#hardtoparse: POS Tagging and Parsing the Twitterverse

Jennifer Foster, Özlem Çetinoğlu, Joachim Wagner, Joseph Le Roux, Stephen Hogan, Joakim Nivre, Deirdre Hogan and Josef van Genabith

AAAI Workshop on Analysing Microtext
San Francisco, August 8th, 2011
What are we doing?

1. Apply off-the-shelf NLP tools to tweets
What are we doing?

2. Investigate the drop in performance
What are we doing?

3. Retrain tools on automatically analysed tweets
What are we doing?

4. Investigate the changes
Why are we doing this?

- POS taggers and parsers assign structure to text based on automated linguistic analysis
Why are we doing this?

• POS taggers and parsers assign structure to text based on automated linguistic analysis
• Useful for MT, QA, Sentiment Analysis, etc.
Why are we doing this?

- POS taggers and parsers assign structure to text based on automated linguistic analysis.
- Useful for MT, QA, Sentiment Analysis, etc.
- Taggers/parsers trained on edited, well-formed text.
I just think he looks like a big baby, and ppl USED to call him that.

been playing with the new Canon EOS 500d and the Nikon D5000 over the weekend.

On Fox: RNC chair sends letter to GOP calling Obama “ARROGANT”
#tcot #sgp #hhrs

LOL!

Man Utd through to the last 8...

Twas okay.

FF > S4

i heart beltran

Bed soon
i heart beltran

been playing with the new Canon EOS 500d and the Nikon D5000 over the weekend.

I just think he looks like a big baby, and ppl USED to call him that

On Fox: RNC chair sends letter to GOP calling Obama “ARROGANT”
#tcot #sgp #hhrs

FF > S4

Twas okay.

LOL!

Man Utd through to the last 8...

Bed soon
I just think he looks like a big baby, and ppl USED to call him that

been playing with the new Canon EOS 500d and the Nikon D5000 over the weekend.

On Fox: RNC chair sends letter to GOP calling Obama “ARROGANT”
#tcot #sgp #hhrs

Twas okay.

LOL!

FF > S4

i heart beltran

Man Utd through to the last 8...

Bed soon
i heart beltran

LOL!

I just think he looks like a big baby, and ppl USED to call him that

On Fox: RNC chair sends letter to GOP calling Obama “ARROGANT”
#tcot #sgp #hhrs

been playing with the new Canon EOS 500d and the Nikon D5000 over the weekend.

Twas okay.

FF > S4

On Fox: RNC chair sends letter to GOP calling Obama “ARROGANT”
#tcot #sgp #hhrs

I just think he looks like a big baby, and ppl USED to call him that

Man Utd through to the last 8...

Bed soon
<table>
<thead>
<tr>
<th>Corpus Name</th>
<th>#Sentences</th>
<th>Average Sentence Length</th>
<th>Median Sentence Length</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TwitterDev</td>
<td>269</td>
<td>11.1</td>
<td>10</td>
<td>6.4</td>
</tr>
<tr>
<td>TwitterTest</td>
<td>250</td>
<td>11.3</td>
<td>10</td>
<td>6.8</td>
</tr>
<tr>
<td>TwitterTrain</td>
<td>1.4 million</td>
<td>8.6</td>
<td>7</td>
<td>6.1</td>
</tr>
</tbody>
</table>
• Why dependencies?
• Why *dependencies*?
• Why *labelled* dependencies?
• Why dependencies?
• Why labelled dependencies?
• Why Stanford labelled dependencies?
Types of Parsers

Raw Text

Phrase Structure Trees

Dependency Trees
Types of Parsers

Raw Text

Phrase Structure Trees

Dependency Trees

PHRASE STRUCTURE PARSER

STANFORD CONVERTER
Types of Parsers

Raw Text

POS TAGGER

POS-Tagged Text

DEPENDENCY PARSER

Dependency Trees
SVMTool and MaltParser models are trained on Wall Street Journal trees.
## Baseline POS Tagging Results

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSJ22</td>
<td>96.3</td>
</tr>
<tr>
<td>TwitterDev</td>
<td>84.1</td>
</tr>
</tbody>
</table>
## POS Tag Confusions

<table>
<thead>
<tr>
<th>Gold/System</th>
<th>Frequency</th>
<th>Gold/System</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNP/NN</td>
<td>59</td>
<td>VBZ/NNS</td>
<td>8</td>
</tr>
<tr>
<td>NN/NNP</td>
<td>54</td>
<td>UH/NNP</td>
<td>7</td>
</tr>
<tr>
<td>NNP/JJ</td>
<td>29</td>
<td>RB/NN</td>
<td>7</td>
</tr>
<tr>
<td>NNP/VB</td>
<td>10</td>
<td>NNP/CD</td>
<td>7</td>
</tr>
<tr>
<td>JJ/NN</td>
<td>10</td>
<td>NN/VB</td>
<td>6</td>
</tr>
<tr>
<td>UH/NN</td>
<td>8</td>
<td>VB/NN</td>
<td>6</td>
</tr>
<tr>
<td>JJ/NNP</td>
<td>8</td>
<td>VB/NNP</td>
<td>6</td>
</tr>
<tr>
<td>NNP/NNS</td>
<td>8</td>
<td>VBP/VB</td>
<td>6</td>
</tr>
<tr>
<td>NNPS/NNS</td>
<td>8</td>
<td>RP/IN</td>
<td>6</td>
</tr>
</tbody>
</table>
### POS Tag Confusions

<table>
<thead>
<tr>
<th>Gold/System</th>
<th>Frequency</th>
<th>Gold/System</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNP/NN</td>
<td>59</td>
<td>VBZ/NNS</td>
<td>8</td>
</tr>
<tr>
<td>NN/NNP</td>
<td>54</td>
<td>UH/NNP</td>
<td>7</td>
</tr>
<tr>
<td>NNP/JJ</td>
<td>29</td>
<td>RB/NN</td>
<td>7</td>
</tr>
<tr>
<td>NNP/VB</td>
<td>10</td>
<td>NNP/CD</td>
<td>7</td>
</tr>
<tr>
<td>JJ/NN</td>
<td>10</td>
<td>NN/VB</td>
<td>6</td>
</tr>
<tr>
<td>UH/NN</td>
<td>8</td>
<td>VB/NN</td>
<td>6</td>
</tr>
<tr>
<td>JJ/NNP</td>
<td>8</td>
<td>VB/NNP</td>
<td>6</td>
</tr>
<tr>
<td>NNP/NNS</td>
<td>8</td>
<td>VBP/VB</td>
<td>6</td>
</tr>
<tr>
<td>NNPS/NNS</td>
<td>8</td>
<td>RP/IN</td>
<td>6</td>
</tr>
</tbody>
</table>
## POS Tagging Examples

<table>
<thead>
<tr>
<th>i</th>
<th>heart</th>
<th>beltran</th>
</tr>
</thead>
<tbody>
<tr>
<td>FW</td>
<td>NN</td>
<td>NN</td>
</tr>
</tbody>
</table>
POS Tagging Examples

i     heart    beltran
FW    NN      NN

LOL    !
NNP    .
POS Tagging Examples

i    heart    beltran
FW   NN      NN

LOL   
NNP

Man    Utd    through    to    the    last    8    ...
NNP    NNP    IN           TO    DT    JJ    CD  :
## Baseline Parsing Results

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Labelled/Unlabelled Attachment Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVMTool POS Tags</td>
</tr>
<tr>
<td>WSJ22</td>
<td>88.0/90.6</td>
</tr>
<tr>
<td>TwitterDev</td>
<td>67.3/73.6</td>
</tr>
<tr>
<td></td>
<td>Gold POS Tags</td>
</tr>
<tr>
<td>WSJ22</td>
<td>90.0/91.6</td>
</tr>
<tr>
<td>TwitterDev</td>
<td>78.3/81.6</td>
</tr>
</tbody>
</table>
Parsing Example

i
FW

heart
NN

beltran
NN

nn
nn

nn
Parser Uptraining

• Use a more accurate parser, $P1$, to provide training material for a less accurate parser, $P2$ (Petrov et al. 2010)
Parser Uptraining

• Use a more accurate parser, $P1$, to provide training material for a less accurate parser, $P2$ (Petrov et al. 2010)

• Why not just use $P1$?
Parser Uptraining

• Use a more accurate parser, $P_1$, to provide training material for a less accurate parser, $P_2$ (Petrov et al. 2010)

• Why not just use $P_1$?
$P_2$ is faster.
Two Types of Uptraining

• *Vanilla*

\[ P1 = \text{Charniak and Johnson (C&J) reranking parser} \]

\[ P2 = \text{MaltParser} \]
Two Types of Uptraining

• *Vanilla*

\[ P1 = \text{Charniak and Johnson (C&J) reranking parser} \]
\[ P2 = \text{MaltParser} \]

• *Domain-adapted*

\[ P1 = \textit{Self-trained} \text{ version of C&J (McClosky et al. 2006)} \]
\[ P2 = \text{MaltParser} \]
Uptaining Results

67.33 % LAS

71.94 % LAS

67.33 % LAS
 Parsing Example

```
          i
         PRP

      heart
     NN

   dobj

    beltran
   NN

  nsubj
```
Uptraining Error Analysis

- Uptraining helps all dependency types
Uptraining Error Analysis

- Uptraining helps all dependency types
- Apart from nn and amod
Uptraining Error Analysis

• Uptraining helps all dependency types
• Apart from nn and amod
• Domain-adapted uptraining particularly helps neg, dobj and xcomp
Upt raining Error Analysis

• Upt raining helps all dependency types
• Apart from nn and amod
• Domain-adapted uptraining particularly helps neg, dobj and xcomp
• But doesn’t particularly help analysis of coordination (cc and conj), and prepositional phrases (pobj and prep)
Upt raining Error Analysis

• Upt raining helps all dependency types
• Apart from nn and amod
• Domain-adapted upt raining particularly helps neg, dobj and xcomp
• But doesn’t particularly help analysis of coordination (cc and conj), and prepositional phrases (pobj and prep)
• And doesn’t help at all with ccomp
Summary

• Apply off-the-shelf POS tagging and dependency parsing models to the language of Twitter
  – 12% drop in POS tagging accuracy
  – knock-on effect on dependency parsing accuracy (20% drop)
Summary

• Apply off-the-shelf POS tagging and dependency parsing models to the language of Twitter
  – 12% drop in POS tagging accuracy
  – knock-on effect on dependency parsing accuracy (20%)

• Uptraining recovers some of the gap (4% LAS improvement on TwitterTest)
What next?

• Other sources of uptraining material
What next?

• Other sources of uptraining material
• Model combination (Petrov 2010, Surdeanu and Manning 2010)
What next?

- Other sources of uptraining material
- Model combination (Petrov 2010, Surdeanu and Manning, 2010)
- Twitter-specific resources (Gimpel et al. 2011)
#thanksforlistening