Words, Meanings and Emotions

Rada Mihalcea and Carlo Strapparava

University of North Texas
FBK-Irst - Istituto per la Ricerca Scientifica e Tecnologica
rada@cs.unt.edu, strappa@itc.it
Goal

- Affective analysis of text is a relatively new area of research
- Important for many NLP applications
  - Opinion mining
  - Market analysis
  - Affective user interfaces
  - E-learning environments
- Goal of the tutorial: overview techniques for affective content detection and generation
Outline

1. Computational Humor
   - Humor generation (Carlo)
   - Humor recognition (Rada)

2. Affective Text
   - Lexical resources (Carlo)
   - Annotation of emotions in text (Rada)
   - Dancing with words (Carlo)
   - Emotions in blogs (Rada)
Society needs humour

- Humor is a powerful generator of emotions
- It has an impact on people's psychological state, directs their attention, influences the processes of memorization and of decision-making (i.e. companies hire ‘humour consultants’)
- E.g. the persuasive effect of humor and emotions is well known and widely employed in advertising.
- Computational Humour can deliver something useful
- Deep modelling of humour in all its facets is not for the near future: humour is AI-complete
- Complete modelling of humour processes is not always required
- CH leads to falsifiable theories: can be tested on human subjects

⇒ Humour is infectious: contagious laughter in Tanganyika, started in a group of schoolgirls and rapidly rose to epidemic proportions, infecting adjacent communities. It required the closing of the schools and it lasted for six months.
Is computational humour realistic?

- Deep modelling of humour in all its facets is not for near future
- But not always complete modelling of humour processes is required
  - E.g. wordplays, lexicon-based semantic opposition, ambiguity, ...

A bit of Marxism ..... in the sense of Marx Brothers :-)

Mrs Teasdale: This is a gala day for you.
Firefly (Groucho): Well, a gal a day is enough for me.
  I don’t think I could handle any more.
Computational humour for edutainment and IT

- To provide comic relief/reward
- To stimulate the attention
- To favor long-term memorization
- To enhance learning experience (positive feelings towards learning when humor is included)
- To stimulate creativity
Theories of humor

- Cognitive (incongruity, contrast)
  Focus: *stimulus*

- Social (superiority, hostility, derision, disparagement)
  Focus: *interpersonal effects*

- Psychoanalytical (relief, release, liberation, sublimation)
  Focus: *audience’s reaction*
### Individual differences

**Personality studies** (see W. Ruch)

<table>
<thead>
<tr>
<th>Conservative Attitudes</th>
<th>Liberal/Radical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intolerance of minorities, militarism, religious fundamentalism, education, traditional family ideology, capitalistic attitudes, property/money, law and order attitude, punitiveness, conventional values</td>
<td>Conservative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>General Inhibitedness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superego strength, inhibition of aggression, self-control, rigidity, need for order, anthropomistic, sexually not permissive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Uncertainty Avoidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intolerance of ambiguity, avoiding new and complex experience, prefers simplicity and symmetry, conventional vocational interests, liking of simple, non fantastic art</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Depressivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressiveness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social Desirability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frankness, &quot;lying&quot;, low frankness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Fluid&quot; intelligence, speed of closure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefers complexity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Openness to Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefers new experience</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonconformism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not obedient, low social desirability, &quot;lying&quot;, frank</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Depressivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social Desirability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frank</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger</td>
</tr>
</tbody>
</table>

- **Incongruity-Resolution Humor (INC-RES)**
  - Low appreciation characterized by
  - Conservative Attitudes: Inhibitedness, Uncertainty Avoidance, Depressivity, Social Desirability, Intelligence
  - Openness to Experience: Intelligence, Complexity

- **Nonsense Humor (NON)**
  - High appreciation characterized by
  - Depressivity, Openness to Experience, Social Desirability, Intelligence, Complexity
  - Nonconformism, Age
Individual differences (2)

**Sexual Libido**
- Low appreciation: weak
  - sexual desire, experience and activity, positive attitude to sex, hedonistic and pleasure-seeking, not prudish, easily excited
- High appreciation: strong
  -

**Tough-mindedness**
- Low: tender-minded
  - tough-mindedness, masculinity, dominance, disinhibition, "undersocialized", need for power, technical interests, low ranking of values freedom, equality, world at peace
- High: tough-minded
  -

**Extraversion**
- Low: introvert
  - activity, sociability, positive emotion
- High: extravert
  -

**Gender**
- Low: female
  - biological, psychological
- High: male
  -

**General Aversiveness**
- Low aversiveness: characterized by
  -
- High aversiveness: characterized by
  - neuroticism, anxiety, depressivity, nervousness, guilt proneness, low ego strength, sexual dissatisfaction, sexual prudishness

**Emotional Lability**
- Low: characterized by
  -
- High: characterized by
  - neuroticism, anxiety, depressivity, nervousness, guilt proneness, low ego strength, sexual dissatisfaction, sexual prudishness

**Tender-mindedness**
- Low: characterized by
  -
- High: characterized by
  - tender-mindedness, intraceptive (social, religious, and aesthetic) value orientation, low technical interests, disinhibition, moral and interpersonal values high, low competence or self-actualization values
Requirements for a successful humorous system

- recognize situations appropriate for humor
- choose a suitable kind of humor for the situation
- generate an appropriately humorous output

incongruity $\rightarrow$ ambiguity and NLP
Work on computational humour

- Research on linguistics and pragmatics of humor [e.g. Attardo and Raskin]
- Speculative writings in AI [e.g. Minsky, Hofstadter]
- Some efforts on building computational humor prototypes. For example:
  - Humour Production
    - JAPE [Binsted & Ritchie] It generates punning riddles, from a linguistic model of pun schemata, e.g. “What do you call a murderer with fiber? A cereal killer”
  - Humour Recognition
    - [Mihalcea & Strapparava 2005] investigated machine learning techniques to distinguish between humorous and non-humorous text
HAHA-Acronym

- HAHA-Acronym has been a Future Emerging Technologies (FET) European project ([http://haha.itc.it](http://haha.itc.it))
- Goal: realization of an acronym re-analyzer and generator as proof of concept in a focalized but non restricted context
- various existing resources for NLP adapted for humor + some strategies for yielding humorous output

⇒ O. Stock & C. Strapparava “Getting serious about the development of computational humor” Proceedings of IJCAI 2003
HAHAcronym Resources

- Lexicon (full English lexicon)
- Lexical knowledge base (WordNet Domains)
- Pronunciation dictionary
- Parser and grammar
- Algorithms (for humour effects)
- Slanting dictionary
- ...
WordNet as a lexical knowledge base

- WordNet is an on-line lexical reference system whose design is inspired by *psycholinguistic* theories of human lexical memory.
- Synonym sets, representing underlying concepts (~100,000). Different relations link the synonym sets.
- IRLST extensions
  - Multilinguality (synset-aligned)
  - Domain labels on synsets (e.g. *Medicine*, *Architecture*, *Sport*)
250 Domain labels collected from dictionaries

Four level hierarchy (Dewey Decimal Classification)
Domain labels annotation in WordNet

- Integrate taxonomic and domain oriented information
  - Cross hierarchy
    - doctor#2 [Medicine] --> person#1
    - hospital#1 [Medicine] --> location#1
  - Cross category relations: operate#3 [Medicine]
  - Cross language information

- Reduce polysemy
Use of domain label annotations

- Theories of humour suggest:
  - incongruity, semantic field opposition, apparent contradiction, absurdity

- We have defined:
  - an independent structure of domain opposition i.e. Religion vs. Technology, Sex vs. Religion, etc...
  - algorithms to detect semantic mismatches between word meaning and sentence meaning (i.e. acronym and its expansion)
Bipolar adjective structure

- Swift
  - Alacritous
  - Prompt
  - Quick
  - Rapid

- Fast
  - Similar to

- Dilatory
  - Sluggish
  - Tardy
  - Laggard

- Slow
  - Leisurely
Rhymes

- The HAHA acronym prototype takes into account the rhyme structure of words
- CMU pronouncing dictionary, reorganized with a suitable indexing
- Over 125,000 words and their transcriptions
- Mappings from words to their pronunciations in the given phoneme set
Slanting Dictionary

- A collection of hyperbolic, epistemic, emotive adjectives, adverbs and nouns
  - Ex. abnormally, abstrusely, adorably, exceptionally, exorbitantly, exponentially, extraordinarily, voraciously, weirdly, wonderfully ...

- Useful when it is not possible to exploit other more meaningful strategies
Heuristics in HAHAcronym

- **Using WordNet**
  - *Semantic field opposition*: e.g. Technology vs. Religion
  - *Antonymy* (for adjectives): e.g. “high” vs. “humble”
  - *Exploiting the hierarchy*:
    - e.g. detecting geographic names/adjectives
    - hypernyms/hyponyms in the generation phase
Heuristics in HAHacronym (2)

- Using general lexical resources
  - Strict rhyme and “light” rhyme
  - slanting dictionaries

- Syntactic strategies
  - e.g. keep the main head fixed
Acronym re-analysis

1. Acronym parsing and construction of logical form
2. Choice of what to keep unchanged
3. Look up for possible substitutions, e.g. exploiting semantic field oppositions
4. Granting phonological analogy and rhyme
5. Exploitation of WordNet antonymy clustering
6. Use of slanting dictionary as a last resource
Acronym re-analysis: the architecture
Examples: re-analysis

"Massachusetts Institute of Technology"

- Slanting dictionary
- Keep the head of NP fixed
- Semantic field opposition

"Mythical Institute of Theology"

FBI - Federal Bureau of Investigation
=> Feral Bureau of Intimidation

GPD - Gross Domestic Product
=> Godless Dietetic Product

PDA - Personal Digital Assistant
=> Penitential Demoniacal Assistant
Acronym generation

- Additional constraint: resulting acronyms to be words of the dictionary (APPLE is good, IBM not)
- Input: WN synsets and some minimal structural indication (e.g. the semantic head)
- Primary strategy: consider as potential acronyms words that are in ironic relation with input concepts
- Impose a syntactic structure and expand the acronym preserving coherence among semantic fields
An example: generation

“fast” “processor”

Select the synsets:
E.g. processor in the sense of CPU

Looking for a funny opposite attribute.
This may be a proposal for the acronym

INERT

Establish a syntactic structure and fill it

Inconclusive Non_parallel Electronic_equipment for Rapid Toggle
Choices

- Ranking of possible re-analyses so that the funnier ones appear at the top
- System is flexible and novel strategies can be added
- Ranking a priori is easy
- Ranking a posteriori is difficult and involves modeling humor appreciation
Evaluation

- Success thresholds stated in the project proposal
- Evaluation carried out by Salvatore Attardo at Youngstown University
- A panel of 40 university students, all native speakers of English, homogeneous for age, and mixed for gender and race
Evaluation results

- About 80 reanalyzed and 80 generated acronyms were tested.
- Also a test with randomly generated acronyms (only syntactic rules were operational)

<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Successful</th>
<th>Success Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>52.87%</td>
<td>45%</td>
</tr>
<tr>
<td>Re-analysis</td>
<td>69.81%</td>
<td>60%</td>
</tr>
<tr>
<td>Random re-analysis</td>
<td>7.69%</td>
<td></td>
</tr>
</tbody>
</table>
HAHAAcronym competes with humans

- HAHAAcronym participated in a contest about (human) production of acronyms, organized in by RAI, the Italian National Broadcasting Service.
- The system won the jury’s special prize!
Possible developments of practical impact

- Educational software for children: word-meanings exploration;
- A system that uses humor as means to promote products and to get user's attention in electronic commerce;
- An explorative environment for advertising professionals (e.g. “thirst come, thirst served” for a soft drink);
- A names generator for products and merchandise
More acronyms

- NATO - *North Atlantic Treaty Organization*
  Noisy Anglophilic Torpidity Organization

- NSF - *National Science Foundation*
  National Somnolence Foundation
  National Science Flirtation
  National Somnolence Fornication

- AAA - *American Automobile Association*
  Antediluvian Automobile Association

- IBM - *International Business Machine*
  Illusional Baroqueness Machine

- GSMC - *Global System for Mobile Communication*
  Gastronomical System for Male Consolation
ITS - Intelligent Tutoring Systems

- Impertinent Tutoring Systems
- Indecent Toying Systems

"intelligent" "tutoring"

- FAINT

Folksy Acritical Instruction for Nescience Teaching

- NAÏVE

Negligent At-large Instruction for Vulnerable Extracurricular-activity

- VOID

Visceral Overflowing Instruction for Degree-program
HAHA Acronym at AI conferences

- AAAI - American Association for Artificial Intelligence
  => Antediluvian Association for Artificial Imprudence

- IJCAI - International Joint Conference on Artificial Intelligence
  => Irrational Joint Conference on Antenuptial Intemperance
Outline

1. Computational Humor
   - Humor generation (Carlo)
   - Humor recognition (Rada)

2. Affective Text
   - Lexical resources (Carlo)
   - Annotation of emotions in text (Rada)
   - Dancing with words (Carlo)
   - Emotions in blogs (Rada)
Questions:

1. Can we build very large data sets of humorous texts?
2. Are humorous and serious texts separable?
   - Can we automatically distinguish between humorous and non-humorous texts?
   - Does this hold for different data sets?
3. What are the distinctive features of humour?
   - Can we identify salient features of verbal humour?
   - Do they hold across data sets?
4. Can humour improve human-computer interaction?
Data for humour recognition

- Required to *learn* and *test* models of humour
- Positive examples = humorous text
- Negative examples = non-humorous text
- Desiderata:
  - Large data sets
    - To test variation of performance with data
  - Humorous text should differ only in comic effect – force classifiers to identify humour-specific features
    - Chose non-humorous data similar in content and style with humorous data
  - Different data sets
    - To test consistency
Humorous data (1/2)

- Focus on two types of humour
  - One-liners
    - “He who smiles in a crisis has found someone to blame”
      - Short sentence, simple syntax
      - Deliberate use of rhetoric devices (alliteration, rhyme)
      - Frequent use of creative language
      - Comic effect
  - How to get 10,000+ one-liners
    - Websites or mailing lists typically include no more than 10 – 100 one-liners
  - Web-based bootstrapping
    - Start with a few manually selected seeds
    - Identify a list of Web pages including at least one seed
    - Parse Web pages and find new one-liners
    - Repeat
Web-based bootstrapping

- Bootstrapping
  - Risks addition of noise in the data
  - Requires constraints to guide bootstrapping process

- Thematic constraint (1)
  - Webpage content
  - Look for indicators of humour in the URL
    - oneliner, one-liner, humor, joke

- Stylistic constraint (2)
  - Exploit HTML structure to identify enumerations including the seed one-liner
    - `<li>` Take my advice, I don’t use it anyway
    - `<li>` 42.7 percent of all statistics are made up on the spot.
Web-based bootstrapping

- Two bootstrapping iterations = 24,000 one-liners
- Remove duplicates
  - String similarity based on longest common subsequence
- Final set of 16,000 one-liners
- Verification:
  - Random sample of 200 one-liners
  - 18 noisy entries = 9% noise
Humorous data (2/2)

- Daily news stories from: “The Onion”
  - “the best source of humour out there” (Jeff Grienfield, CNN)

  - Canadian Prime Minister Jean Chrétien and Indian President Abdul Kalam held a subdued press conference in the Canadian Capitol building Monday to announce that the two nations have peacefully and sheepishly resolved a dispute over their common border. "We are - well, I guess proud isn't the word - relieved, I suppose, to restore friendly relations with India after the regrettable dispute over the exact coordinates of our shared border," said Chrétien, who refused to meet reporters' eyes as he nervously crumpled his prepared statement. "The border that, er... Well, I guess it turns out that we don't share a border after all." Chrétien then officially withdrew his country's demand that India hand over a 20-mile-wide stretch of land that was to have served as a demilitarized buffer zone between the two nations."

- 1,125 news articles from August 2005 – March 2006
  - 1,000-10,000 characters
Serious data

- EVERYWHERE! (almost)
- Data similar in structure and composition to the humorous text
  - Make the humour-recognition task more difficult (& real)
  - Allow the classifiers to identify humour-specific features

- For the one-liners:
  - Sentences of 10 – 15 words
  - Similar to one-liners with respect to creativity and intent
  - Mix of *Reuters* titles, *proverbs*, *British National Corpus*, sentences from *Open Mind Common Sense*
Serious data

- Reuters titles
  - Phrased to catch the readers attention
  - Reuters newswire 1996 – 1997
  - “Silver fixes at two-month high, but gold lags”.

- Proverbs
  - From online proverb collection
  - Memorable sayings, considered true by many people
  - “Beauty is in the eye of the beholder”.

- Text
  - British National Corpus
  - Most similar sentences, using vectorial similarity with tf.idf weighting
  - “The train arrives three minutes early”.
Serious data

- For the news articles:
  - Documents with a length of 1,000-10,000 characters
  - Mix of *Los Angeles Times, Foreign Broadcast Information Service, British National Corpus*
Learning to recognize humour

Hypothesis: “We can apply machine learning techniques to distinguish between humorous and non-humorous text”

- Data
  - Positive / negative examples

- Features
  - Content / Style

- Learning algorithms
  - Naïve Bayes / SVM / …
Features

- Style:
  - Rhetorical devices
  - Attention-catching sounds

- Content:
  - Specific vocabulary
Stylistic features

- Inspired from linguistic theories of humour
  - (Attardo 1994)
- Focus on features that can be implemented with current resources
  - Alliteration
  - Antonymy
  - Slang
Alliteration

- Phonetic properties: alliteration, word repetition, rhyme, producing a comic effect
  - Similar devices are used in wordplay, newspaper headlines, advertisement

- Examples
  - “Veni, Vidi, Visa: I came, I saw, I did a little shopping”.
  - “Infants don’t enjoy infancy like adults do adultery”.

- Identify and count alliteration/rhyme chains using the CMU pronunciation dictionary
Antonymy

- Humor often relies on incongruity and contradiction
  - Antonymy is a form of incongruity that can be identified

- Examples:
  - “A clean desk is a sign of a cluttered desk drawer”.
  - “Always try to be modest and be proud of it”.

- Identify antonyms using WordNet:
  - Nouns, verbs, adjectives, adverbs
Adult slang

- A popular form of humour
- Can be identified through the detection of sexual-oriented vocabulary
- Examples:
  - “The sex was so good that even the neighbors had a cigarette”
  - “Artificial insemination: procreation without recreation”
- Use WordNet – Domains to build a lexicon with all synsets marked with the domain “sexuality”
  - Remove words with high polysemy (> 3)
Experiments

- Data set: 16,000 one-liners + 16,000 “serious” sentences
- Apply the stylistic features to humour-recognition
- Features act as heuristics
  - Require a threshold
- Learn a decision tree using 1000 (x 2) positive and negative examples
- Evaluate on remaining 15,000 (x 2) examples
- 10 trials
## Results

<table>
<thead>
<tr>
<th>Heuristic</th>
<th>Oneliners Reuters</th>
<th>Oneliners BNC</th>
<th>Oneliners Proverbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alliteration</td>
<td>74.31%</td>
<td>59.34%</td>
<td>53.30%</td>
</tr>
<tr>
<td>Antonymy</td>
<td>55.65%</td>
<td>51.40%</td>
<td>50.51%</td>
</tr>
<tr>
<td>Adult slang</td>
<td>52.74%</td>
<td>52.39%</td>
<td>50.74%</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td><strong>76.73%</strong></td>
<td><strong>60.63%</strong></td>
<td><strong>53.71%</strong></td>
</tr>
</tbody>
</table>

### Notes:
- A combination of features provides the best results
- Alliteration is the most useful feature
- Reuters titles are the most different with respect to one-liners
- Proverbs are the most similar
Context-based features

- Formulate humour recognition as a text classification problem
  - Data
    - Positive (humorous) / negative (serious) examples
  - Features
    - N-grams
  - Learning algorithms
    - Naïve Bayes / SVM
    - 10-fold cross-validation
Classification results

<table>
<thead>
<tr>
<th>Classifier</th>
<th>One-liners</th>
<th>News articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>79.69%</td>
<td>88.00%</td>
</tr>
<tr>
<td>SVM</td>
<td>79.23%</td>
<td>96.80%</td>
</tr>
</tbody>
</table>

- Significant improvement over the 50% baseline
- Better discrimination for news stories – longer size
Characteristics of humour

- What are the distinctive features of humour?
  - Identify the most salient features for humorous text
  - Classify these features into categories

- Feature list
  - Start with the score generated by the Naïve Bayes classifier
  - \( \text{Humorous score} = \frac{\text{score in humorous text}}{\text{total score}} \)
    - Score close to 1 => features specific to the humorous text
    - Score close to 0 => features specific to the non-humorous text
  - Extract the 1,500 most discriminatory features
    - Occurring at least 100 times in the entire corpus
Characteristics of verbal humour

- Observed by analyzing the features extracted from the one-liners

- Human-centric vocabulary
  - *you, I, man, woman, guy*
    - *you* occurs in more than 25% of the one-liners
    - “You can always find what you are not looking for.”

- Negation
  - *doesn’t, isn’t, don’t*
    - “If at first you don’t succeed, skydiving is not for you.”

- Negative orientation
  - words with negative orientation: *bad, illegal, wrong*
    - “When everything comes your way, you are in the wrong lane.”
Characteristics of verbal humour

- **Professional communities**
  - *lawyers, programmers, policemen*
    - “It was so cold last winter, that I saw a lawyer with his hands in his own pockets.”

- **Human “weakness”**
  - *ignorance, stupidity, trouble, beer, drink, lie*
    - “Only adults have trouble with child-proof bottles.”
Two main features

- Human centeredness
  - Human-centric vocabulary
  - Professional communities
  - Human weakness

- Polarity orientation
  - Negation
  - Negative orientation
  - Human weakness
Human centeredness

- Measure the weight of the most salient features with respect to a semantic class
  - Score of semantic class = sum of the corresponding features normalized with the size of the class
  - E.g. I (0.88), me (0.65), myself (0.55) => 0.69

- Top 1,500 most discriminatory features

- Four semantic classes
  - persons: WordNet hierarchy subsumed by person#n#1
  - social groups: hierarchy subsumed by social_group#n#1
  - social relations: hierarchies of relative#n#1 and relationship#n#1
  - personal pronouns
Human centeredness: One-liners

<table>
<thead>
<tr>
<th></th>
<th>Humour</th>
<th>Non-humour</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>68</td>
<td>32</td>
</tr>
<tr>
<td>SR</td>
<td>63.1</td>
<td>36.9</td>
</tr>
<tr>
<td>SG</td>
<td>50.2</td>
<td>49.8</td>
</tr>
<tr>
<td>P</td>
<td>53</td>
<td>47</td>
</tr>
</tbody>
</table>
Human centeredness: News articles

![Bar chart showing human centeredness in news articles with categories PP, SR, SG, and P, comparing humour and non-humour]
Polarity orientation

- Measure the orientation of humorous text
- Tool for semantic analysis
  - 10,662 positive/negative short fragments (Pang & Lee)
  - Naïve Bayes classifier
  - 78.15% 10-fold cross validation
- Reference: the orientation of non-humorous texts
  - 56% of the non-humorous sentences are labeled as negative
  - 67% of the non-humorous news-articles are negative
Polarity orientation of humour

![Bar chart showing the polarity orientation of humour in one-liners and news articles]

- **One-liners**
  - Negative: 71.75%
  - Positive: 28.25%

- **News articles**
  - Negative: 90.05%
  - Positive: 9.95%
Humour for computer applications

- Find the most appropriate joke for a given context
  - Text semantic similarity
  - LSA, WordNet-based

- Determine the affective orientation of text
  - Avoid the use of humorous text for negative/sad situations
  - Automatic classification of affect
Fun Email

- Add humorous one-liners to email
  - Modification of Squirrel Mail email client
  - find the text’s semantic orientation and ignore the email if adding humor would be inappropriate
    - Automatic classification of text as happy/sad
  - extract the last 30 percent of text from the email body
    - similarity is computed with respect to the topic of the last part of the email
  - compare the email’s LSA vector with those of the one-liners, and identify the most similar one-liner
HLT-NAACL 2007 Computational Approaches to Figurative Language: Call for Papers

Figurative language, such as metaphor, metonymy, idioms, personification, simile among others, is in abundance in natural discourse. It is an effective apparatus to heighten effect and convey various meanings, such as humor, irony, sarcasm, affection, etc. Figurative language can be found not only in fiction, but also in everyday speech, newspaper articles, research papers, and even technical reports.

[....]

Important Dates:

Paper submission deadline: January 18, 2007
Notification of acceptance for papers: February 22, 2007
Camera ready papers due: March 1, 2007
Workshop Date: April 26, 2007

You will be six months behind schedule on your first day
Fun Email

- 10 emails covering different topics
- Add motto
  - Version 1: basic (none)
  - Version 2: random one-liner addition
  - Version 3: contextualized one-liner addition
- 13 users ranked the emails on a 10-point scale on four dimensions:
  - entertainment (the email was entertaining)
  - appropriateness (the motto was appropriate)
  - intelligence (the email program behaved intelligently)
  - adoption (I would use the email program myself)
Fun Email
Questions (revisited)

1. Can we build very large data sets of humorous texts?

2. Are humorous and serious texts separable?
   - Can we automatically distinguish between humorous and non-humorous texts?
   - Does this hold for different data sets?

3. What are the distinctive features of humour?
   - Can we identify salient features of verbal humour?
   - Do they hold across data sets?

4. Can humour improve human-computer interaction?
Question 1

Can we build very large data sets of humorous texts?

- Automatic Web-based bootstrapping of collections of humorous text
  - Thematic and stylistic constraints
  - Very large collection of one-liners
- Crawling of existing collections
  - Humorous news articles
Question 2

Are humorous and serious texts separable?

- Can we automatically distinguish between humorous and non-humorous texts?
  - Humorous and non-humorous data-sets are clearly separable
  - 80-95% accuracy in 10-fold cross-validation experiments
- Does this hold for different data sets?
  - Significant improvements over the 50% baseline observed for two data sets:
    - One-liners: 80%
    - News stories: 80-95%
Question 3

What are the distinctive features of humour?

- Analysis of linguistic features revealed two important characteristics
  - **Human-centeredness**: human-related semantic classes found dominant in humorous text as compared to non-humorous text
  - **Negative orientation**: humorous texts found predominantly “negative”
    - Properties validated through large-scale analysis on two different data sets

- Humour as “natural therapy” where tensions related to negative scenarios concerning us humans are relieved through laughter
Question 4

Can humour improve human-computer interaction?

- Automatic humorous additions in FunEmail
  - Entertaining and appropriate
  - Interest in adopting the application
Current work on computational humour

- Try to exploit further lexical semantics techniques
- E.g. (Bucaria, 2004) “Lexical and syntactic ambiguity as a source of humor”

⇒ Lexical, syntactic, phonological ambiguity

- Advertisement, News Headlines
The case of news headlines

- **Lexical ambiguity**
  - Men recommended more clubs for wives
  - Stadium air conditioning fails - Fans protest
  - Doctor testifies in horse suit
  - Queen Mary having bottom scraped

- **Syntactic ambiguity**
  - Lawyers give poor free legal advice
  - Babies are what the mother eats
  - Man eating piranha mistakenly sold as pet fish

- **Referential ambiguity**
  - Autos killing 110 a day: let’s resolve to do better
Initial steps for producing humorous expressions

- **Overall goal:**
  Realization of an environment for the production of *creative and humorous expressions.* e.g. newspaper titles, advertisements, ...

- **Current achievements:**
  - some basic and general techniques for *automatic* creation of *emotional language*;
  - indications for humorous expressions as variation of existing texts
Outline

1. Computational Humor
   - Humor generation (Carlo)
   - Humor recognition (Rada)

2. Affective Text
   - Lexical resources (Carlo)
   - Annotation of emotions in text (Rada)
   - Dancing with words (Carlo)
   - Emotions in blogs (Rada)
Emotion and texts: motivation

- Future of HCI is in themes such as entertainment, emotions, aesthetic pleasure, motivation, attention, engagement, etc.
- Automatically produce what human graphic designers sometime manually do for TV/Web presentations (e.g. advertisements, news titles, …)
- Studying the relation between natural language and affective information and dealing with its computational treatment is becoming crucial.
Affective lexical resources

- What an emotion is?
  - Notoriously it is a difficult problem.
    - Many approaches: facial expressions (Ekman), action tendencies (Frijda), physiological activity (Ax), ...

- Emotions, of course, are not linguistic things
- However the most convenient access we have to them is *through the language*

- *Ortony et al. (1987)* introduced the problem
  => an analysis of 500 words taken from literature on emotions. The words are then organized in a taxonomy.
Some affective lexical resources

- General Inquirer (Stone et al.)
- SentiWordNet (Esuli and Sebastiani)
- Affective Norms for English Words (ANEW) (Bradley and Lang)
- WordNet Affect (Strapparava and Valitutti)
The General Inquirer is basically a mapping tool. It maps each text file with counts on dictionary-supplied categories.

The currently distributed version combines the "Harvard IV-4" dictionary content-analysis categories, the "Lasswell" dictionary content-analysis categories, and five categories based on the social cognition work of Semin and Fiedler, making for 182 categories in all.

Each category is a list of words and word senses.

Currently, the category "negative" is our largest with 2291 entries.

http://www.wjh.harvard.edu/~inquirer/
GI marker categories

- Emotions (EMOT): anger, fury, distress, happy, etc.
- Frequency (FREQ): occasional, seldom, often, etc.
- Evaluative Adjective (EVAL): good, bad, beautiful, hard, easy, etc.
- Dimensionality Adjective (DIM): big, little, short, long, tall, etc.
- Position Adjective (POS): low, lower, upper, high, middle, first, fourth (ordinal numbers) etc.
- Degree Adverbs (DEG): very, extremely, too, rather, somewhat...

...
SentiWordNet (Esuli and Sebastiani, 2006) is a lexical resource in which each synset $s$ of WordNet is associated to three numerical scores $Obj(s)$, $Pos(s)$ and $Neg(s)$, describing how Objective, Positive, and Negative the terms contained in the synset are.

Positive - Negative and Subjective - Objective classification

The three scores are derived by combining the results produced by a committee of eight ternary classifiers.
SentiWordNet

Visualization of the opinion related properties of the term estimable

Adjective
3 senses found.

estimable(1)
deserving of respect or high regard

P = 0.75, N = 0, O = 0.25

honorable(5) good(4) respectable(2) estimable(2)
deserving of esteem and respect; "all respectable companies give guarantees"; "ruined the family's good name"

P = 0.625, N = 0.25, O = 0.125

computable(1) estimable(3)
may be computed or estimated; "a calculable risk"; "computable odds"; "estimable assets"

P = 0, N = 0, O = 1

(c) Andrea Esuli 2005 - andrea.esuli@isi.cnr.it
Affective Norms for English Words (ANEW)

- ANEW was developed to provide a set of normative emotional ratings for a large number of words in English.
- A set of verbal materials that have been rated in terms of pleasure, arousal, and dominance, manually annotated.

### Affective Norms for English Words. All Subjects

<table>
<thead>
<tr>
<th>Description</th>
<th>Word No.</th>
<th>Valence Mean (SD)</th>
<th>Arousal Mean (SD)</th>
<th>Dominance Mean (SD)</th>
<th>Word Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>abduction</td>
<td>621</td>
<td>2.76 (2.06)</td>
<td>5.53 (2.43)</td>
<td>3.49 (2.38)</td>
<td>1</td>
</tr>
<tr>
<td>abortion</td>
<td>622</td>
<td>3.50 (2.30)</td>
<td>5.39 (2.80)</td>
<td>4.59 (2.54)</td>
<td>6</td>
</tr>
<tr>
<td>absurd</td>
<td>623</td>
<td>4.26 (1.82)</td>
<td>4.36 (2.20)</td>
<td>4.73 (1.72)</td>
<td>17</td>
</tr>
<tr>
<td>abundance</td>
<td>624</td>
<td>6.59 (2.01)</td>
<td>5.51 (2.63)</td>
<td>5.80 (2.16)</td>
<td>13</td>
</tr>
<tr>
<td>abuse</td>
<td>1</td>
<td>1.80 (1.23)</td>
<td>6.83 (2.70)</td>
<td>3.69 (2.94)</td>
<td>18</td>
</tr>
<tr>
<td>acceptance</td>
<td>625</td>
<td>7.98 (1.42)</td>
<td>5.40 (2.70)</td>
<td>6.64 (1.91)</td>
<td>49</td>
</tr>
</tbody>
</table>
ANEW: the annotation procedure

Happy vs. Unhappy

Exited vs. Calm

Controlled vs. In-control

Give score, for each word
Affective semantic similarity

- All words can potentially convey affective meaning
- Even those not directly related to emotions can evoke pleasant or painful experiences
- Some of them are related to the individual story
- But for many others the affective power is part of the collective imagination (e.g. mum, ghost, war, ...)
- cfr. Ortony & Clore

Affective words

- Direct affective words that refer directly to emotional states (e.g. fear, love, ...)
- Indirect affective words that have an indirect reference (e.g. monster, cry, ...)
- Many words can potentially convey affective meaning
- For the second group of words the affective power can be induced automatically form large corpora of texts (e.g. British National Corpus, ~100 millions of words)
WordNet Affect

- We built an affective lexical resource, essential for affective computing, computational humor, text analysis, etc.
- It is a lexical repository of the *direct affective words*
- The resource, named WordNet-Affect, started from WordNet, through selection and labeling of synsets representing affective concepts.
WordNet

- WordNet is an on-line lexical reference system whose design is inspired by psycholinguistic theories of human lexical memory.
- English nouns, verbs, adjectives and adverbs are organized into synonym sets (synsets), each representing one underlying lexical concept.
- IRST extensions: multilinguality and Domain Labels (WordNet Domains).
Analogy with WordNet domains

- In WordNet Domains each synset has been annotated with a domain label (e.g. Sport, Medicine, Politics) selected from a set of 200 labels hierarchically organized.

- In WordNet-Affect we have an additional hierarchy of affective domain labels (independent from the domain labels) with which the synsets representing affective concepts are annotated.
A-Labels and some examples

<table>
<thead>
<tr>
<th>A-Label</th>
<th>Examples of Synsets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EMOTION</strong></td>
<td>noun &quot;anger#1&quot;, verb &quot;fear#1&quot;</td>
</tr>
<tr>
<td><strong>MOOD</strong></td>
<td>noun &quot;animosity#1&quot;, adjective &quot;amiable#1&quot;</td>
</tr>
<tr>
<td><strong>TRAIT</strong></td>
<td>noun &quot;aggressiveness#1&quot;, adjective &quot;competitive#1&quot;</td>
</tr>
<tr>
<td><strong>COGNITIVE STATE</strong></td>
<td>noun &quot;confusion#2&quot;, adjective &quot;dazed#2&quot;</td>
</tr>
<tr>
<td><strong>PHYSICAL STATE</strong></td>
<td>noun &quot;illness#1&quot;, adjective &quot;all_in#1&quot;</td>
</tr>
<tr>
<td><strong>HEDONIC SIGNAL</strong></td>
<td>noun &quot;hurt#3&quot;, noun &quot;suffering#4&quot;</td>
</tr>
<tr>
<td><strong>EMOTION-ELICITING SITUATION</strong></td>
<td>noun &quot;awkwardness#3&quot;, adjective &quot;out_of_danger#1&quot;</td>
</tr>
<tr>
<td><strong>EMOTIONAL RESPONSE</strong></td>
<td>noun &quot;cold_sweat#1&quot;, verb &quot;tremble#2&quot;</td>
</tr>
<tr>
<td><strong>BEHAVIOUR</strong></td>
<td>noun &quot;offense#1&quot;, adjective &quot;inhibited#1&quot;</td>
</tr>
<tr>
<td><strong>ATTITUDE</strong></td>
<td>noun &quot;intolerance#1&quot;, noun &quot;defensive#1&quot;</td>
</tr>
<tr>
<td><strong>SENSATION</strong></td>
<td>noun &quot;coldness#1&quot;, verb &quot;feel#3&quot;</td>
</tr>
</tbody>
</table>

Freely available (for research purposes) at http://wndomains.itc.it
New extensions of WN-affect

- **Specialization of the Emotional Hierarchy.**
  For the present work we provide a specialization of the a-label *Emotion*

- **Stative/Causative tagging.**
  Concerning mainly the adjectival interpretation

- **Valence Tagging.**
  Positive/Negative dimension
Emotional hierarchy

- With respect to WN-Affect, we provided some additional a-labels, hierarchically organized starting from the a-label *Emotion*
- About 1637 words / 918 synsets

<table>
<thead>
<tr>
<th>Synset</th>
<th>A-Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>joy, joyousness, joyfulness</td>
<td>EMOTION → POSITIVE → GENERAL-JOY → JOY</td>
</tr>
<tr>
<td>scare, panic_attack</td>
<td>EMOTION → NEGATIVE → NEGATIVE-FEAR → SCARE</td>
</tr>
<tr>
<td>surprise</td>
<td>EMOTION → AMBIGUOUS → SURPRISE</td>
</tr>
<tr>
<td>indifference</td>
<td>EMOTION → NEUTRAL → NEUTRAL-UNCONCERN → INDIFFERENCE</td>
</tr>
</tbody>
</table>
Valence tagging

- Distinguishing synsets according to emotional valence
  - Positive emotions (joy#1, enthusiasm#1),
  - Negative emotions (fear#1, horror#1),
  - Ambiguous, when the valence depends on the context (surprise#1),
  - Neutral, when the synset is considered affective but not characterized by valence (indifference#1)
Affective semantic similarity

- We needed a technique for evaluating the affective weight of *indirect affective words*
- The mechanism is based on *similarity* between *generic terms* and *affective lexical concepts*
- We estimated term similarity from a large scale corpus (BNC ~ 100 millions of words)
Homogeneous representations

- In the Latent Semantic Space, we can represent in a homogeneous way
  - Words
  - Texts
  - Synsets

- Each text (and synsets) can be represented in the LSA space exploiting a variation of the *pseudo-document* methodology
  => summing up the normalized LSA vectors of all the terms contained in it
LSA space

\[ \text{synset} = w_1 + w_2 + w_3 \]

\[ \text{term} = w_1 \]

- Similarity: \textit{cosine} among vectors
Affective synset representation

- Thus an affective synset (and then an emotional category) can be represented in the Latent Semantic Space.
- We can compute a similarity measure among terms and affective categories.
- Ex. the term “gift” is highly related (in BNC) with the emotional categories:
  - Love (with positive valence)
  - Compassion (with negative valence)
  - Surprise (with ambiguous valence)
  - Indifference (with neutral valence)
An example: *university*

<table>
<thead>
<tr>
<th>Related emotional terms</th>
<th>Positive emotional category</th>
</tr>
</thead>
<tbody>
<tr>
<td>university</td>
<td>Enthusiasm</td>
</tr>
<tr>
<td>professor</td>
<td>Sympathy</td>
</tr>
<tr>
<td>scholarship</td>
<td>Devotion</td>
</tr>
<tr>
<td>achievement</td>
<td>Encouragement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Related emotional terms</th>
<th>Negative emotional category</th>
</tr>
</thead>
<tbody>
<tr>
<td>university</td>
<td>Downheartedness</td>
</tr>
<tr>
<td>professor</td>
<td>Antipathy</td>
</tr>
<tr>
<td>study</td>
<td>Isolation</td>
</tr>
<tr>
<td>scholarship</td>
<td>Melancholy</td>
</tr>
</tbody>
</table>
Affective synset similarity

- The adjective **terrific** is polysemous
  - a sense of \{fantastic, howling, marvelous, rattling, terrific, tremendous wonderful\}
    - *extraordinarily good*.
    - most similar to the positive emotion **Joy**
  - a sense of \{terrific, terrifying\} - *causing extreme terror*.
    - most similar to the negative emotion **Distress**
## News titles

- E.g. the affective weight of some news titles

<table>
<thead>
<tr>
<th>News titles (Google-news)</th>
<th>Emotion</th>
<th>Word with highest affective weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review: `King Kong' a giant pleasure</td>
<td>Joy</td>
<td>pleasure#n</td>
</tr>
<tr>
<td>Romania: helicopter crash kills four people</td>
<td>Fear</td>
<td>crash#v</td>
</tr>
<tr>
<td>Record sales suffer steep decline</td>
<td>Sadness</td>
<td>suffer#v</td>
</tr>
<tr>
<td>Dead whale in Greenpeace protest</td>
<td>Anger</td>
<td>protest#v</td>
</tr>
</tbody>
</table>
Affective evaluative expressions

- We defined the *affective weight* the similarity value between an emotional vector and an input term vector.
- Given a term (i.e. university), ask for related terms that have a positive affective valence, possibly according to some emotional category.
- Given two terms, check if they are semantically related, with respect to some emotional category.
Other examples

- Given in input a target term and a valence value
  - select the corresponding emotional category with maximum affective weight
  - produce a noun phrase, using the target term modified by an evaluative term (e.g. by a causative adjective)

- Input: gun, negative valence
  - => emotional category: Horror
  - "frightening gun"
Possible Applications

- **Computer Assisted Creativity**
  - Automatic personalized advertisement, Computational Humor, persuasive communication

- **Verbal Expressivity of Embodied Conversational Agents**
  - Intelligent dynamic word selection for appropriate conversation

- **Sentiment Analysis**
  - Text categorization according to affective relevance, opinion analysis
Summing up

1. WordNet-Affect provides the representation of direct affective terms
2. LSA from the BNC gives a measure of the similarity between direct affective terms and generic terms
Summing up

- Some resources and functionalities for dealing with affective evaluative terms
- An affective hierarchy as an extension of WordNet-Affect lexical database, including emotion, causative/stative and valence tagging
- A semantic similarity mechanism acquired in an unsupervised way from a large corpus, providing relations among concepts and emotional categories
Outline

1. Computational Humor
   - Humor generation (Carlo)
   - Humor recognition (Rada)

2. Affective Text
   - Lexical resources (Carlo)
   - Annotation of emotions in text (Rada)
   - Dancing with words (Carlo)
   - Emotions in blogs (Rada)
Annotation of emotions in text

- **Semeval 2007 task**
- Emotion classification of news headlines
- Headlines typically consist of few words and often written to “provoke” emotions (e.g. to attract reader’s attention)
  - Affective/emotional features probably present
- Suitable for use in automatic emotion recognition
Data and objective

- Corpus
  - News titles from the web sites Google News, CNN, New York Times, BBC over a period of time of 3 months
  - Development set of 250 headlines
  - Test set of 1,000 annotated headlines

Thailand attacks kill three, injure 70

Women face greatest threat of violence at home, study finds

Prehistoric lovers found locked in eternal embrace

Male sweat boosts women's hormone levels
Data and objective

- **Objective**
  
  - Provided a set of predefined six emotion labels (*Anger, Disgust, Fear, Joy, Sadness, Surprise*) classify the titles with
    - the appropriate emotion label and/or
    - a positive/negative valence indication
  
  - Emotion labeling and valence classification are seen as independent tasks
  
  - The task was carried out in an unsupervised setting
  
  - We want to emphasize emotion lexical semantics, avoid biasing towards simple text categorization
Data and objective

- **Other Data**
  - Participants were free to use any resources they want
  - We provide a set of words extracted from WordNet-Affect (Strapparava and Valitutti, 2004), relevant to the six emotions of interest
  - Links to other possibly useful resources on the Web - e.g. SentiWordNet (Esuli and Sebastiani, 2006)
Data annotation

- We developed a web-based annotation interface:
  - One headline at time, six slide bars for emotions and one slide bar for valence
  - Interval for emotion annotations [0,100], while [-100, 100] for valence annotations (0 means neutral)

  ⇒ Finer-grained scale than typical 0/1 annotations
Data annotation

- Six annotators
- Presence of words or phrases with emotional content, as well the overall feeling invoke by the headline
- Inter-annotator agreement: Pearson correlation measure

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>49.55</td>
</tr>
<tr>
<td>Disgust</td>
<td>44.51</td>
</tr>
<tr>
<td>Fear</td>
<td>63.81</td>
</tr>
<tr>
<td>Joy</td>
<td>59.91</td>
</tr>
<tr>
<td>Sadness</td>
<td>68.19</td>
</tr>
<tr>
<td>Surprise</td>
<td>36.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Valence</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valence</td>
<td>78.01</td>
</tr>
</tbody>
</table>
Evaluations

- Fine-grained evaluation
  - Pearson between system scores and gold standard, averaged over all the headlines in the data set

- Coarse-grained evaluation
  - Each emotion annotation was mapped in a 0/1 classification 0= [0,50) and 1=[50,100], and valence annotation into -1/0/1 -1=[-100,-50] 0=(-50,50) 1=[50,100]
  - Then accuracy, precision and recall wrt the possible classes
## Participating systems

- **Five teams with**
  - Five systems for valence classification
  - Three systems for emotion labeling

<table>
<thead>
<tr>
<th>Teams/ Contact</th>
<th>Emotion Labeling</th>
<th>Valence Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordia University</td>
<td></td>
<td>- CLaC</td>
</tr>
<tr>
<td>- Alina Andreevskai</td>
<td></td>
<td>- CLaC-NaïveBayes</td>
</tr>
<tr>
<td>Swedish Institute of Computer Science - Magnus Sahlgren</td>
<td></td>
<td>- SICS</td>
</tr>
<tr>
<td>Swarthmore College</td>
<td>- SWAT</td>
<td>- SWAT</td>
</tr>
<tr>
<td>- Phil Katz</td>
<td></td>
<td>- UPAR7</td>
</tr>
<tr>
<td>University Paris 7</td>
<td>- UPAR7</td>
<td>- UPAR7</td>
</tr>
<tr>
<td>- Francois-Regis Chaumartin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Alicante</td>
<td>- UA</td>
<td></td>
</tr>
<tr>
<td>- Zornitsa Kozareva</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Participant Systems

<table>
<thead>
<tr>
<th>System</th>
<th>Approach</th>
<th>Main Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>- CLaC</td>
<td>Unsupervised knowledge-based system</td>
<td>- Sentiment words</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Valence shifters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- set of rules</td>
</tr>
<tr>
<td>- CLaC-NaïveBayes</td>
<td>Supervised corpus-based</td>
<td>- additional corpus manually annotated</td>
</tr>
<tr>
<td>- SICS</td>
<td>Word space model + seed words</td>
<td>- LA times corpus</td>
</tr>
<tr>
<td>- SWAT</td>
<td>Supervised</td>
<td>- Roget Thesaurus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- additional 1000 headlines manually annotated</td>
</tr>
<tr>
<td>- UPAR7</td>
<td>Rule-based system with linguistic approach</td>
<td>- Stanford parser</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SentiWordNet</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- WordNet-Affect</td>
</tr>
<tr>
<td>- UA</td>
<td>Point-wise Mutual Information</td>
<td>- search engines</td>
</tr>
</tbody>
</table>
## Results

### System results for valence annotations

<table>
<thead>
<tr>
<th></th>
<th>Fine $r$</th>
<th>Acc.</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLaC</td>
<td>47.70</td>
<td>55.10</td>
<td>61.42</td>
<td>9.20</td>
<td>16.00</td>
</tr>
<tr>
<td>UPAR7</td>
<td>36.96</td>
<td>55.00</td>
<td>57.54</td>
<td>8.78</td>
<td>15.24</td>
</tr>
<tr>
<td>SWAT</td>
<td>35.25</td>
<td>53.20</td>
<td>45.71</td>
<td>3.42</td>
<td>6.36</td>
</tr>
<tr>
<td>CLaC-NB</td>
<td>25.41</td>
<td>31.20</td>
<td>31.18</td>
<td>66.38</td>
<td>42.43</td>
</tr>
<tr>
<td>SICS</td>
<td>20.68</td>
<td>29.00</td>
<td>28.41</td>
<td>60.17</td>
<td>38.60</td>
</tr>
</tbody>
</table>

### System results for emotion labeling

<table>
<thead>
<tr>
<th></th>
<th>Coarse $r$</th>
<th>Acc.</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWAT</td>
<td>24.51</td>
<td>92.10</td>
<td>12.00</td>
<td>5.00</td>
<td>7.06</td>
</tr>
<tr>
<td>UA</td>
<td>23.20</td>
<td>86.40</td>
<td>12.74</td>
<td>21.6</td>
<td>16.03</td>
</tr>
<tr>
<td>UPAR7</td>
<td>32.33</td>
<td>93.60</td>
<td>16.67</td>
<td>1.66</td>
<td>3.02</td>
</tr>
<tr>
<td>Disgust</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWAT</td>
<td>18.55</td>
<td>97.20</td>
<td>0.00</td>
<td>0.00</td>
<td>-</td>
</tr>
<tr>
<td>UA</td>
<td>16.21</td>
<td>97.30</td>
<td>0.00</td>
<td>0.00</td>
<td>-</td>
</tr>
<tr>
<td>UPAR7</td>
<td>12.85</td>
<td>95.30</td>
<td>0.00</td>
<td>0.00</td>
<td>-</td>
</tr>
<tr>
<td>Fear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWAT</td>
<td>32.52</td>
<td>84.80</td>
<td>25.00</td>
<td>14.40</td>
<td>18.27</td>
</tr>
<tr>
<td>UA</td>
<td>23.15</td>
<td>75.30</td>
<td>16.23</td>
<td>26.27</td>
<td>20.06</td>
</tr>
<tr>
<td>UPAR7</td>
<td>44.92</td>
<td>87.90</td>
<td>33.33</td>
<td>2.54</td>
<td>4.72</td>
</tr>
<tr>
<td>Joy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWAT</td>
<td>26.11</td>
<td>80.60</td>
<td>35.41</td>
<td>9.44</td>
<td>14.91</td>
</tr>
<tr>
<td>UA</td>
<td>2.35</td>
<td>81.80</td>
<td>40.00</td>
<td>2.22</td>
<td>4.21</td>
</tr>
<tr>
<td>UPAR7</td>
<td>22.49</td>
<td>82.20</td>
<td>54.54</td>
<td>6.66</td>
<td>11.87</td>
</tr>
<tr>
<td>Sadness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWAT</td>
<td>38.98</td>
<td>87.70</td>
<td>32.50</td>
<td>11.92</td>
<td>17.44</td>
</tr>
<tr>
<td>UA</td>
<td>12.28</td>
<td>88.90</td>
<td>25.00</td>
<td>0.91</td>
<td>1.76</td>
</tr>
<tr>
<td>UPAR7</td>
<td>40.98</td>
<td>89.00</td>
<td>48.97</td>
<td>22.02</td>
<td>30.38</td>
</tr>
<tr>
<td>Surprise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWAT</td>
<td>11.82</td>
<td>89.10</td>
<td>11.86</td>
<td>10.93</td>
<td>11.78</td>
</tr>
<tr>
<td>UA</td>
<td>7.75</td>
<td>84.60</td>
<td>13.70</td>
<td>16.56</td>
<td>15.00</td>
</tr>
<tr>
<td>UPAR7</td>
<td>16.71</td>
<td>88.60</td>
<td>12.12</td>
<td>1.25</td>
<td>2.27</td>
</tr>
</tbody>
</table>

System results for valence annotations

System results for emotion labeling