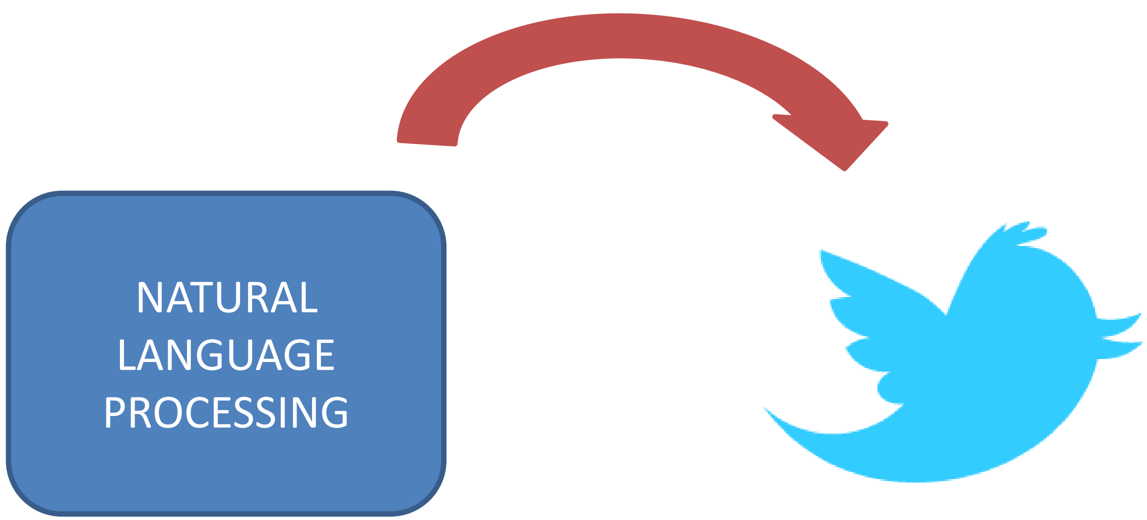
#hardtoparse: POS Tagging and Parsing the Twitterverse

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

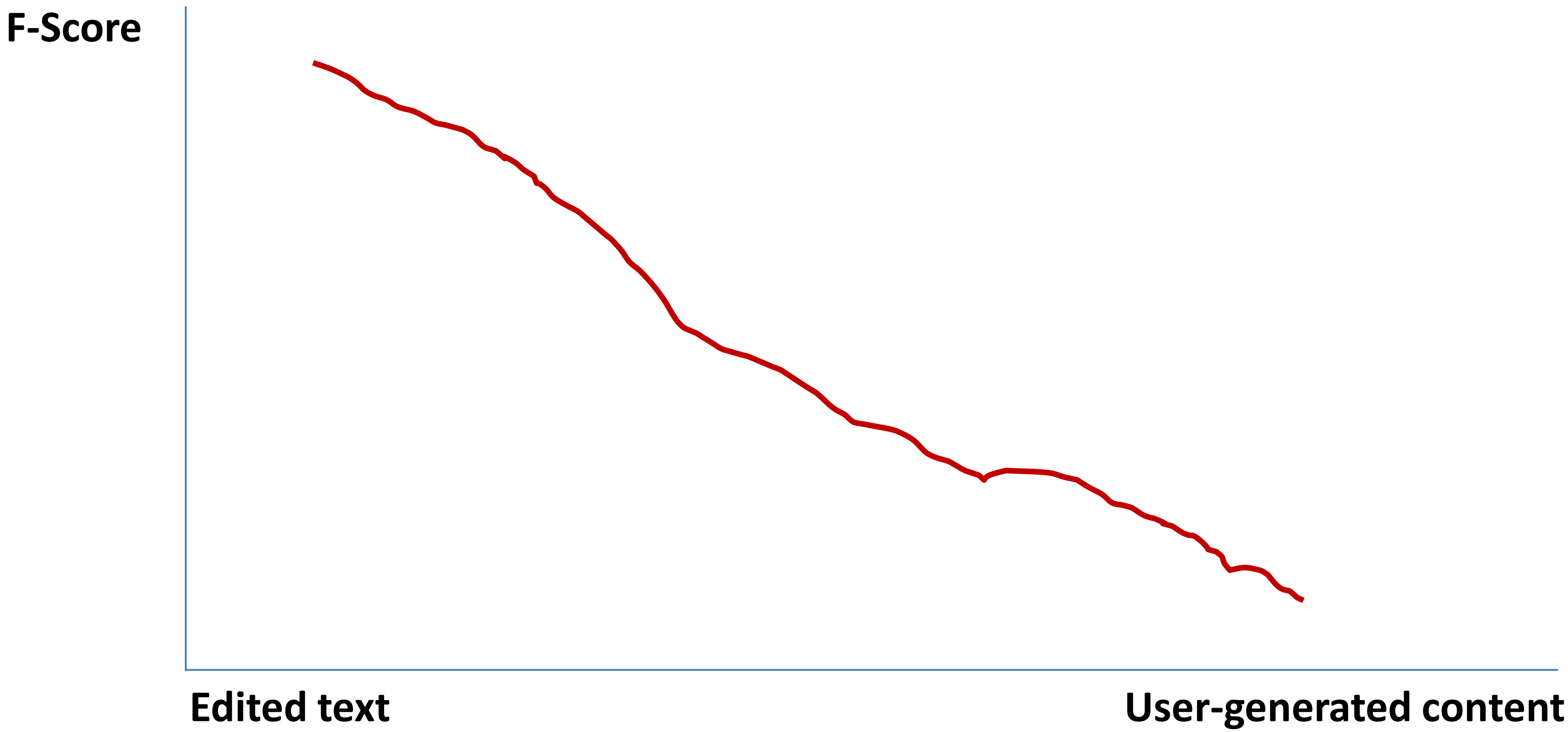
What are we doing?

1. Apply off-the-shelf NLP tools to tweets



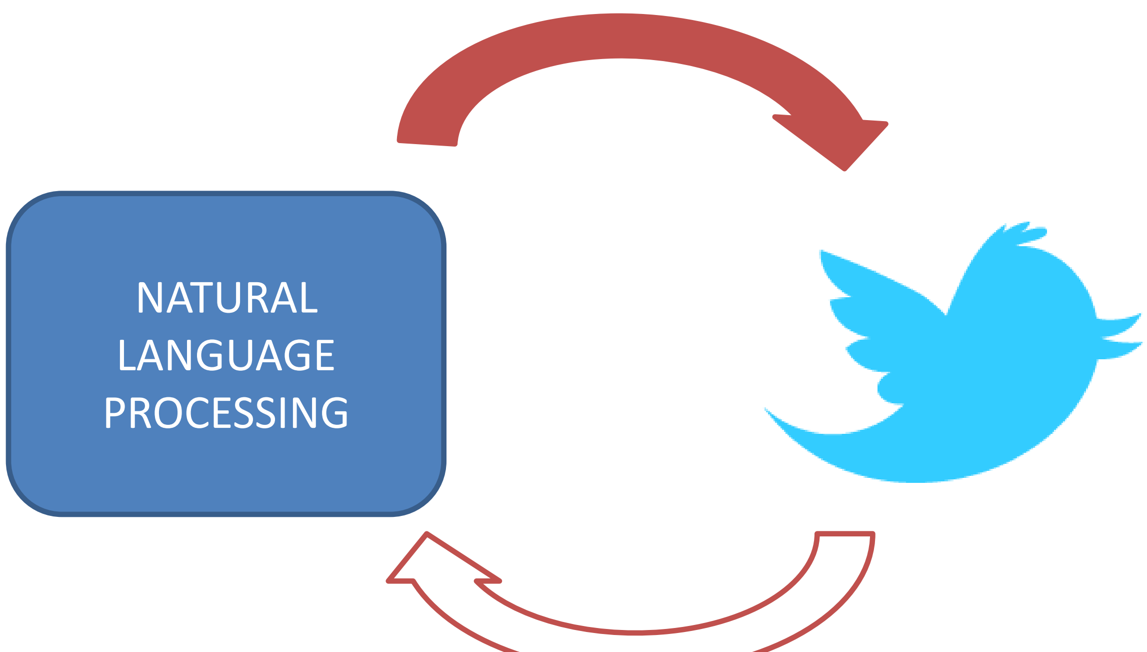
What are we doing?

2. Investigate the drop in performance



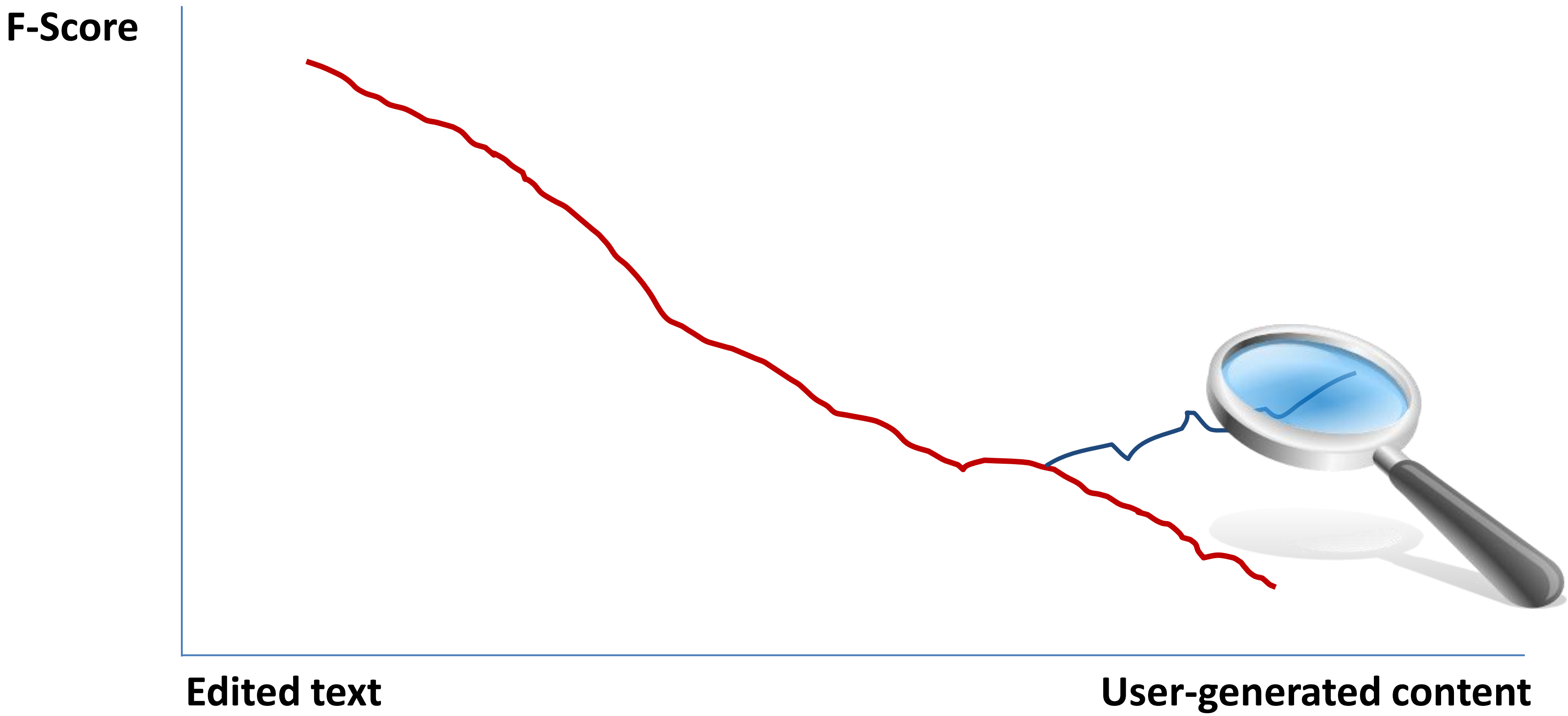
What are we doing?

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 heart beltran
On Fox: RNC chair sends letter to GOP calling Obama “ARROGANT” ” #tcot #sgp #hhrs
Twas okay.
FF > S4
LOL!
been playing with the new Canon EOS 500d and the Nikon D5000 over the weekend.
Man Utd through to the last 8…
I just think he looks like a big baby , and ppl USED to call him that
Bed soon

i heart beltran

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

Twas okay.

FF > S4

LOL!

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

Man Utd through to the last 8…

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

Bed soon

i heart beltran

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

Twas okay.

FF > S4

LOL!

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

Man Utd through to the last 8…

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

Bed soon

i heart beltran

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

Twas okay.

FF > S4

LOL!

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

Man Utd through to the last 8…

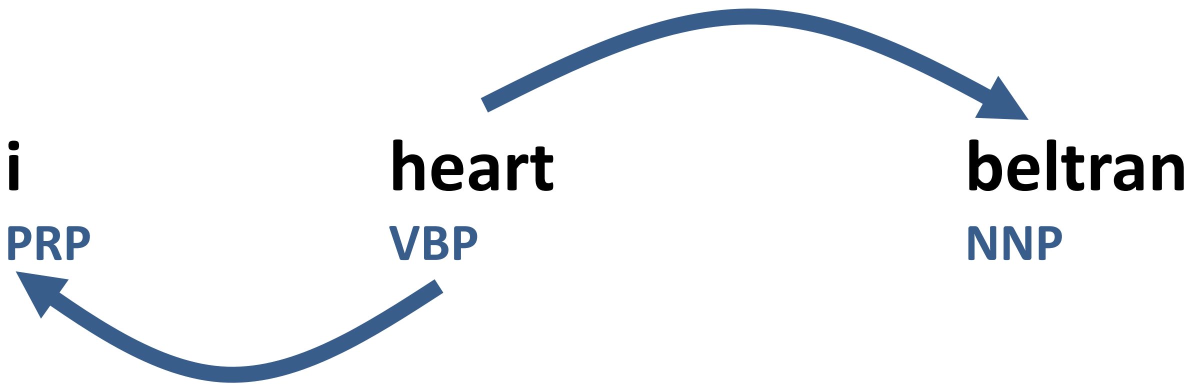
I just think he looks like a big baby , and ppl USED to call him that

Bed soon

Datasets

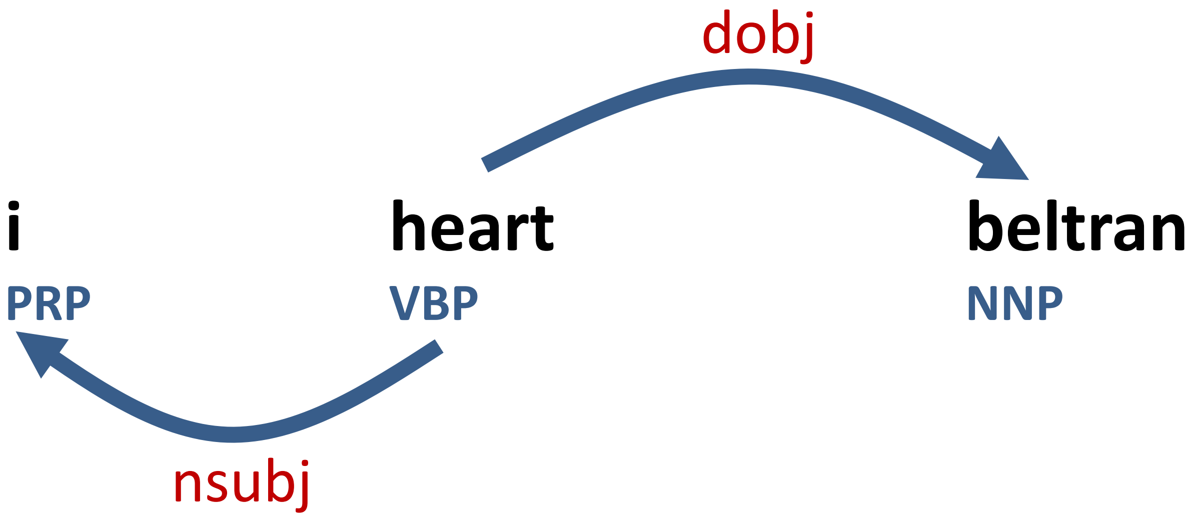
Corpus Name	#Sentences	Average Sentence Length	Median Sentence Length	Std. Deviation
TwitterDev	269	11.1	10	6.4
TwitterTest	250	11.3	10	6.8
TwitterTrain	1.4 million	8.6	7	6.1

Parser Output



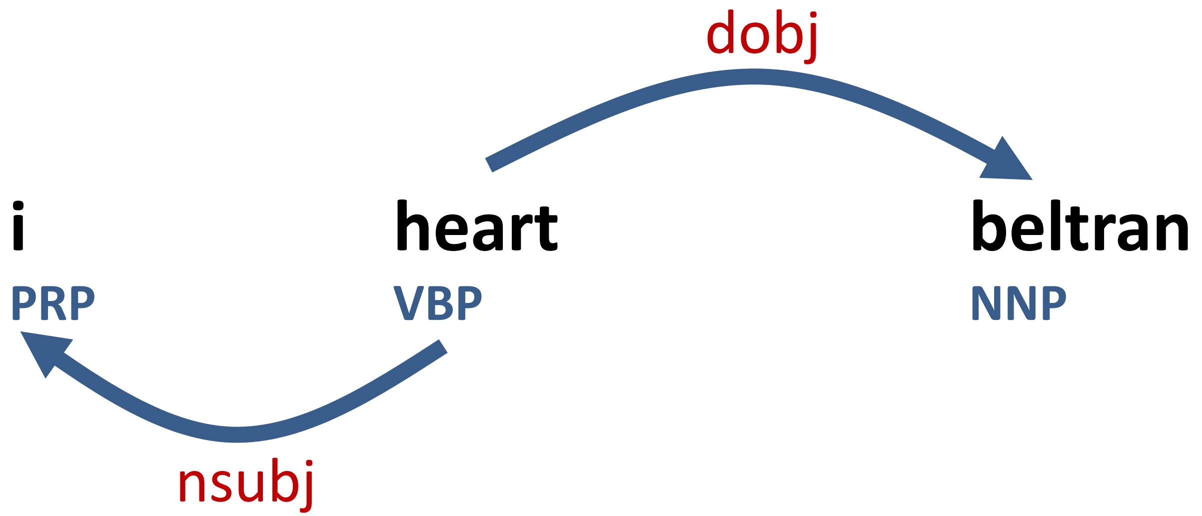
- Why *dependencies*?

Parser Output



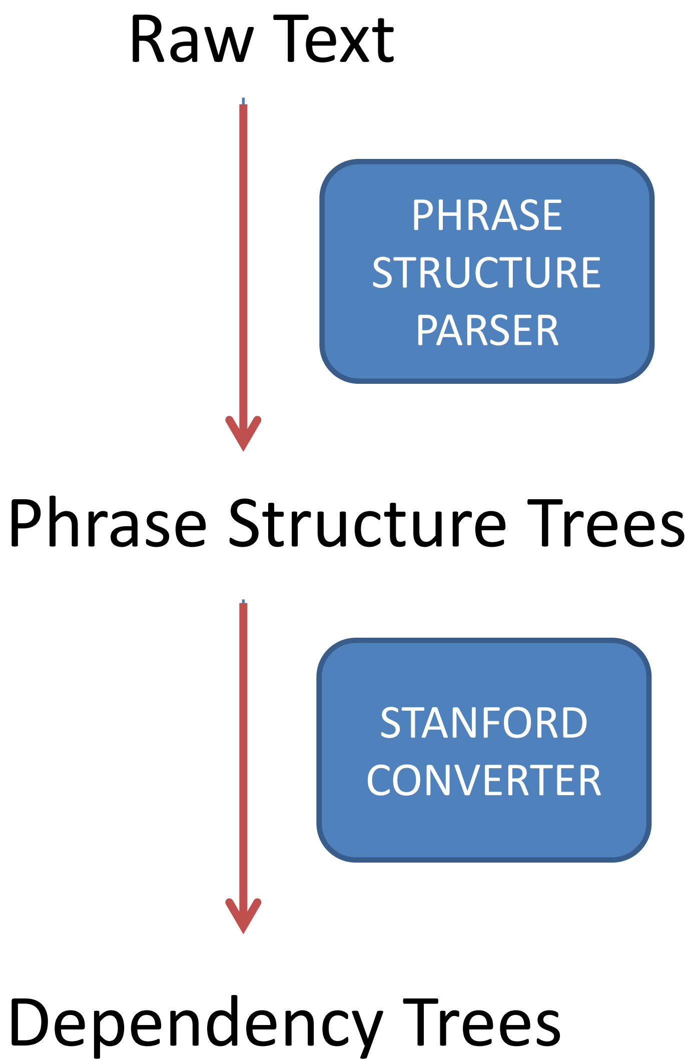
- Why *dependencies*?
- Why *labelled* dependencies?

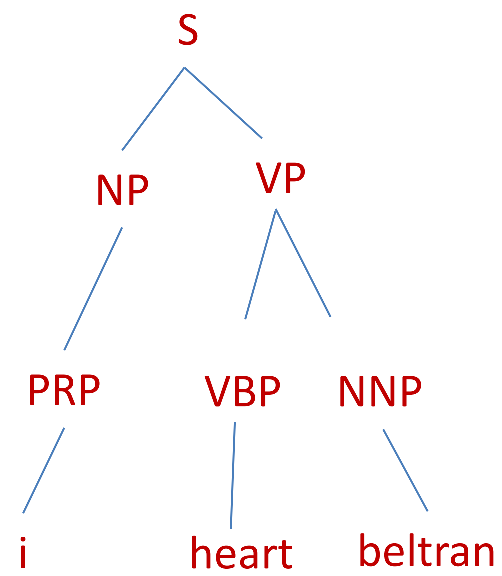
Parser Output



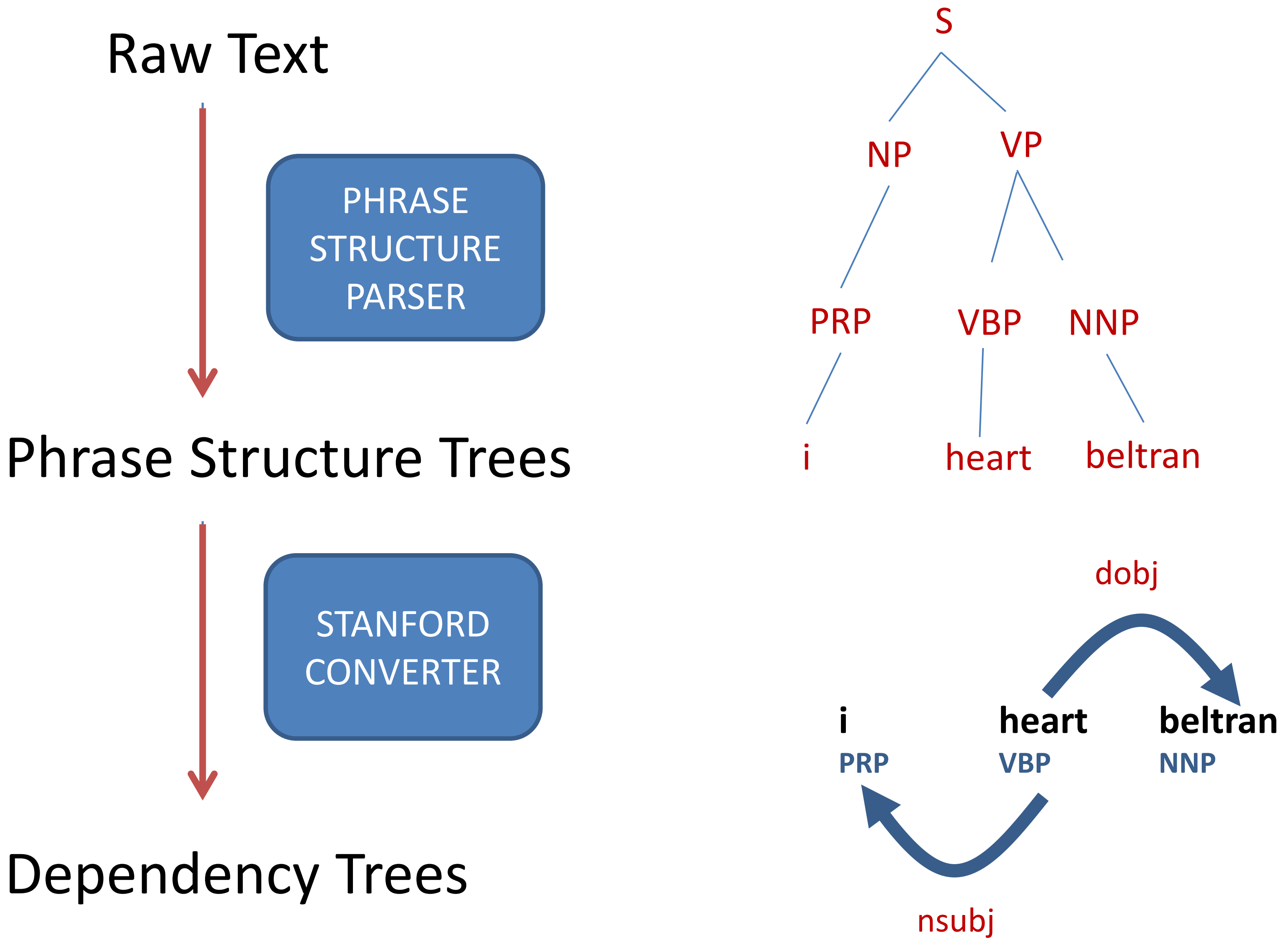
- Why *dependencies*?
- Why *labelled* dependencies?
- Why *Stanford* labelled dependencies?

Types of Parsers

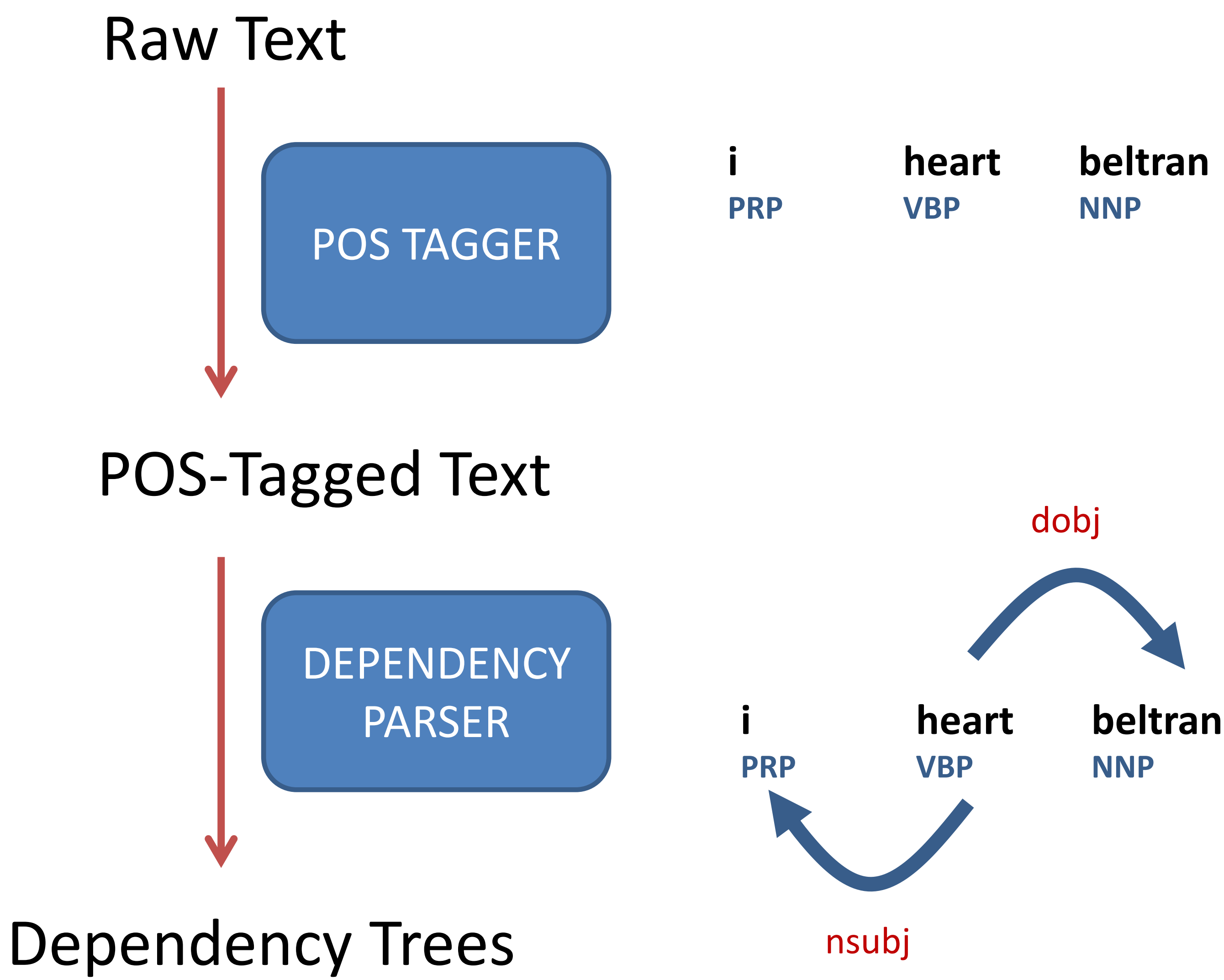




Types of Parsers



Types of Parsers



Baseline Model

SVMTool and MaltParser models are trained on Wall Street Journal trees.

Baseline POS Tagging Results

Corpus	Accuracy
WSJ22	96.3
TwitterDev	84.1

POS Tag Confusions

Gold/System	Frequency	Gold/System	Frequency
NNP/NN	59	VBZ/NNS	8
NN/NNP	54	UH/NNP	7
NNP/JJ	29	RB/NN	7
NNP/VB	10	NNP/CD	7
JJ/NN	10	NN/VB	6
UH/NN	8	VB/NN	6
JJ/NNP	8	VB/NNP	6
NNP/NNS	8	VBP/VB	6
NNPS/NNS	8	RP/IN	6

POS Tag Confusions

Gold/System	Frequency	Gold/System	Frequency
NNP/NN	59	VBZ/NNS	8
NN/NNP	54	UH/NNP	7
NNP/JJ	29	RB/NN	7
NNP/VB	10	NNP/CD	7
JJ/NN	10	NN/VB	6
UH/NN	8	VB/NN	6
JJ/NNP	8	VB/NNP	6
NNP/NNS	8	VBP/VB	6
NNPS/NNS	8	RP/IN	6

POS Tagging Examples

i	**heart**	**beltran**
FW	**NN**	**NN**

POS Tagging Examples

i	**heart**	**beltran**
FW	**NN**	**NN**

LOL	**!**
NNP	**.**

POS Tagging Examples

i	**heart**	**beltran**					
FW	NN	NN					

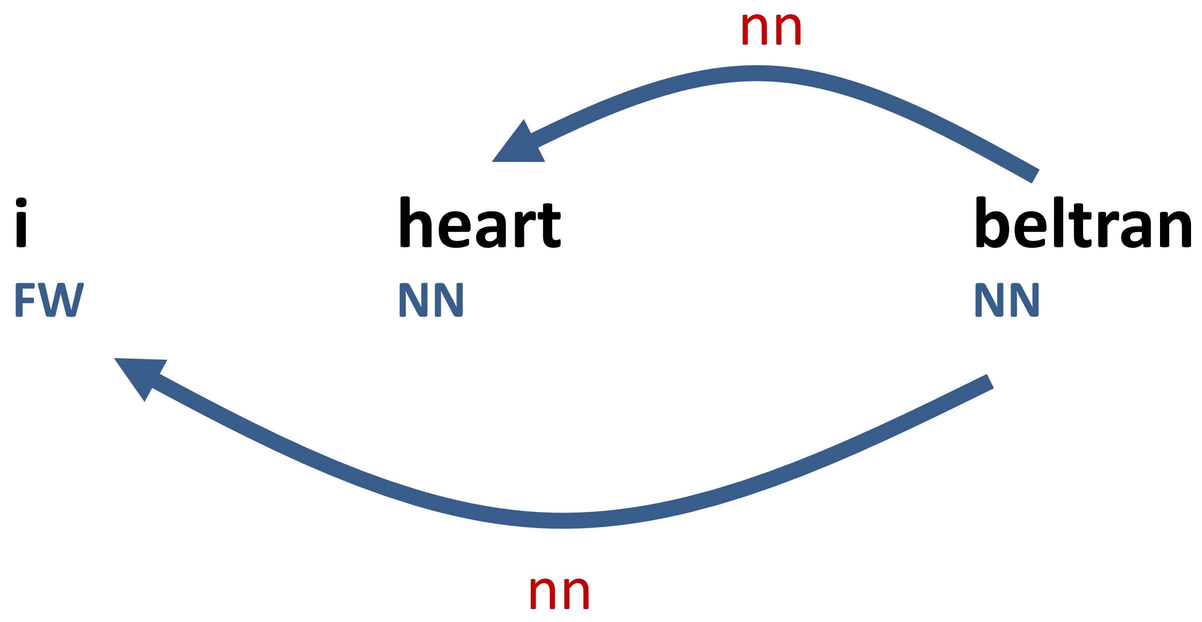
LOL	**!**
NNP	.

Man	**Utd**	**through**	**to**	**the**	**last**	**8**	**…**
NNP	NNP	IN	TO	DT	JJ	CD	:

Baseline Parsing Results

Corpus	Labelled/Unlabelled Attachment Accuracy
SVMTool POS Tags	
WSJ22	88.0/90.6
TwitterDev	67.3/73.6
Gold POS Tags	
WSJ22	90.0/91.6
TwitterDev	78.3/81.6

Parsing Example



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, *P1*, to provide training material for a less accurate parser, *P2* (Petrov et al. 2010)

- Why not just use *P1*?

  P2 is faster.

Two Types of Uptraining

- *Vanilla*

P1 = Charniak and Johnson (C&J) reranking parser

P2 = MaltParser

Two Types of Uptraining

- *Vanilla*

P1 = Charniak and Johnson (C&J) reranking parser

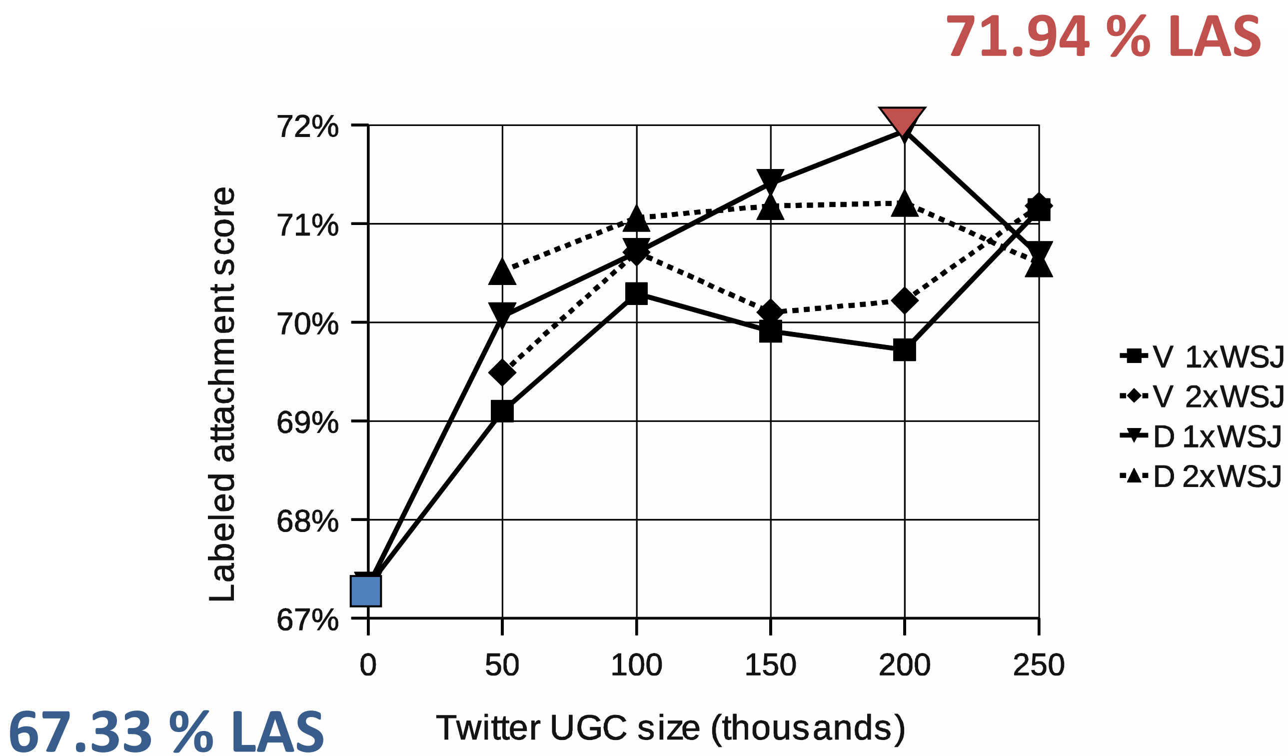
P2 = MaltParser

- *Domain-adapted*

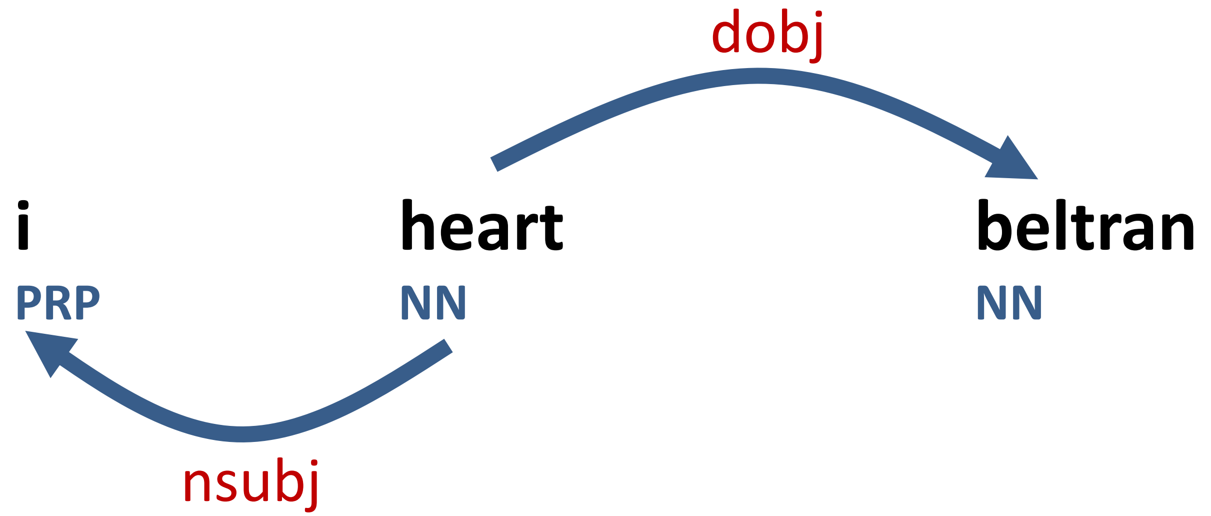
P1 = *Self-trained* version of C&J (McClosky et al. 2006)

P2 = MaltParser

Uptraining Results



Parsing Example



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**

Uptraining Error Analysis

- Uptraining 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**)

Uptraining Error Analysis

- Uptraining 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**)
- 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)

UNIVERSITÉ DE LA MÉDITERRANÉE AIX-MARSEILLE II

DCU

#thanksforlistening

UPPSALA UNIVERSITET

CNGL
Centre for Next Generation Localisation