Detection and Correction of Determiner and Preposition Errors

Task Overview

- Errors in determiners and prepositions
- **Focus**: 6 errors types – unnecessary (UD, UT), missing (MD, MT) and requiring replacement (RD, RT) determiners and prepositions
- **Data**: non-native, errorful, partially annotated
- **Evaluation**: detection, recognition and correction

Data and Preprocessing

Training Set

- 1000 files
- Metadata included
- Years 2000-2001
- 16 L1s
- Error rate = 44.18
- Gold standard not revised
- Standoff annotations provided

Test Set

- 100 files
- Metadata not included
- Years 1993-2009
- 27 L1s
- Error rate = 39.77
- Gold standard revised

Raw Training Data Format

Listening talks by the famous writers was a opportunity...

Standoff Annotation

```
<edit end="823" file="0519" index="0008" part="1" start="823" type="MT">
    <original> <empty/> </original>
    <corrections> <correction>to </correction> </corrections>
</edit>
<edit end="836" file="0519" index="0009" part="1" start="832" type="UD">
    <original>the </original>
    <corrections> <correction> <empty/> </correction> </corrections>
</edit>
```

Adopted In-line Format

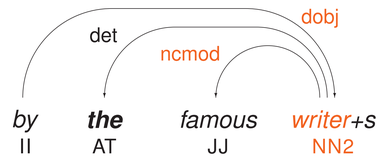
Raw data parsed with the RASP parser, error annotations added:

```
<lemma affix="ing" e="822" lem="Listen" pos="VVG" s="813" >Listening </lemma>
<lemma corEnd="823" corStart="823" corType="MT" corr="to "/>
<lemma affix="s" e="828" lem="talk" pos="NN2" s="823" >talks </lemma>
<lemma e="831" lem="by" pos="II" s="829" >by </lemma>
<lemma corEnd="836" corStart="832" corType="UD" corr="" e="835" lem="the"
pos="AT" s=" 832" >the </lemma>
<lemma e="842" lem="famous" pos="JJ" s="836" >famous </lemma>
<lemma affix="s" e="850" lem="writer" pos="NN2" s="843" >writers </lemma>
<lemma affix="ed" e="854" lem="be" pos="VBDZ" s="851" >was </lemma>
<lemma e="856" lem="a" pos="AT1" s="855" >a </lemma>
<lemma e="868" lem="opportunity" pos="NN1" s="857" >opportunity </lemma>
```

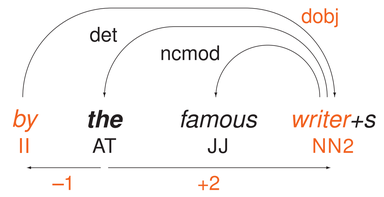
Feature Extraction

Determiners

High Precision system

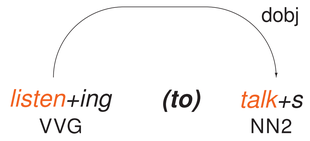


Naïve Bayes systems



Prepositions

All systems



High Precision Correction System

- **Aimed at** high precision
- **Types of errors**: MD and RT only
- **Method**: seen $\geqslant$ 2 times in the full CLC, $\geqslant$ 75% with an error

Results

- **Training set**

	Detection			Recognition			Correction		
	R	P	F	R	P	F	R	P	F
All	5.54	**81.08**	10.37	5.32	**77.95**	9.97	4.90	**71.70**	9.17
MD	8.74	**82.98**	15.82	8.74	**82.98**	15.82	8.39	**79.57**	15.17
RT	9.82	**75.59**	17.38	9.70	**74.71**	17.17	8.63	**66.47**	15.28

- **Test set**

	Detection			Recognition			Correction		
	R	P	F	R	P	F	R	P	F
All	4.86	**76.67**	9.15	4.65	**73.33**	8.75	4.65	**73.33**	8.75
MD	7.63	**83.33**	13.99	7.63	**83.33**	13.99	7.63	**83.33**	13.99
RT	8.05	**66.67**	14.37	8.05	**66.67**	14.37	8.05	**66.67**	14.37

Naïve Bayes Systems

- **Aimed at** detecting and correcting errors in:
  - **Det**: *none*, *the*, *a/an*
  - **Prep**: *none*, *in*, *of*, *for*, *to*, *at*, *with*, *on*, *about*, *from*, *by*, *after*.
- **Upper bound** for recall:
  - **Training set**: *recognition* = 91.95, *correction* = 86.24
  - **Test set**: *recognition* = 93.20, *correction* = 86.39
- **Method**: unadapted and adapted Naïve Bayes classifiers.

Results

Best-performing runs:

- `Run 3` – a combination of word-specific classifiers
- `Run 7` – unadapted, trained on the full CLC classifier

- **Training set**

		Detection			Recognition			Correction		
		R	P	F	R	P	F	R	P	F
All	3	50.09	27.54	35.52	45.99	25.23	32.57	28.78	15.80	20.39
Det	3	58.50	62.22	60.22	57.41	61.07	59.11	46.33	49.25	47.68
Prep	3	42.63	16.53	23.81	36.18	14.04	20.22	13.74	5.35	7.70
All	7	62.92	11.60	19.59	52.29	9.61	16.24	34.32	6.31	10.66
Det	7	58.65	8.11	14.24	53.90	7.43	13.06	40.61	5.60	9.84
Prep	7	65.68	16.89	26.87	50.87	13.09	20.82	28.92	7.44	11.84

- **Test set**

		Detection			Recognition			Correction		
		R	P	F	R	P	F	R	P	F
All	3	19.24	12.10	14.86	14.59	9.18	11.27	5.71	3.59	4.41
Det	3	6.11	11.29	7.93	5.24	9.68	6.80	5.24	9.68	6.80
Prep	3	29.10	11.31	16.28	23.36	9.08	13.07	6.15	2.39	3.44
All	7	56.66	11.59	19.24	42.29	8.69	14.43	27.27	5.58	9.26
Det	7	51.09	8.53	14.63	44.10	7.37	12.63	35.37	5.91	10.13
Prep	7	59.43	15.41	24.47	40.98	10.63	16.88	19.67	5.10	8.10

Discussion

- Differences in the training and test data → change in performance on the test set
- More consistent results are obtained for classifiers trained on full CLC (e.g., `Run 0` and `Run 7`) → more general training data helps classifiers generalize better

We thank Cambridge ESOL, a division of Cambridge Assessment for a partial grant to the first author and a research contract with iLexIR Ltd. We also thank them and Cambridge University Press for granting us access to the CLC for research purposes.

UNIVERSITY OF CAMBRIDGE

Ekaterina Kochmar (ek358@cl.cam.ac.uk)
Øistein E. Andersen (and@ilexir.co.uk)
Ted Briscoe (ejb@cl.cam.ac.uk)