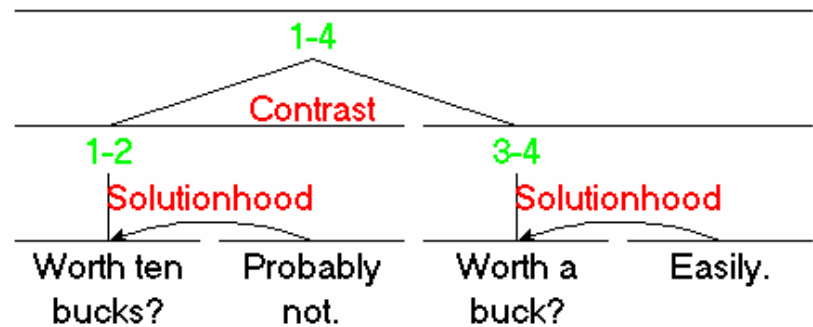
# Question-answer, parallelism



Julia Roberts without a doubt, is a good actress depending on what film she does.... Sure, her smile will make you go to the theaters but does it help the film. No.



Worth ten bucks?  
Probably not.  
Worth a buck?  
Easily.



# Performance: discourse structure

- Used an existing discourse parser that extracts relations at the sentence level (Soricut and Marcu 2003)
- Parses are not quite accurate
- Voll and Taboada (2007)

Method	Percent correct
All adjectives	63%
Topic sentences only	69%
Topic sentences, normalized	73%
Discourse structure	61%
Discourse structure, normalized	69%

# Current research

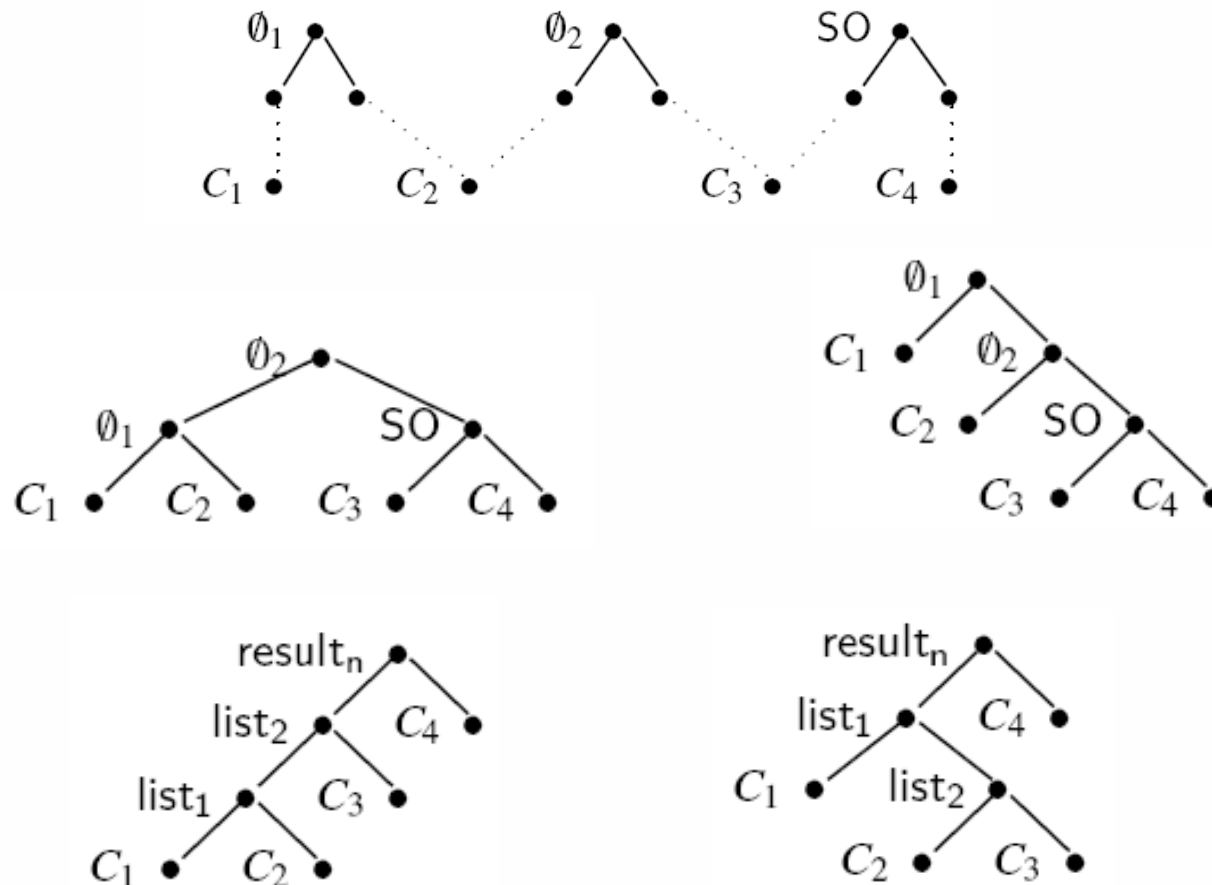
- Parsing rhetorical relations
- Annotation according to Appraisal theory

# Current research: Parsing rhetorical relations

- Segment text into units
- Find cues that indicate a relation
  - Discourse markers (*if, therefore, in summary, however*)
  - Syntactic relations
  - Semantic relations (event types for verbs)
  - Lexical chains
  - Text structure (beginning, middle, end)
- Classify relations
- Build trees for the text
- Discard unsuitable trees

# An Example

John is stubborn ( $C_1$ ). His sister is stubborn ( $C_2$ ). His parents are stubborn ( $C_3$ ). So, these people are continually arguing ( $C_4$ ).



# Appraisal Theory

- Out of work in **Systemic Functional Linguistics** (Martin 2000, White 2003)
- Explores how speakers and writers pass judgement on:
  - People
  - Events, facts
  - Objects, language
- An Appraisal-based analysis can help us discern
  - the type of opinion expressed
  - its intensity
  - the author's commitment to the opinion

# Appraisal types

Attitude

Affect (emotional) – *I feel x*

love Dr. Seuss stories; enjoy reading them;  
upset; didn't really care for her; kinda like her hair;  
I was in denial

Judgement (ethical) – *He/she is x*

subversive art teacher; independent filmmaker;  
control freak; shameless; typical workaholic mom  
who has no time for discipline

Appreciation (aesthetic) – *It is x*

mesh of a film; plot holes; noteworthy performances;  
totally generic; not that funny; easy on the eyes

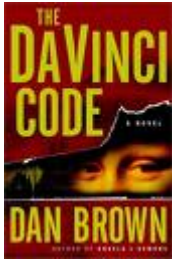
# Sentiment orientation + Appraisal

- We can extract the SO (a numerical value), plus a sense of what type of appraisal is in the text

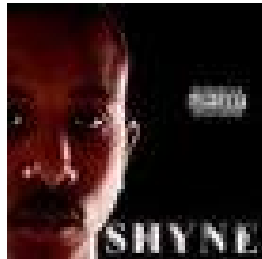


- Currently, annotation work
- With a view toward machine learning approaches
- Annotation by Patrick Larrivee-Woods

# What we can and cannot do automatically



- For an example of how marketing hype can overcome critical judgment and influence popular taste, read "The Da Vinci Code".



- ... this is Great Music to Play While Getting ready to go out, because hearing this garbage makes you want to leave as fast as possible.



- Speaking of kids in the back-seat, this little truck actually has lots of leg room...for kids.

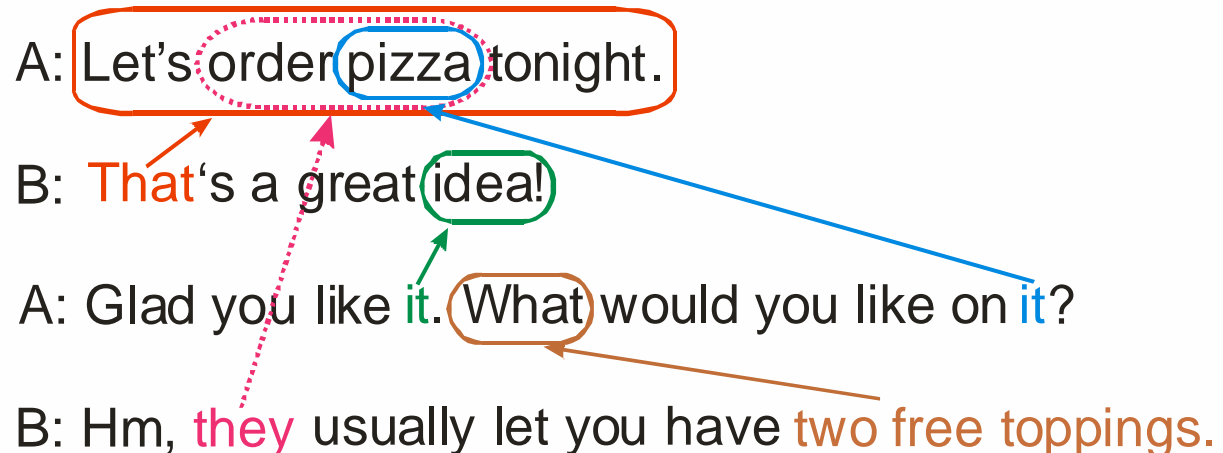
# Related project: literary reputation



- Three authors, very popular in their time; not popular now
  - John Galsworthy, Arnold Bennett, Marie Corelli
- Three authors, unknown or with low sales in their time; popular now
  - Virginia Woolf, Joseph Conrad, D.H. Lawrence
- Extract information about how they were reviewed across time, and map that to their perceived reputation
- Collaboration with Mary Ann Gillies (English, SFU) and Paul McFetridge (Linguistics, SFU)
- Grad students Graham Lyons and Robert Outtrim
- Taboada, Gillies and McFetridge (2006)

# Other (less related) research

- Anaphora and reference, Spanish and English
- Application of Centering Theory
  - Interaction of local and global structure
  - Focus of attention and referring expressions



# References

- Egg, M. & G. Redeker (2007) Underspecified discourse representation. In A. Benz and P. Kühnlein (Eds.), *Constraints in Discourse* (pp. to appear). Amsterdam and Philadelphia: John Benjamins.
- Hatzivassiloglou, V. & K. McKeown (1997) Predicting the semantic orientation of adjectives. In *Proceedings of 35th Meeting of the Association for Computational Linguistics* (pp. 174-181). Madrid, Spain.
- Mann, W.C. & S.A. Thompson (1988) Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8 (3), 243-281.
- Martin, J.R. (2000) Beyond exchange: Appraisal systems in English. In S. Hunston and G. Thompson (Eds.), *Evaluation in Text: Authorial Distance and the Construction of Discourse* (pp. 142-175). Oxford: Oxford University Press.
- Pang, B., L. Lee & S. Vaithyanathan (2002) Thumbs up? Sentiment classification using Machine Learning techniques. In *Proceedings of Conference on Empirical Methods in NLP* (pp. 79-86).
- Soricut, R. & D. Marcu (2003) Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of Human Language Technology and North American Association for Computational Linguistics Conference (HLT-NAACL'03)*. Edmonton, Canada.
- Taboada, M. & J. Grieve (2004) Analyzing appraisal automatically. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07)* (pp. 158-161). Stanford University, CA.
- Taboada, M. & W.C. Mann (2006) Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies*, 8 (3), 423-459.
- Taboada, M., C. Anthony & K. Voll (2006) Creating semantic orientation dictionaries. In *Proceedings of 5th International Conference on Language Resources and Evaluation (LREC)* (pp. 427-432). Genoa, Italy.
- Taboada, M., M.A. Gillies & P. McFetridge (2006) Sentiment classification techniques for tracking literary reputation. In *Proceedings of LREC Workshop, "Towards Computational Models of Literary Analysis"* (pp. 36-43). Genoa, Italy.
- Turney, P. (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of 40th Meeting of the Association for Computational Linguistics* (pp. 417-424).
- Voll, K. & M. Taboada (2007) Not all words are created equal: Extracting semantic orientation as a function of adjective relevance. In *Proceedings of the 20th Australian Joint Conference on Artificial Intelligence* (pp. to appear). Gold Coast, Australia.
- White, P.R.R. (2003) *An Introductory Course in Appraisal Analysis*, from [www.grammatics.com/appraisal](http://www.grammatics.com/appraisal)
- Zaenen, A. & L. Polanyi (2004) Contextual valence shifters. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS-04-07)* (pp. 106-111). Stanford University, CA.

# Thumbs Up or Thumbs Down? Detecting Sentiment and Opinion Automatically

Maite Taboada  
mtaboada@sfu.ca

<http://www.sfu.ca/~mtaboada/>

<http://www.sfu.ca/~mtaboada/research/nserc-project.html>

<https://wiki.sfu.ca/research/sentiment/index.php/ProjectDescription>