

# CSLI in CoNLL-2013 Shared Task on Grammatical Error Correction

**Dan Flickinger and Jiye Yu**

CSLI, Stanford University

DELPH-IN Summit, St. Wendel

29 July 2013

# Overview of Shared Task

- Follows HOO 2011/2012 tasks on error detection and correction
- Development and test data from Cambridge Learner Corpus
- Five error types:
  - determiner
  - preposition
  - noun number
  - verb form
  - subject-verb agreement



# Sample development data

```
830 3 1 0 We PRP 3 nsubj (ROOT(S(S(NP*)
830 3 1 1 do VBP 3 aux (VP*
830 3 1 2 not RB 3 neg *
830 3 1 3 want VB 14 ccomp (VP*
830 3 1 4 to TO 5 aux (S(VP*
830 3 1 5 spend VB 3 xcomp (VP*
830 3 1 6 time NN 5 dobj (NP*
830 3 1 7 and CC 6 cc *
830 3 1 8 effort NN 6 conj *)
830 3 1 9 to TO 10 aux (S(VP*
830 3 1 10 remember VB 5 xcomp (VP*
830 3 1 11 password JJ 10 acomp (ADJP*))))))))))
```



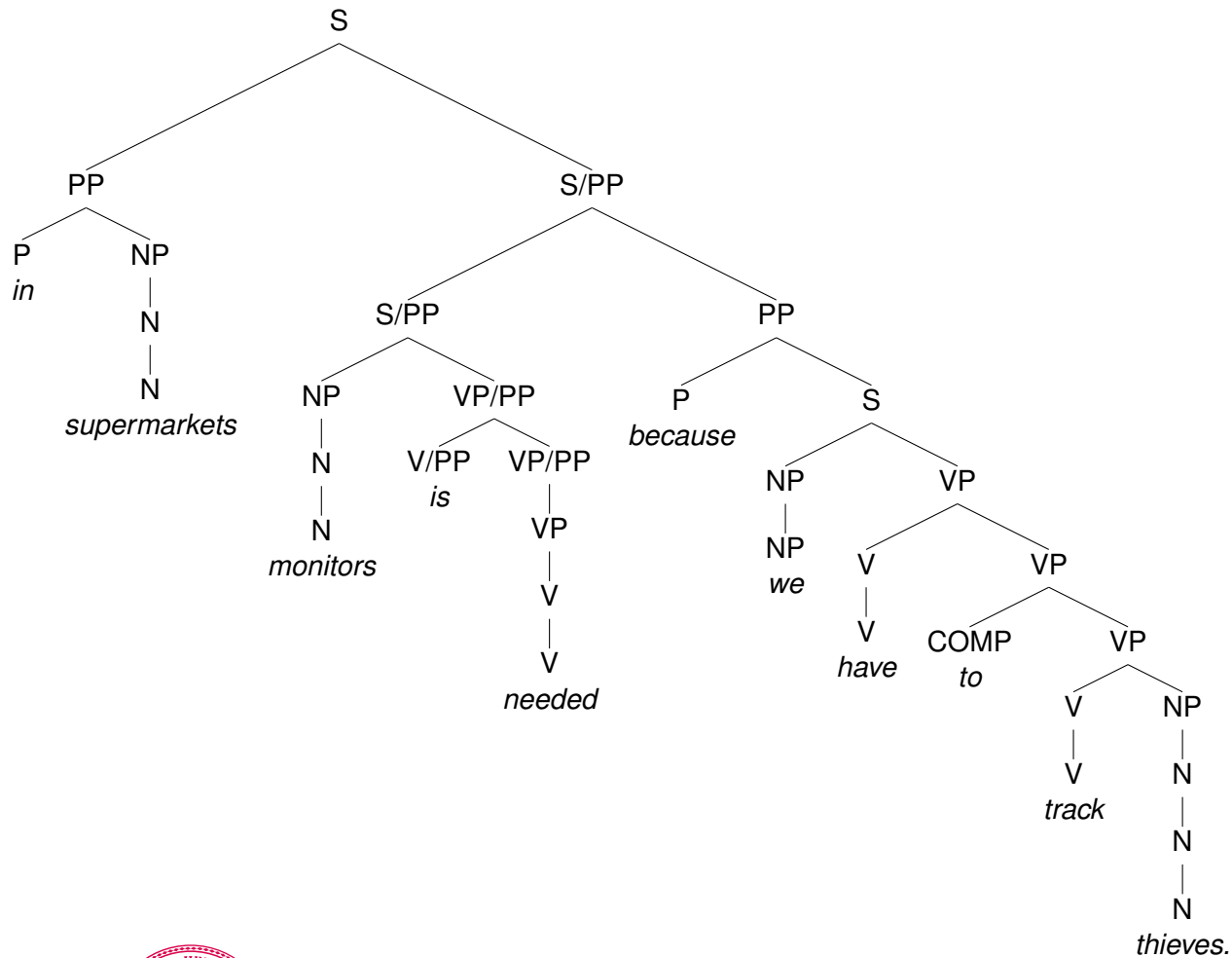
# Our method

- Convert CONLL format to one sentence/line, keeping tokenization
- Parse each sentence using PET with ERG-mal
- If robust and one of five error types, apply correction script
- Produce as output a 'corrected' human-readable document
- Evaluate using task-supplied tool (Dahlmeier & Ng 2012)  
Measures distance from hand-corrected output



# An example from the corpus

*In supermarkets monitors is needed because we have to track thieves .*



# Derivation tree

```
hd-aj_scp.c 0 11 [ root.robust.s ]
flr-hd_nwh-nc-pp.c 0 5
  hd-cmp_u.c 0 2
    in/p_np_i-reg 0 1 "in"
    hdn_bnp.c 1 2
      n_pl_olr 1 2
        supermarket_n1/n_-_c 1 2
          "supermarkets"
sb-hd_nmc.c 2 5
  hdn_bnp.c 2 3
    n_pl_olr 2 3
      monitor_n1/n_-_c 2 3 "monitors"
hd-cmp_u.c 3 5
  be.c.is.rbst 3 4 "is"
  hd_xaj-int-vp.c 4 5
    hd_optcmp.c 4 5
      v_pas_odlr 4 5
        need_v1/v_np 4 5 "needed"
hd-cmp_u.c 5 11
  because/p_cp_s 5 6 "because"
sb-hd_nmc.c 6 11
  hdn_bnp-qnt.c 6 7
    we/n_-_pr-we 6 7 "we"
hd-cmp_u.c 7 11
  v_n3s-bse_ilr 7 8
    have_to1/v_vp_ssr 7 8 "have"
hd-cmp_u.c 8 11
  to.c_prop/cm_vp.to 8 9 "to"
hd-cmp_u.c 9 11
  v_n3s-bse_ilr 9 10
    track_v1/v_np* 9 10 "track"
  hdn_bnp.c 10 11
    period_plr 10 11
      n_pl_olr 10 11
        thief_n1/n_-_c 10 11 "thieves."
```

# Sample output

. . .

In crime investigation , the truth of evidence is necessary , which can only be made by photos , videos or small tracks .

**In supermarkets monitors are needed because we have to track thieves .**

In front of ATMs we need monitors to avoid intentional crime .

We need CCTV in every public corner to track suspicious .

Only in this way can we get evidence of bad movement to protect others .

. . .



# Correction script (task-specific)

- For each ERG error code, effect the appropriate repair substitution, deletion, insertion
- If multiple candidate corrections, rank via N-gram-based resource
- If multiple errors in sentence, repairs apply in given order  
*The equipments have arrived => The equipment has arrived*
- Inflection changed output tokens using NLTK tools





# Results

Precision	25.50%
Recall	4.69 %
F1	7.92 %



## Results: Best case

Precision	25.50%
Recall	4.69 %
F1	7.92 %

In autopsy, we found two misanalyzed nominals with high frequency in test corpus:

*life expectancy* (91 occurrences) and *population aging/ageing* (40 occurrences)

After correction of these two and other minor correction script infelicities:

Precision	47.5%
Recall	13.2 %
F1	20.7 %



# Results for all systems

Rank	Team	R	P	F <sub>1</sub>
1	UIUC	23.49	46.45	31.20
2	NTHU	26.35	23.80	25.01
3	HIT	16.56	35.65	22.61
4	NARA	18.62	27.39	22.17
5	UMC	17.53	28.49	21.70
6	STEL	13.33	27.00	17.85
7	SJT1	10.96	40.18	17.22
8	CAMB	10.10	39.15	16.06
9	IITB	4.99	28.18	8.48
10	STAN	4.69	25.50	7.92
11	TOR	4.81	17.67	7.56
12	KOR	3.71	43.88	6.85
13	TILB	7.24	6.25	6.71
14	SZEG	3.16	5.52	4.02
15	UAB	1.22	12.42	2.22
16	SAAR	1.10	27.69	2.11
17	SJT2	0.24	13.33	0.48



# Affiliations

Team ID	Affiliation
CAMB	University of Cambridge
HIT	Harbin Institute of Technology
IITB	Indian Institute of Technology, Bombay
KOR	Korea University
NARA	Nara Institute of Science and Technology
NTHU	National Tsing Hua University
SAAR	Saarland University
SJT1	Shanghai Jiao Tong University (Team #1)
SJT2	Shanghai Jiao Tong University (Team #2)
STAN	Stanford University
STEL	Stellenbosch University
SZEG	University of Szeged
TILB	Tilburg University
TOR	University of Toronto
UAB	Universitat Autònoma de Barcelona
UIUC	University of Illinois at Urbana-Champaign
UMC	University of Macau



# Issue: Inconsistent manual markup of dev corpus

Comparing our system's output on student errors in first 500 sentences of the development corpus to the manual markup:

Alteration	# of Sentences
Both match	34
Missing gold	26
Differing correction	25
Wrong alteration	28



# Lessons learned

- Examine the quality of the task data carefully at the outset
- Choose the parts of the task relevant to research interests
- Allow sufficient time to engineer the task-specific dev/test harness
- For any future such task, develop a robust-inclusive treebank

