DISFLUENCIES

A flight to $\underbrace{\text{um}}_{\text{FP}} \underbrace{\text{Berlin}}_{\text{RM}} \underbrace{\text{Imean}}_{\text{IM}} \underbrace{\text{Munich}}_{\text{RP}}$ on Tuesday

 $\hookrightarrow A$ flight to Munich on Tuesday

- ▶ BULATS learner English (<60s prompted test recordings);
- ▶ 'pilot' dataset: 1000 candidates, 19085 recordings;
- ▶ L1 Gujarati, Hindi, Urdu, Thai, Spanish, Portuguese, etc;
- CEFR proficiency levels $A2 \Leftrightarrow C2$.

- ▶ BULATS learner English (<60s prompted test recordings);
- ▶ 'pilot' dataset: 1000 candidates, 19085 recordings;
- ▶ L1 Gujarati, Hindi, Urdu, Thai, Spanish, Portuguese, etc;
- CEFR proficiency levels $A2 \Leftrightarrow C2$.

- ▶ BULATS learner English (<60s prompted test recordings);
- ▶ 'pilot' dataset: 1000 candidates, 19085 recordings;
- ▶ L1 Gujarati, Hindi, Urdu, Thai, Spanish, Portuguese, etc;
- CEFR proficiency levels $A2 \Leftrightarrow C2$.

- ► BULATS learner English (<60s prompted test recordings);
- ▶ 'pilot' dataset: 1000 candidates, 19085 recordings;
- ▶ L1 Gujarati, Hindi, Urdu, Thai, Spanish, Portuguese, etc;
- CEFR proficiency levels $A2 \Leftrightarrow C2$.

- ► BULATS learner English (<60s prompted test recordings);
- ▶ 'pilot' dataset: 1000 candidates, 19085 recordings;
- ▶ L1 Gujarati, Hindi, Urdu, Thai, Spanish, Portuguese, etc;
- CEFR proficiency levels $A2 \Leftrightarrow C2$.

- ▶ BULATS learner English (<60s prompted test recordings);
- crowdsourced transcriptions;
- ▶ gold-standard: hand-annotated 5667 token subset, errors & disfluencies;
- $\blacktriangleright~B1, B2, C1:$ c.1800 tokens each.

- ► BULATS learner English (<60s prompted test recordings);
- crowdsourced transcriptions;
- ▶ gold-standard: hand-annotated 5667 token subset, errors & disfluencies;
- ▶ B1, B2, C1: c.1800 tokens each.

- ► BULATS learner English (<60s prompted test recordings);
- crowdsourced transcriptions;
- ▶ gold-standard: hand-annotated 5667 token subset, errors & disfluencies;
- ▶ B1, B2, C1: c.1800 tokens each.

SPOKEN LANGUAGE: NATIVE SPEAKER

- ► Switchboard (SWB) U.S. English telephone conversations;
- ▶ Used in previous work –

System	Disfl.F
Zwarts & Johnson 2011 ACL	83.8
Qian & Liu 2013 NAACL	84.1
Rasooli & Tetreault 2014 EACL	82.6
Honnibal & Johnson 2014 (Redshift) TACL	84.1

Dependency parser: we have a buffer (the sentence to parse), a stack (starts empty), and we want to build a set of arc relations. We use 4(+1) operations: –

- ► Shift moves first word from buffer to stack
- ► REDUCE pops stack
- ► RIGHT ARC draws arc from stack to buffer, shifts buffer
- ▶ LEFT ARC draws arc from buffer to stack, pops stack
- ► (Edit ...)





































How are actions selected?

- ▶ selecting next action is a classification task,
- ▶ averaged perceptron to score action sequences,
- ▶ uses beam search to select next best step(s).

(Zhang & Clark 2008 EMNLP, Huang & Sagae 2010 ACL, Zhang & Nivre 2011 ACL)

How do we score each possible action?

- ▶ Features are defined by 73 templates from Zhang & Nivre (2011), plus extensions by H&J, which refer to 12 context tokens around word;
- ▶ e.g. $S_{0p}N_{0wp}$, POS-tag for first word on stack plus word and POS-tag for first word on buffer.

Introducing the EDIT action, to remove disfluent items : -

- ► Shift, Reduce, Right Arc, Left Arc
- ► EDIT on detection of disfluency, removes words and their dependencies.
- ▶ use extra features introduced by H&J to detect disfluency.

His1 company2 went3 broke4 i_mean5 went6 bankrupt7



















Following H&J:

- ► TASK 1: dependency relations
 - ► correctly identify head of each token
 ↔ UAS: unlabelled attachment score
- ► TASK 2: disfluency detection
 - correctly apply EDIT transition to RM tokens \hookrightarrow Disfl.F: disfluency F-measure = $2(\frac{p \times r}{r})$

Following H&J:

► TASK 1: dependency relations

▶ correctly identify head of each token
 → UAS: unlabelled attachment score

► TASK 2: disfluency detection

• correctly apply EDIT transition to RM tokens \hookrightarrow Disfl.F: disfluency F-measure = $2(\frac{p \times r}{r})$

Following H&J:

- ► TASK 1: dependency relations
 - ▶ correctly identify head of each token
 → UAS: unlabelled attachment score
- ► TASK 2: disfluency detection

• correctly apply EDIT transition to RM tokens \hookrightarrow Disfl. F: disfluency F-measure = $2(\frac{p \times r}{r})$

Following H&J:

- ► TASK 1: dependency relations
 - ► correctly identify head of each token
 → UAS: unlabelled attachment score
- ► TASK 2: disfluency detection

• correctly apply EDIT transition to RM tokens \hookrightarrow Disfl.*F*: disfluency *F*-measure = $2(\frac{p \times r}{p+r})$

Following H&J:

- ► TASK 1: dependency relations
 - ► correctly identify head of each token
 → UAS: unlabelled attachment score
- ► TASK 2: disfluency detection
 - correctly apply EDIT transition to RM tokens \hookrightarrow Disfl.F: disfluency F-measure = $2(\frac{p \times r}{p+r})$

EXPERIMENT: RESULTS

Treebank	Tokens	UAS	$\mathrm{Disfl.P/R}$	Disfl.F
SWB:TEST	45,405	90.5	n/a	84.1
SWB:dev	45,381	90.9	92.3/76.5	83.7
BULATS B1 B2 C1	5667 1895 1879 1893	91.1 88.9 91.2 93.0	82.6/33.6 85.3/31.4 79.2/33.2 83.8/37.3	47.8 45.9 46.8 51.6

And the people from twenty five to fifty years old

$$\underbrace{ \underset{\rm RM}{\text{they have the percentage of them}}}_{\rm RP}$$

 \hookrightarrow And the people from twenty five to fifty years old the percentage...

My colleagues will advise me

$\underbrace{\underbrace{\text{for the working stra they will guide for}_{RM}}_{\text{RM}}$

 \hookrightarrow My colleagues will advise me how to behave in the office

$\underbrace{\text{That}}_{\text{RM}} \underbrace{\text{I want they maybe they maybe help for my}}_{\text{RM}}$

 $\underbrace{\text{they maybe}}_{\text{RP}} \text{took my advice}$

 \hookrightarrow That they maybe took my advice

