Figurative Language Processing: Mining Underlying Knowledge from Social Media

Antonio Reyes and Paolo Rosso
Natural Language Engineering Lab – EliRF
Universidad Politécnica de Valencia
{areyes,prosso}@dsic.upv.es

Outline

- Introduction

- Objective

- Our approach

- Some results

- Final remarks

Introduction (1)

- *Figurative language* refers to second meanings, which are produced by altering the usual referents or concepts.

- Unlike *literal language*, the former takes advantage of linguistic devices, such as metaphor, analogy, ambiguity, irony, and so on, in order to project more complex meanings

- Pragmatic challenge, not only for computers, but for humans as well.

Introduction (2)

- Different linguistic strategies are used to produce the effect; e.g., *ambiguity* and *alliteration* regarding humor; *similes* regarding irony.

  - Children in the back seats of cars cause accidents, but accidents in the back seats of cars cause children.

  - His research is about as ground-breaking as a foam jackhammer.

Challenge

- Figurative language implies information not *grammatically* expressed to be able to decode its underlying meaning: if this information is not unveiled, the real meaning is not accomplished and the figurative effect is lost.

- For instance, a joke. The amusing effect sometimes relies on not given information. If such information is not filled, the result is a bad, or better said, a misunderstood joke.

Objective (1)

- Our goal aims at showing how two specific domains of figurative language —*humor* and *irony*, may be automatically handled by means of considering linguistic devices, such as ambiguity and incongruity, and metalinguistic devices, such as polarity and emotional scenarios.

- We especially focus on analyzing how underlying knowledge, which relies on shallow and deep linguistic layers, may represent relevant information to automatically identify figurative usages of language.

Objective (2)

- In particular, and contrary to the most of the researches which deal with figurative language, we aim at identifying figurative usages regarding language in *social media*.

- Therefore, we do not focus on analyzing *prototypical* jokes nor *literary* examples of irony

- Rather, we try to find patterns in texts whose intrinsic characteristics and targets are different to the ones described in the specialized literature.

- For instance, web comments, product reviews, or tweets.

Motivation (1)

- Humor

- Automatic Recognition

  – Patterns to characterize humor

- Wide range of phenomena underlies humor

  – Cognitive, cultural, social, linguistic …

- Focus on ambiguity

Motivation (2)

- Irony

- Sentiment analysis and opinion mining tasks

  − Hints to represent ironic contents

- Negative and positive opinions are "easily" identifiable.

  − Fine-grained knowledge might be mined

- Like humor, irony cannot be defined as the sum of features nor with a single schema

Humor processing

- Lexical features regarding ambiguity

  - Structural (language models)

  - Morpho-syntactic

  - Syntactic

  - Semantic

Short texts

- Sometimes I need what only you can provide: your absence

- USA is a nation of laws: badly written and randomly enforced.

- He has no enemies, but is intensely disliked by his friends.

- A 16-year-old girl bought herself a very tiny bikini.

- Speak kind words and you will hear wonderful echoes.

- A conservative is a man who believes that nothing should be done for the first time.

- Love is a fire. Whether it will warm your heart or burn down your house, you can never tell.

- Your primary care physician is wearing the pants you gave to Goodwill last month.

Long texts

- A man and his wife were spending the day at the zoo. She was wearing a loose fitting, pink dress, sleeveless with straps. He was wearing his usual jeans and T-shirt. As they walked through the ape exhibit, they passed in front of a large, silverback gorilla. Noticing the wife, the gorilla went crazy. He jumped on the bars, and holding on with one hand and 2 feet he grunted and pounded his chest with his free hand. He was obviously excited at the pretty lady in the pink dress. The husband, noticing the excitement, thought this was funny. He suggested that his wife tease the poor fellow some more by puckering her lips and wiggling her bottom. She played along and the gorilla got even more excited, making noises that would wake the dead. Then the husband suggested that she let one of her straps fall to show a little more skin. She did and the gorilla was about to tear the bars down. Now show your thighs and sort of fan your dress at him, he said. This drove the gorilla absolutely crazy, and he started doing flips. Then the husband grabbed his wife, ripped open the door to the cage, flung her in with the gorilla and slammed the cage door shut. Now. Tell him you have a headache.

Experiments

- Lexical features

  - Perplexity

  - Morphosyntactic ambiguity

  - Syntactic parsing

  - Senses

Perplexity

- Structural ambiguity

- Language models

- Predictability

- Given w, probability to predict $w + 1$

- SRILM toolkit

Ppl

- Perplexity on language modeling

$$PP(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i|w_{i-1})}}$$

PP = The weighted average branching factor of a language. The branching factor of a language is the number of possible next words that can follow any word (Jurafsky).

Morphosyntactic Ambiguity

- POS tagging

  - Different thresholds

- A word can play several syntactic functions

  - Perquè els tontos no entren a la cuina? Perchè ha un pot que diu **sal**!

  - sal = noun, verb

- Triggers of funny interpretations

POS tagging

- It is the process of marking up the words in a text (corpus) as corresponding to a particular part of speech (Wikipedia).

- Viterbi algorithm,

- Constraint Grammar,

- Baum-Welch algorithm (forward-backward algorithm)

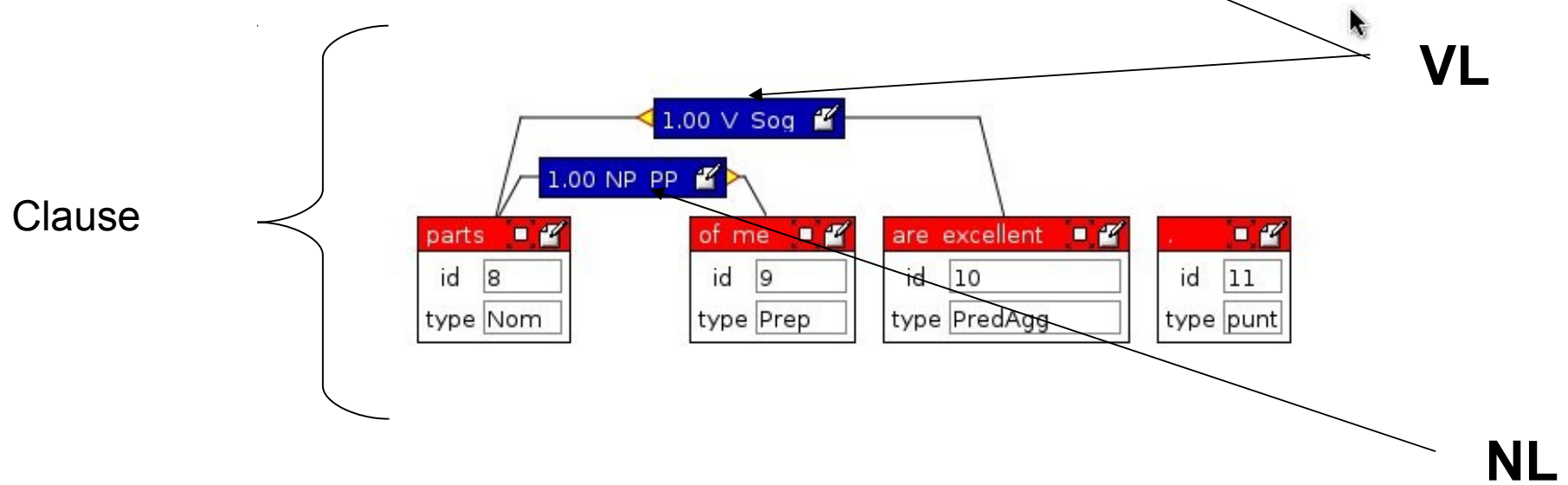- Hidden Markov model and visible Markov model
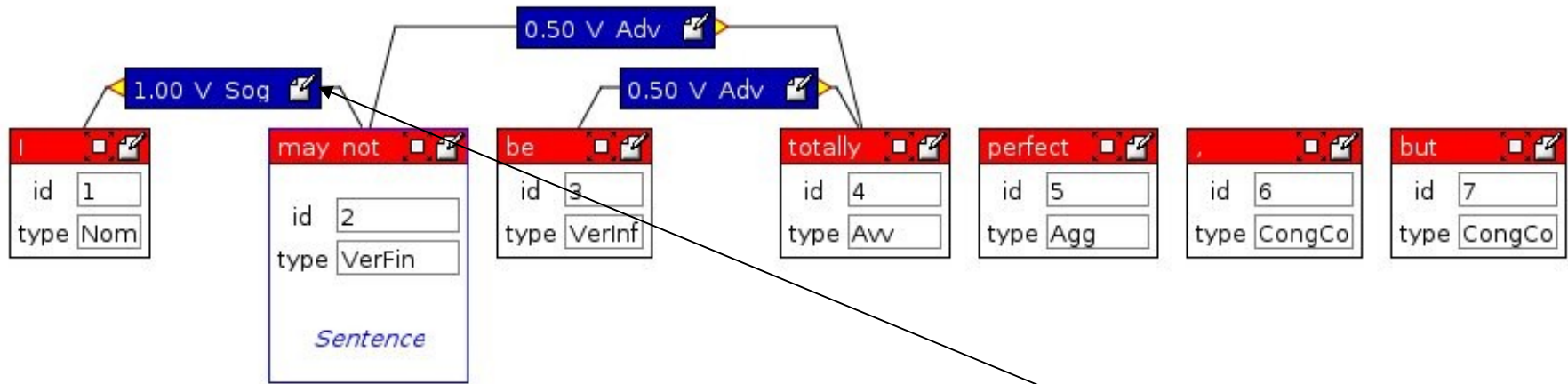
Syntactic Ambiguity

- Syntactic parser
  - The process of analysing a text to determine its grammatical structure with respect to a given grammar (Wikipedia).

- How complex is the syntactic structure?
  - Children in the back seats of cars cause accidents, but accidents in the back seats of cars cause children.
  - Food companies are well aware of the economic implications of reversing the obesity epidemic.

- Sentence Complexity

Formula

- Sentence Complexity

$$SC = \frac{\sum (VL, NL)}{\sum Cl}$$

where *VL* and *NL* are the number of verbal and nominal links respectively, divided by the number of clauses (Cl) (Basili and Zanzotto, 2002).

0.50 V Adv

1.00 V Sog

0.50 V Adv

	may not	be	totally	perfect	,	but
id 1	id 2	id 3	id 4	id 5	id 6	id 7
type Nom	type VerFin	type VerInf	type Avv	type Agg	type CongCo	type CongCo

Sentence

VL

Clause

1.00 V Sog

1.00 NP PP

parts	of me	are excellent	,
id 8	id 9	id 10	id 11
type Nom	type Prep	type PredAgg	type punt

NL

I may not be totally perfect, but parts of me are excellent

Senses Ambiguity

- WordNet (Fellbaum)
- Lexical database
- Lexicon organised in *synsets*

```
>>> N['dog']
dog(n.)
>>> N['dog'].getSenses()
('dog' in {noun: dog, domestic dog, Canis familiaris},
 'dog' in {noun: frump, dog}, 'dog' in {noun: dog},
 'dog' in {noun: cad, bounder, blackguard, dog, hound, heel},
 'dog' in {noun: pawl, detent, click, dog},
 'dog' in {noun: andiron, firedog, dog, dogiron})
```

Senses

- Mean of Senses

    – WordNet

http://wordnet.princeton.edu/

    – Categories N, ADJ, ADV

$$\frac{\sum (W \in C)}{\sum S_c}$$

where W is all the words belonging to a category C, and S is the number of senses for C

Experiments: Blogosphere

- Sense dispersion

- Templates

- Clusters

- Sentiment profiling

- Affective profiling

Sense Dispersion

- Hypernym distance
- WordNet relations

$$\delta(w_s) = \frac{1}{P(|s|, 2)} \sum_{s_i, s_j \in S} d(s_i, s_j)$$

$$\delta_{TOT} = \sum_{w_s \in W} \delta(w_s)$$

http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o7=&o5=&o1=1&o6=&o4=&o3=&r=1&s=brake&i=35&h=10112312312312312231231230   ▾   wordnet

Noun

$$\delta_{\text{TOT}}(w_s) = 7$$

- **S:** (n) **brake** (a restraint used to slow or stop a vehicle)
  - *direct hyponym* / *full hyponym*
  - *direct hypernym* / *inherited hypernym* / *sister term*
    - **S:** (n) restraint, constraint (a device that retards something's motion) *"the car did not have proper restraints fitted"*
      - **S:** (n) device (an instrumentality invented for a particular purpose) *"the device is small enough to wear on your wrist"; "a device intended to conserve water"*
        - **S:** (n) instrumentality, instrumentation (an artifact (or system of artifacts) that is instrumental in accomplishing some end)
          - **S:** (n) artifact, artefact (a man-made object taken as a whole)
            - **S:** (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
              - **S:** (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*

1º common hypernym

$\delta(w_1)= 6$ ⟶
                - **S:** (n) physical entity (an entity that has physical existence)
                  - **S:** (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

  - *derivationally related form*
- **S:** (n) **brake** (any of various ferns of the genus Pteris having pinnately compound leaves and including several popular houseplants)
  - *member holonym*
  - *direct hypernym* / **inherited hypernym** / *sister term*
    - **S:** (n) fern (any of numerous flowerless and seedless vascular plants having true roots from a rhizome and fronds that uncurl upward; reproduce by spores)
      - **S:** (n) pteridophyte, nonflowering plant (plants having vascular tissue and reproducing by spores)
        - **S:** (n) vascular plant, tracheophyte (green plant having a vascular system: ferns, gymnosperms, angiosperms)
          - **S:** (n) plant, flora, plant life ((botany) a living organism lacking the power of locomotion)
            - **S:** (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
              - **S:** (n) living thing, animate thing (a living (or once living) entity)
                - **S:** (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                  - **S:** (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*

1º common hypernym

$\delta(w_2)= 8$ ⟶
                    - **S:** (n) physical entity (an entity that has physical existence)
                      - **S:** (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Keyness

- Bag of keywords

- Keyness value: It compares the word frequencies in a text against their occurrences in a much larger corpus (reference corpus)

- Reference corpus (Google N-grams)

- Values are computed taking into account the Log Likelihood test.

Templates

- Mutual information

- Two or more words produce new meanings

  – *por* – isolated meaning

  – *favor* – isolated meaning

  – **por favor** (template) – new meaning

- High values

Clusters

- Cluto (Karypis)

http://glaros.dtc.umn.edu/gkhome/views/cluto/

- SenseClusters (Kulkarni and Pedersen)

http://search.cpan.org/dist/Text-SenseClusters/

- Sets of common elements

Discriminating Items

Exp.	Space	Cl. Method	Cr. Function	LSA	Order	Cluststop	Clusters
1	Vector	RB/Direct	UPGMA	Yes	Bi/Co	All	2
2	Similarity	Agglo/RBR/Graph	H2	Yes	Uni/Bi	Gap	2
3	Vector	Direct	H2	Not	Co	None	20
4	Vector	RB	I2	Not	Bi	Pk	27

Beyond lexical knowledge

- Humor is NOT only a linguistic phenomenon

- Language is NOT only grammar

- More knowledge to represent more features

Beyond lexical knowledge (2)

God must love stupid people...He made so many of them

Deeper knowledge

- By means of analyzing language is possible to find information related to:

  - Subjective knowledge

    - Sentiments

    - Opinions

    - Emotions

    - Attitudes

    - etc.

Sentiment analysis and humor

- Humor profiles negative polarity

- Necessary to identify what are the elements which trigger the negative information

- Hints to study irony in humor

Irony detection

- Taking into account that humor, in many cases, profiles negative aspects through irony for producing its effect it would be possible to take advantage of this information.

- Some of the features related to humor may be useful for other purposes, for instance, Opinion Mining or Sentiment Analysis.

Theoretical problems

- How to determine irony?

- Is there any pattern?

- Where to find examples?

Practical problems

- Data

- Quantitative and qualitative

- Resources

Two simple examples

A more complicated one

Ten Reasons Gay Marriage is Un-American

1. Being gay is not natural. Real Americans always reject unnatural things like eyeglasses, polyester, and air conditioning.

2. Gay marriage will encourage people to be gay, in the same way that hanging around tall people will make you tall.

3. Legalizing gay marriage will open the door to all kinds of crazy behavior. People may even wish to marry their pets because a dog has legal standing and can sign a marriage contract.

4. Straight marriage has been around a long time and hasn't changed at all; women are still property, blacks still can't marry whites, and divorce is still illegal.

5. Straight marriage will be less meaningful if gay marriage were allowed; the sanctity of Britany Spears' 55-hour just-for-fun marriage would be destroyed.

6. Straight marriages are valid because they produce children. Gay couples, infertile couples, and old people shouldn't be allowed to marry because our orphanages aren't full yet, and the world needs more children.

7. Obviously gay parents will raise gay children, since straight parents only raise straight children.

8. Gay marriage is not supported by religion. In a theocracy like ours, the values of one religion are imposed on the entire country. That's why we have only one religion in America.

9. Children can never succeed without a male and a female role model at home. That's why we as a society expressly forbid single parents to raise children.

10. Gay marriage will change the foundation of society; we could never adapt to new social norms. Just like we haven't adapted to cars, the service-sector economy, or longer life spans.

A theoretical (manual) approach

- Incongruity

  – God must love stupid people. He made so many of them.

- Logic

  – If speed kills, then Windows users may live forever.

- Sarcasm

  – I've got the body of a god ... unfortunately it's Buddha.

- Unexpected situations

  – I'm on a thirty day diet. So far, I have lost 15 days.

Solving practical problems

- Personal examples? Subjective, slow

- Internet: a lot of pages talking about irony, but few examples

- Many images

- Looking for text...

Exploiting Web 2.0

- WWW
  - User-generated tags

- Amazon
  - Viral effect

- Twitter
  - Users hashtags

cpumaster

I blew out my knee making love to my fiance a week before I had to go to the recruiters office.

tags: sex irony humor air force injury knee love [add]

2008-05-08 17:37:40 / Rating: 372.75 /

Lock

I'm the high school dropout with purple hair, yet I'm the only one out of my group of friends with a job.

tags: irony [add]

2008-04-29 12:11:07 / Rating: 239.5 /

Interviewee

1 hour and 45 minutes, a talkative bald man, a crack-head with a faxing issue, and 13 color copies later, I am still hours away from ready for tomorrow.

tags: interview resume irony humor [add]

2008-04-25 10:22:44 / Rating: 204.75 /

OminousGasPains

When I wrenched my back unpacking the massage table, her anniversary gift suddenly became "ours."

tags: humor anniversary injury gift irony [add]

2008-04-15 10:05:44 / Rating: 277 /

Corr... | Inbo... | Trad... | Trad... | Wel... | The ... | site:l... | Expl... | Wel... | EM ... | N-G... | Iron... | Sarc... | Bill ... | de ... | Iron... | A... ✕

http://www.amazon.com/Mountain-Three-Wolf-Short-Sleeve/product-reviews/B002HJ377A/ref=dp_top_cm_cr_acr_txt?ie=UTF8&showVie ▾

Google

amazon.com

Hello, Antonio. We have **recommendations** for you. (**Not Antonio**?)

FREE 2-Day Shipping: **See details**

Antonio's Amazon.com | 🎁 **Today's Deals** | **Gifts & Wish Lists** | **Gift Cards**

Your Account | Help

Shop All Departments | Search | Clothing & Accessories ▾ | | **GO** | 🛒 **Cart** | **Wish List**

Clothing & Accessories | Women's | Men's | Juniors' | Young Men's | Kids' & Baby | Denim Shop | Shoes | Accessories | On Sale Now

Customer Reviews
The Mountain Three Wolf Moon Short Sleeve Tee

1,694 Reviews

5 star: ▬▬▬ (1,361)
4 star: ▬ (119)
3 star: ▬ (50)
2 star: ▬ (35)
1 star: ▬ (129)
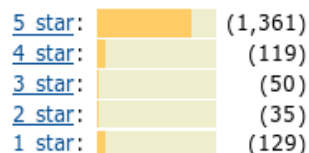
Average Customer Review
⭐⭐⭐⭐½ (**1,694 customer reviews**)

Share your thoughts with other customers

Create your own review

Search Customer Reviews

[   ] **GO!**

☑ Only search this product's reviews

This product

The Mountain Three Wolf Moon Short Sleeve Tee
by The Mountain
$19.95 **$9.94** - **$19.99**

Add to Wish List | ▶ See buying options

A D V E R T I S E M E N T

The most helpful favorable review

18,339 of 18,511 people found the following review helpful:

⭐⭐⭐⭐⭐ **Dual Function Design**
This item has wolves on it which makes it intrinsically sweet and worth 5 stars by itself, but once I tried it on, that's when the magic happened. After checking to ensure that the shirt would properly cover my girth, I walked from my trailer to Wal-mart with the shirt on and was immediately approached by women. The women knew from the wolves on my shirt that I, like a...
Read the full review ›
Published 15 months ago by B. Govern

› See more **5 star**, **4 star** reviews

Vs.

The most helpful critical review

1,864 of 2,022 people found the following review helpful:

⭐☆☆☆☆ **I think some of the benefits are exagerrated**
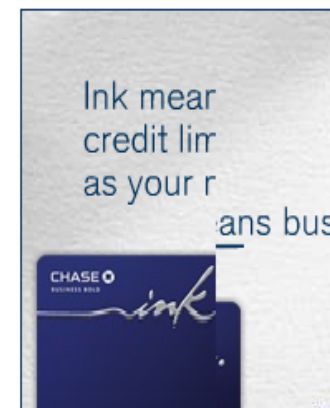So I got this wolf shirt because of, you know, the sweet wolves on it.

However, having owned this shirt for three weeks now and having tried it out in a variety of situations, both formal and informal, I'm beginning to believe that some of the benefits ---- as described by other reviewers ---- are exaggerated. For example, not ONE supermodel has approached...
Read the full review ›
Published 9 months ago by Go Down, Moses

› See more **3 star**, **2 star**, **1 star** reviews

🔍 120%

Corr...   Inbo...   Trad...   Trad...   Wel...   The ...   site:I...   Expl...   Wel...   EM ...   N-G...   Iron...   Sarc...   Bill ...   de ...   Iron...   ⓣ "... ✕   ✚

http://search.twitter.com/search?q=&ands=&phrase=%23irony&ors=&nots=&tag=&lang=en&from=&to=&ref=&near=&within=15&units= | twitter advance search

Show tweets written in:
English

13 minutes ago from *web* · Reply · View Tweet

tomo1340: RT @windowsphone Phone 7 Series: everything you ever wanted to know http://bit.ly/cXy2vR (expand) #WP7 - article clearly states unanswered Qs #irony
26 minutes ago from *HTC Peep* · Reply · View Tweet

Trending topics:

· #nowplaying
· #idontunderstandwhy
· #handsdown
· #WeLoveJonas
· #imgivingup
· Lent
· Ash Wednesday
· Nate Robinson
· Justin Bieber
· Haiti

itsmeARD: RT @DQBond For lent, how many people are going to give up something that the bible already forbids? (eg. sex, drugs, etc). #irony -exactly
28 minutes ago from *Twitterrific* · Reply · View Tweet

Dftya: canceling appointment. too sick to go to the doctor's. #irony
39 minutes ago from *web* · Reply · View Tweet

DigitalKadi: @nickelnm I thought he was going to be like the rest of my kids so I chose Aiden & he is the only mellow one! #irony
about 1 hour ago from *web* · Reply · View Tweet · 💬 Show Conversation

Nifty queries:

· cool filter:links
· "is down"
· movie :)
· "happy hour" near:SF
· #haiku
· "listening to"
· love OR hate
· flight :(

ryanbalas: I think its interesting that a lot of "micro budget" and "no budget" film fests have the most expensive submission fees. #film #irony
about 1 hour ago from *Twitterrific* · Reply · View Tweet

davepetri: Sitting in the Eye Dr's waiting room waiting to be seen. I believe there is irony there somewhere. #irony, #waiting
about 1 hour ago from *TweetDeck* · Reply · View Tweet

MEETorDIE: @gregobrien nice LONG rant on the site. Funny that you don't like the govt to spend money but you don't mind wasting your company's. #irony
about 2 hours ago from *web* · Reply · View Tweet

120%

Corpora

- Amazon

  - ~3,500 ironic reviews

- One-sentence

  - ~ 10,000 ironic statements

- Twitter

  - ~20,000 ironic tweets

Irony (1)

- Literature divides two primaries classes of irony:

  − verbal and situational

- The most of theories agree on the main property of the first one: verbal irony conveys an *opposite* meaning; i.e., a speaker says something that seems to be the opposite of what s/he means.

- By contrast, situational irony is a state of the world which is perceived as ironic; i.e., situations that should not be.

Irony (2)

- Some authors distinguish other types of ironies:

    - dramatic;

    - discursive;

    - tragic;

    - comic;

    - etc.

- We focus on *verbal* irony

Defining verbal irony

- Grice considers that an utterance is ironic if it intentionally violates some conversational maxims.

- Wilson and Sperber assume that verbal irony must be understood as echoic, i.e., as a distinction between use and mention.

- Utsumi suggests an ironic environment, which causes a negative emotional attitude, as a requisite to consider an utterance as ironic.

- Same underlying concept of opposition

- Their computational formalization is quite complex.

First features

- N-grams

- Morpho-grams

- Funny profiling (Humor-specific features)

- Positive/Negative profiling (Polarity)

- Affective profiling

Objective

- Analyzing irony in social media

- There is a "general" idea about what irony is

- Gathering the most discriminating features to represent irony

- Hints about how to automatically deal with irony

Data sets

- Positive data:

  - Amazon

  - User reviews considered as ironic ones by mass & social media (Youtube, BBC, ABC ...)

  - 6 products reviewed

  - 2,861 reviews

Data sets

- Negative data:

  – Amazon (users reviews)

  – SlashDot (web comments labelled with the tag *funny*)

  – TripAdvisor (users reviews on hotels)

  – 3,000 documents per set

- Final corpus contains 11,861 documents

N-grams

- Find frequent sequences of recurrent words which could denote irony

- Order 2 – 7

- Jaccard distance

- TFIDF

Morpho-grams

- Word representation: more abstract

- POS tags instead of words

- Order 2 – 7

- Statistical Substring Reduction

- TFIDF

Funny profiling

- Relevance of some features related to humor

- 3 categories:

  - Sexual data (sex, gay, lesbian)

  - Social relationships (woman, kid, friend)

  - Keyness (google n-grams as reference model)

Positve/Negative profiling

- Importance of negative information to represent ironic contents

- Macquarie Semantic Orientation Lexicon (MSOL)
  - 20,299 items

- Negative category:
  - 22,384 items

Affective profiling

- Affective (cognitive, emotional, psychological …) info is represented by words

- Two representations

  - WordNet-Affect

    - 11 classes

  - Dictionary of affect in language

    - Pleasantness rank

WordNet Affect

- 11 classes
- Based on WordNet relations
- Automatically retrieved
  - emo = emotion (e.g. noun "anger#1", verb "fear#1")
  - moo = mood (e.g. noun "animosity#1", adjective "amiable#1")
  - tra = trait (e.g. noun "aggressiveness#1", adjective "competitive#1")
  - cog = cognitive state (e.g. noun "confusion#2", adjective "dazed#2")
  - phy = physical state (e.g. noun "illness#1", adjective "all_in#1")
  - eds = edonic signal (e.g. noun "hurt#3", noun "suffering#4")
  - sit = emotion-eliciting situation (e.g. noun "awkwardness#3", adjective "out_of_danger#1")
  - res = emotional response (e.g. noun "cold_sweat#1", verb "tremble#2")
  - beh = behaviour (e.g. noun "offense#1", adjective "inhibited#1")
  - att = attitude (e.g. noun "intolerance#1", noun "defensive#1")
  - sen = sensation (e.g. noun "coldness#1", verb "feel#3")

Whissell's dictionary

- ~9,000 words

- Scores for 3 features:

  – Pleasantness

  – Activation

  – Imagery

- Abnormal: 1.0000   2.0000   2.4

- Good: 2.7500   1.9167   1.0

- Flower: 2.7500   1.0714   3.0

Document representation

- Feature vectors

- Representativeness threshold

- Documents normalized

Classification

- Binary classifiers

  – Amazon (+) vs. Amazon (-)

  – Amazon (+) vs. Slashdot (-)

  – Amazon (+) vs. TripAdvisor (-)

- Bayes, SVM, Decision tree

- 10-fold cross validation

Results

- Worst feature: N-grams

- Morpho-grams enhance accuracy (amazon vs. tripAdvisor the best result)

- Pleasantness rank seems to discriminate well

Classification accuracy

	TREE	Pr	Re	F-M	SVM	Pr	Re	F-M	BAYES	Pr	Re	F-M
amazon.allFeatures	73,53%	0,73	0,73	0,73	74,54%	0,75	0,74	0,74	72,18%	0,72	0,72	0,72
amazon.allFeatures.filters	74,14%	0,74	0,74	0,74	75,75%	0,76	0,76	0,76	74,63%	0,75	0,75	0,75
slash.allFeatures	75,76%	0,76	0,76	0,76	76,75%	0,78	0,77	0,77	75,19%	0,77	0,75	0,75
slash.allFeatures.filters	75,12%	0,75	0,75	0,75	73,35%	0,74	0,73	0,73	74,61%	0,75	0,75	0,74
trip.allFeatures	88,38%	0,88	0,88	0,88	89,63%	0,90	0,90	0,90	87,17%	0,87	0,87	0,87
trip.allFeatures.filters	89,05%	0,89	0,89	0,89	89,04%	0,89	0,89	0,89	87,77%	0,88	0,88	0,88

Preliminary results

- No formal patterns.

- N-grams didn't work

- Interesting morpho-syntactic sequences

- Funny and affective features seem to be interesting

- Negative polarity appears quite often in positive data

First conclusions

- Considering the task: good results

- Knowledge for many applications

- Improve features

- Take into account context

- More experiments

- New data

- More problems

Fine-grained features

- 3D and 2D features

    – Signatures

    – Unexpectedness

    – Style

    – Emotional Scenarios

Signatures (2D)

- This feature focuses on exploring irony in terms of underlying linguistic marks.

  1. Typographical marks (punctuation or emoticons)

  2. Discursive marks (terms related to opposition)

- Formally, signatures are textual elements which put in focus certain information.

Unexpectedness (3D)

1. A mean to represent both temporal and contextual imbalances (or incongruity) in the ironic documents.

2. Temporal (degree of opposition in a same document regarding the information profiled in present and past)

   - Divergences related only to verbs

1. Contextual (inconsistencies within a context)

   - Similarity of concepts taking into account their semantic relatedness (Resnik, Leacock & Chodorow).

Style (3D)

- Distinctive manner of expression (fingerprint that determines intrinsic characteristics)

1. Character n-grams (*c-grams*). Order 3 – 5
   - Frequent sequences of morphological information

1. Skip n-grams (*s-grams*). Skips = 2 and 3
   - Entire words which allow arbitrary gaps

1. Polarity s-grams (*ps-sgrams*). Ibid.
   - Sequences of abstract representations on the basis of the s-grams

Emotional Scenarios (1)

- A manner of representing information regarding contents beyond grammar, and beyond positive or negative polarity.

- Characterizing irony in terms of elements which symbolize abstract contents such as sentiments, attitudes, feelings, moods, and so on, in order to define a schema of favorable and unfavorable contexts to express irony.

Emotional Scenarios (3D)

1. Activation: degree of response, either passive or active, that humans have under an emotional state (e.g., burning is more active than basic).

2. Imagery: how difficult is to form a mental picture of a given word (e.g., never is more difficult to be mentally depicted than alcoholic).

3. Pleasantness: degree of pleasure produced by words (e.g., love is more pleasant than money).

Some experiments

- Classification task

- Corpus

  − Twitter

  − Five sets: humor, irony, politics, technology, general

- 10.000 documents per set

- 70% training and 30% test

Results

	Accuracy	Precision	Recall	F-measure
Irony *vs.* Humor	84.33	0.80	0.91	0.85
Irony *vs.* Politics	91.97	0.90	0.95	0.92
Irony *vs.* Technology	88.97	0.87	0.91	0.89
Irony *vs.* General	70.12	0.78	0.56	0.65

Features: Signatures, Unexpectedness, Style, Emotional Scenarios

Final remarks

- Set of features to represent different kinds of patterns from a text regarding figurative language

- They intended to symbolize low and high level properties of figurative language on the basis of formal linguistic elements.

- No single feature is distinctly humorous or ironic, but all of them together provide a useful linguistic inventory for detecting these types of figurative devices at textual level.

- Results are encouraging

Further experiments on Irony

- Negation
  - Negative attitude

- Frames

- Triggers

- Etc.

Coarse or fine-grained?

- Irony -- sarcasm -- satire (*humor* tends to rely all of them)

  - My mother never saw the irony in calling me a son-of-a-bitch.

  - If you find it hard to laugh at yourself, I would be happy to do it for you.

  - Let's pray that the human race never escapes from Earth to spread its iniquity elsewhere.