

Automatic Annotation and Evaluation of Error Types for Grammatical Error Correction

Christopher Bryant, Mariano Felice, Ted Briscoe

Cambridge ALTA

Institute for Automated Language Teaching and Assessment



UNIVERSITY OF
CAMBRIDGE

Overview

- Grammatical Error Correction (GEC)

He only can look at the TV in the night.

He can only watch TV at night.

Overview

- Grammatical Error Correction (GEC)

He only can look at the TV in the night.

He can only watch TV at night.

- Annotated Data

He	only can	look at	the	TV	in	the	night .
He	can only	watch	∅	TV	at	∅	night .
	R:WO	R:VERB	U:DET		R:PREP	U:DET	

- How can we annotate automatically?

Motivation

1. Facilitates system error type evaluation
 - Unannotated hypotheses vs. annotated references
2. Standardise datasets
 - FCE (80 types), NUCLE (28 types), Lang-8 (0 types)
3. Reduce annotator burden
4. Feedback to learners/teachers

1. Edit Extraction

- *Felice et al. 2016. Automatic extraction of learner errors in ESL sentences using linguistically enhanced alignments. COLING.*

Orig	He	only	can	look	at	the	TV	in	the	night	.
Corr	He	can	only	watch	TV	at	night	.			

1. Edit Extraction

- *Felice et al. 2016. Automatic extraction of learner errors in ESL sentences using linguistically enhanced alignments. COLING.*

Orig He only can look at the TV in the night .
Corr He can only watch TV at night .

- Levenshtein

Orig	He	only	can	look	at	the	TV	in	the	night	.
Corr	He	∅	can	∅	only	watch	TV	∅	at	night	.

1. Edit Extraction

- Felice et al. 2016. Automatic extraction of learner errors in ESL sentences using linguistically enhanced alignments. COLING.

Orig He only can look at the TV in the night .
Corr He can only watch TV at night .

- Levenshtein

Orig He only can look at the TV in the night .
Corr He \emptyset can \emptyset only watch TV \emptyset at night .

- Damerau-Levenshtein, linguistic features, merging rules

Orig He only can look at the TV in the night .
Corr He can only watch \emptyset TV at \emptyset night .

- Performance on several datasets: **~85% F1**

2. Edit Classification

- ~50 rules based on automatic markup

Original	Corrected	Type	Rule Info
the	∅	U:DET	POS
∅	in	M:PREP	POS
cat	dog	R:NOUN	POS, Lemma
cat	cats	R:NOUN:NUM	POS, Lemma
eats	has eaten	R:VERB:TENSE	POS, Lemma, Parse
atack	attack	R:SPELL	Wordlist

- Dataset agnostic
- Less subjective

2. Edit Classification

		Operation Tier			
		Type	Missing	Unnecessary	Replacement
Token Tier	Part Of Speech	Adjective	M:ADJ	U:ADJ	R:ADJ
		Adverb	M:ADV	U:ADV	R:ADV
		Conjunction	M:CONJ	U:CONJ	R:CONJ
		Determiner	M:DET	U:DET	R:DET
		Noun	M:NOUN	U:NOUN	R:NOUN
		Particle	M:PART	U:PART	R:PART
		Preposition	M:PREP	U:PREP	R:PREP
		Pronoun	M:PRON	U:PRON	R:PRON
		Punctuation	M:PUNCT	U:PUNCT	R:PUNCT
		Verb	M:VERB	U:VERB	R:VERB
Token Tier	Other	Contraction	M:CONTR	U:CONTR	R:CONTR
		Morphology	—	—	R:MORPH
		Orthography	—	—	R:ORTH
		Other	M:OTHER	U:OTHER	R:OTHER
		Spelling	—	—	R:SPELL
		Word Order	—	—	R:WO
Morphology Tier	Adjective Form	—	—	R:ADJ:FORM	
	Noun Inflection	—	—	R:NOUN:INFL	
	Noun Number	—	—	R:NOUN:NUM	
	Noun Possessive	M:NOUN:POSS	U:NOUN:POSS	R:NOUN:POSS	
	Verb Form	M:VERB:FORM	U:VERB:FORM	R:VERB:FORM	
	Verb Inflection	—	—	R:VERB:INFL	
	Verb Agreement	—	—	R:VERB:SVA	
	Verb Tense	M:VERB:TENSE	U:VERB:TENSE	R:VERB:TENSE	

2. Edit Classification

- 5 participants
- 200 edits in context (100 FCE, 100 CoNLL-2014)

Social network sites provides
provide us with many convenience
conveniences .

R:VERB:SVA - Replacement, subject-verb agreement

GOOD

ACCEPTABLE

BAD

2. Edit Classification

- 5 participants
- 200 edits in context (100 FCE, 100 CoNLL-2014)

Social network sites provides
provide us ~~with~~ many convenience
conveniences .

R:VERB:SVA - Replacement, subject-verb agreement

GOOD

ACCEPTABLE

BAD

• Results

Judgement	1	2	3	4	5	OVERALL
Good	92.00%	89.50%	83.00%	84.50%	82.50%	86.30%
Acceptable	4.00%	6.50%	13.00%	11.00%	15.50%	10.00%
Bad	4.00%	4.00%	4.00%	4.50%	2.00%	3.70%

CoNLL-2014 Shared Task

- 12 Teams: Rules, LMs, Classifiers, SMT, hybrid
- M2 Scorer (F0.5), I-measure, GLEU

Hypothesis	Reference
This are a grammatical sentences .	S This is grammatical sentences .
	A 2-2 Det a
	A 3-4 Nn sentence

Key: TP FP FN

CoNLL-2014 Shared Task

- 12 Teams: Rules, LMs, Classifiers, SMT, hybrid
- M2 Scorer (F0.5), I-measure, GLEU

Hypothesis	Reference
This are a grammatical sentences .	S This is grammatical sentences .
	A 2-2 Det a
	A 3-4 Nn sentence

- Our Proposal

Key: **TP** **FP** **FN**

Hypothesis	Reference
S This is grammatical sentences .	S This is grammatical sentences .
A 1-2 R:VERB:SVA are	A 2-2 M:DET a
A 2-2 M:DET a	A 3-4 R:NOUN:NUM sentence

CoNLL-2014 Shared Task

	CAMB	CUUI	AMU
Approach	Hybrid SMT	Classifiers	SMT
Official	37.33	36.79	35.01

All scores are F0.5

CoNLL-2014 Shared Task

	CAMB	CUUI	AMU
Approach	Hybrid SMT	Classifiers	SMT
Official	37.33	36.79	35.01
M	41.43	24.18	31.08
R	35.18	38.27	34.57
U	25.88	34.02	-
SPELL	66.20	0.00	73.89
PREP	40.14	25.76	34.15
VERB:SVA	41.20	57.53	42.45

All scores are F0.5

CoNLL-2014 Shared Task

	CAMB	CUUI	AMU
Approach	Hybrid SMT	Classifiers	SMT
Official	37.33	36.79	35.01
M	41.43	24.18	31.08
R	35.18	38.27	34.57
U	25.88	34.02	-
SPELL	66.20	0.00	73.89
PREP	40.14	25.76	34.15
VERB:SVA	41.20	57.53	42.45

All scores are F0.5

	CAMB			CUUI		
	P	R	F0.5	P	R	F0.5
M:DET	43.20	51.77	44.68	23.86	45.00	26.34
R:DET	19.33	35.37	21.26	27.03	24.39	26.46
U:DET	43.75	39.90	42.92	36.19	66.37	39.81

Summary

- ERRor ANnotation Toolkit (ERRANT)
 - <https://github.com/chrisjbryant/errant>
- Uses:
 - Annotate parallel English sentences
 - Detailed error type evaluation
 - Dataset standardisation

Summary

- ERRor ANnotation Toolkit (ERRANT)
 - <https://github.com/chrisjbryant/errant>
- Uses:
 - Annotate parallel English sentences
 - Detailed error type evaluation
 - Dataset standardisation

Questions?