

# SemEval-2013 Task 2: Sentiment Analysis in Twitter

Preslav Nakov  
Zornitsa Kozareva  
Alan Ritter

Sara Rosenthal  
Veselin Stoyanov  
Theresa Wilson

# Task 2 - Overview

## Sentiment Analysis

- Understanding how opinions and sentiments are expressed in language
- Extracting opinions and sentiments from human language data

## Social Media

- Short messages
- Informal language
- Creative spelling, punctuation, words, and word use
- Genre-specific terminology (#hashtags) and discourse (RT)

**Task Goal:** Promote sentiment analysis research in Social Media

→ SemEval Tweet Corpus

- Publically available (within Twitter TOS)
- Phrase and message-level sentiment
- Tweets and SMS<sup>1</sup> for evaluating generalizability

<sup>1</sup> From NUS SMS Corpus (Chen and Kan, 2012)

# Task Description

## Two subtasks:

- A. Phrase-level sentiment
- B. Message-level sentiment

Classify as *positive*, *negative*, *neutral/objective*:

- Words and phrases identified as subjective  
[Subtask A]
- Messages (tweets/SMS)  
[Subtask B]

# Data Collection

Extract NEs (Ritter et al., 2011)

Identify Popular Topics (Ritter et al., 2012)

- NEs frequently associated with specific dates

Extract Messages Mentioning Topics

Filter Messages for Sentiment

- Keep if  $\geq$  pos/neg term from SentiWordNet ( $>0.3$ )

Data for Annotation

# Annotation Task

## Mechanical Turk HIT (3-5 workers per tweet)

**Instructions:** Subjective words are ones which convey an opinion. *Given a sentence, identify whether it is objective, positive, negative, or neutral. Then, identify each subjective word or phrase in the context of the sentence and mark the position of its start and end in the text boxes below.* The number above each word indicates its position. The word/phrase will be generated in the adjacent textbox so that you can confirm that you chose the correct range. Choose the polarity of the word or phrase by selecting one of the radio buttons: positive, negative, or neutral. If a sentence is not subjective please select the checkbox indicating that "There are no subjective words/phrases". Please read the examples and invalid responses before beginning if this is your first time answering this hit.

Sentence: friday<sup>1</sup> evening<sup>2</sup> plans<sup>3</sup> were<sup>4</sup> great,<sup>5</sup> but<sup>6</sup> saturday's<sup>7</sup> plans<sup>8</sup> didnt<sup>9</sup> go<sup>10</sup> as<sup>11</sup> expected<sup>12</sup> --<sup>13</sup> i<sup>14</sup> went<sup>15</sup> dancing<sup>16</sup> &<sup>17</sup> it<sup>18</sup> was<sup>19</sup> an<sup>20</sup> ok<sup>21</sup> club,<sup>22</sup> but<sup>23</sup> "terribly"<sup>24</sup> crowded<sup>25</sup> :-(<sup>26</sup>

Overall, the sentence is  Objective  Positive  Negative  Neutral

There are no subjective words/phrases.

Subjective Phrase 1:  to   Positive  Negative  Neutral

Subjective Phrase 2:  to   Positive  Negative  Neutral

Subjective Phrase 3:  to    Positive  Negative  Neutral

Subjective Phrase 4:  to    Positive  Negative  Neutral

[add more phrases >>](#)

# Data Annotations

Final annotations determined using majority vote

Worker 1	I would love to watch Vampire Diaries tonight :) and some Heroes! Great combination
Worker 2	I would love to watch Vampire Diaries tonight :) and some Heroes! Great combination
Worker 3	I would love to watch Vampire Diaries tonight :) and some Heroes! Great combination
Worker 4	I would love to watch Vampire Diaries tonight :) and some Heroes! Great combination
Worker 5	I would love to watch Vampire Diaries tonight :) and some Heroes! Great combination
<b>Intersection</b>	I would love to watch Vampire Diaries tonight :) and some Heroes! Great combination

# Distribution of Classes

## Subtask A

	Train	Dev	Test-TWEET	Test-SMS
Positive	5,895	648	2,734 (60%)	1,071 (46%)
Negative	3,131	430	1,541 (33%)	1,104 (47%)
Neutral	471	57	160 (3%)	159 (7%)
Total			4,635	2,334

## Subtask B

	Train	Dev	Test-TWEET	Test-SMS
Positive	3,662	575	1,573 (41%)	492 (23%)
Negative	1,466	340	601 (16%)	394 (19%)
Neutral/Objective	4,600	739	1,640 (43%)	1,208 (58%)
Total			3,814	2,094

# Options for Participation

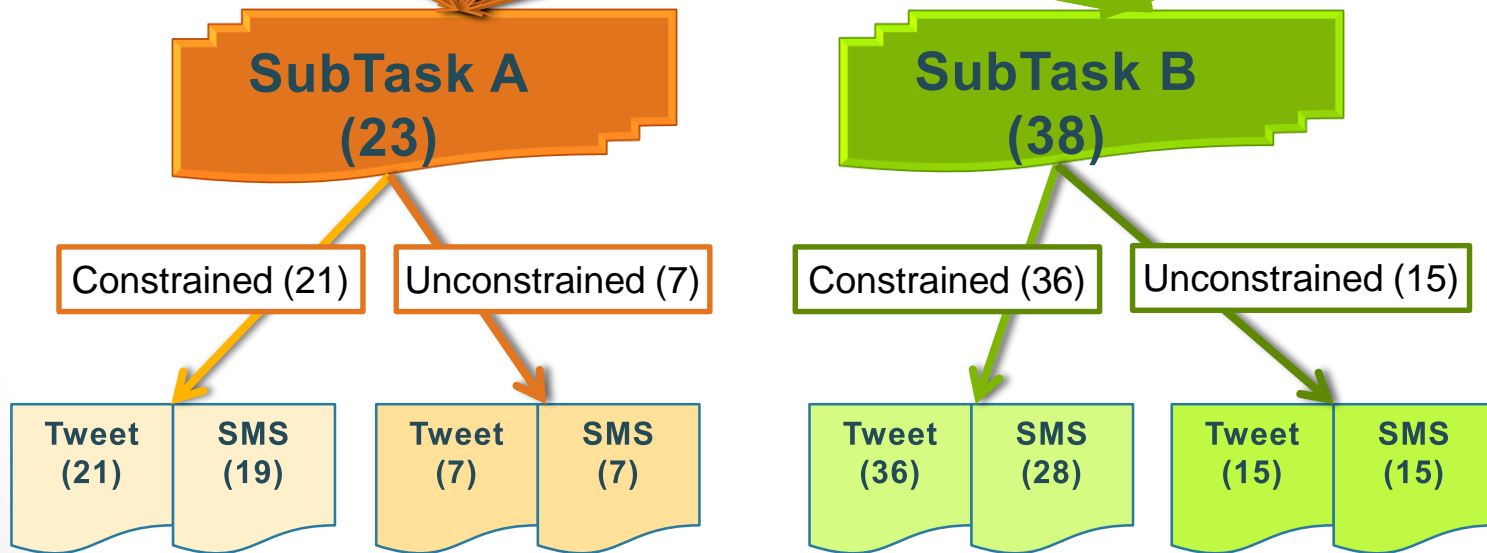
1. **Subtask A** and/or **Subtask B**
2. **Constrained\*** and/or **Unconstrained**
  - Refers to data used for training
3. **Tweets** and/or **SMS**

\* Used for ranking



# Participation

Teams (44)



Submissions (148)

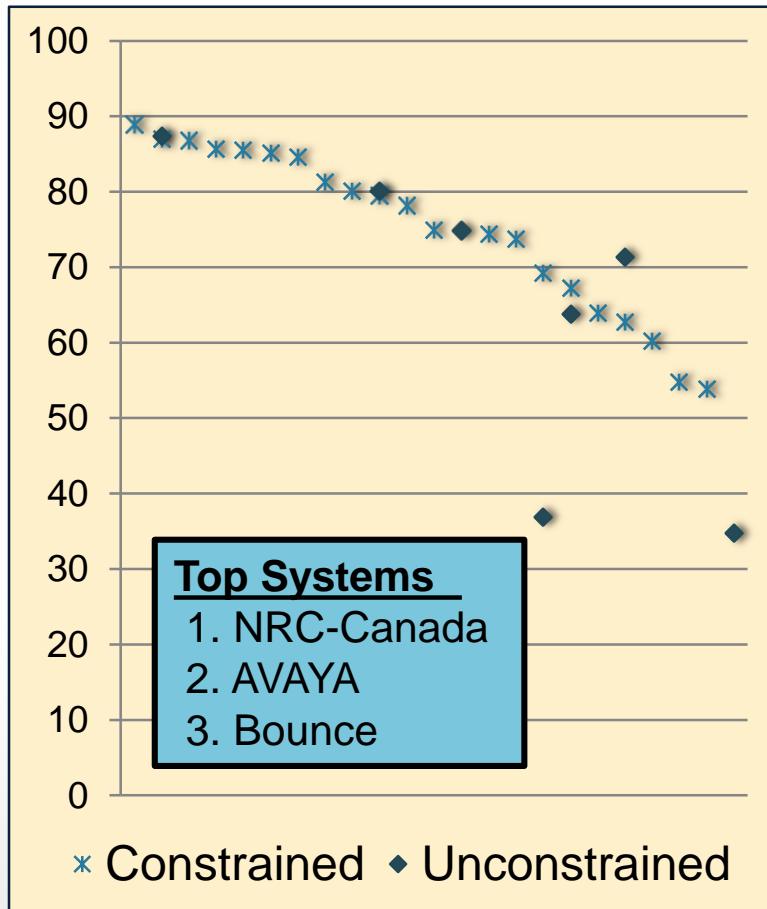
# Scoring

- Recall, Precision, F-measure calculated for pos/neg classes for each run submitted

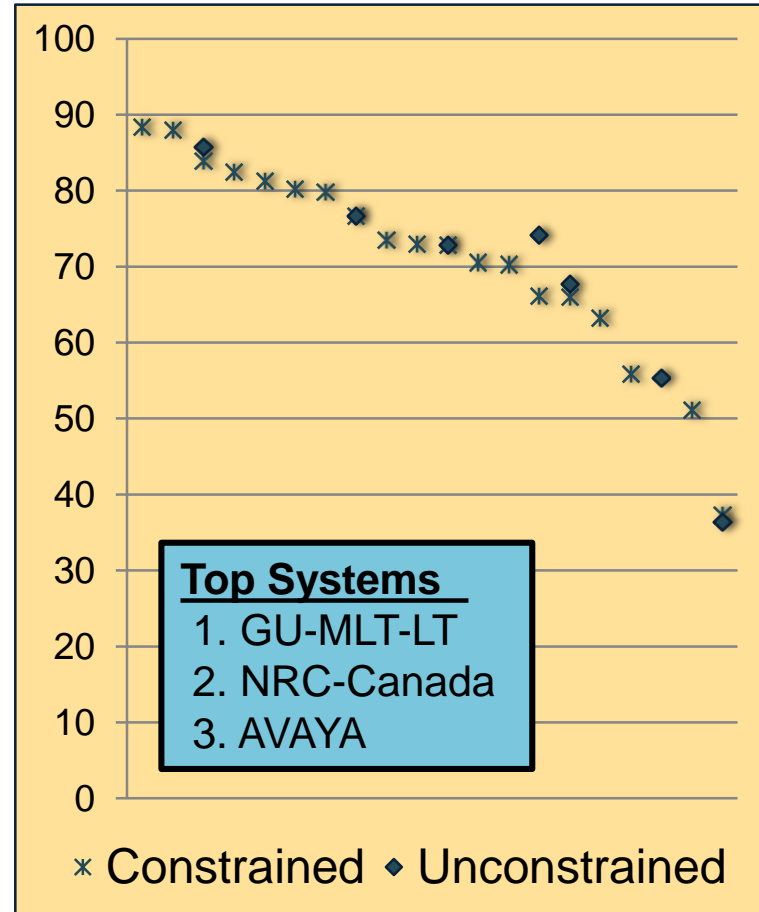
$$\text{Score} = \text{Ave}(\text{Pos F}, \text{Neg F})$$

# Subtask A (words/phrases) Results

## Tweets

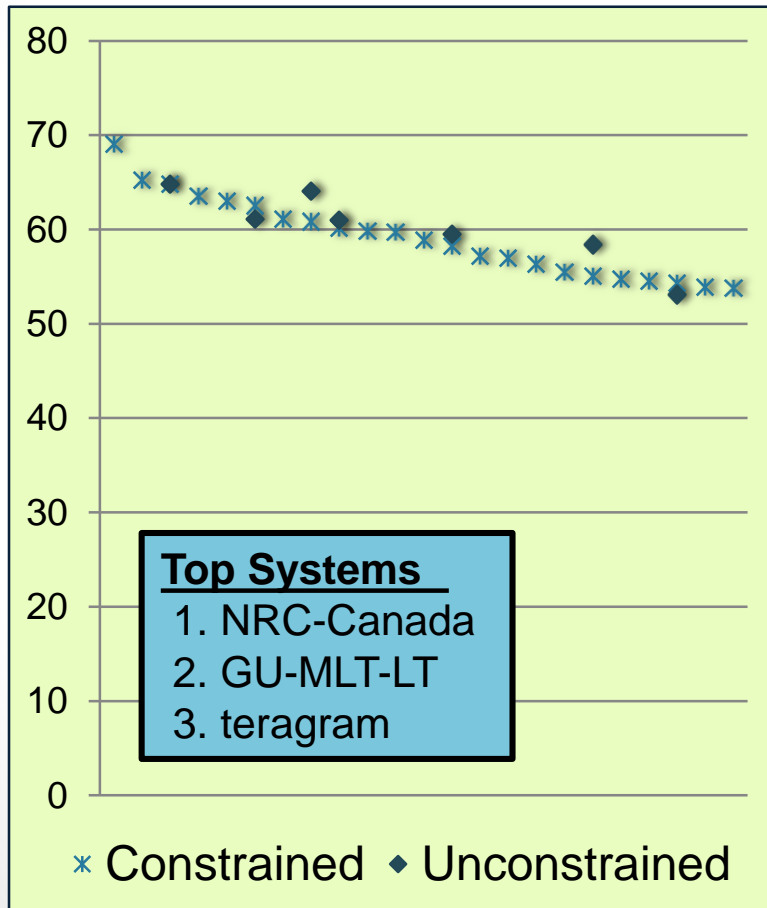


## SMS

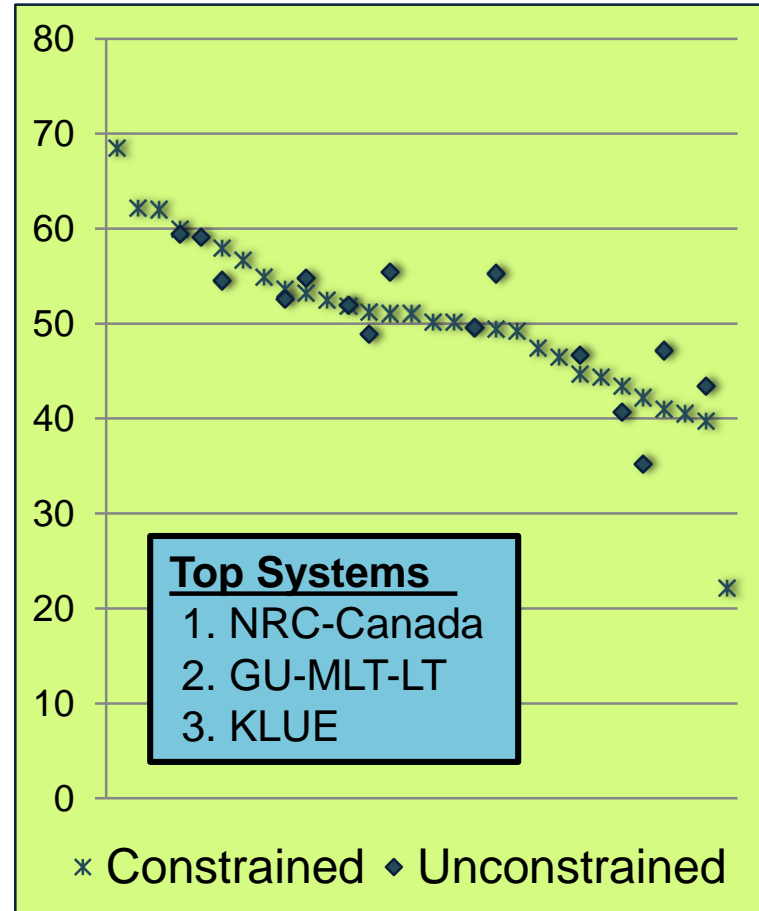


# Subtask B (messages) Results

## Tweets



## SMS



# Observations

Majority of systems were supervised and constrained

- 5 semi-supervised, 1 fully unsupervised

Systems that made best use of unconstrained option:

- **Subtask A:** senti.ue-en
- **Subtask B Tweet:** AVAYA, bwbaugh, ECNUCS, OPTIMA, sinai
- **Subtask B SMS:** bwbaugh, nlp.cs.aueb.gr, OPTIMA, SZTE-NLP

Most popular classifiers

- SVM, MaxEnt, linear classifier, Naive Bayes

# Thank You!

Special thanks to co-organizers:

**Preslav Nakov, Sara Rosenthal, Alan Ritter**  
**Zonitsa Kozareva, Veselin Stoyanov**

## SemEval Tweet Corpus

- Funding for annotations provided by:
  - JHU Human Language Technology Center of Excellence
  - ODNI IARPA