Automatic Annotation and Evaluation of Error Types for Grammatical Error Correction

#### Christopher Bryant, Mariano Felice, Ted Briscoe



Institute for Automated Language Teaching and Assessment



#### **Overview**

#### Grammatical Error Correction (GEC)

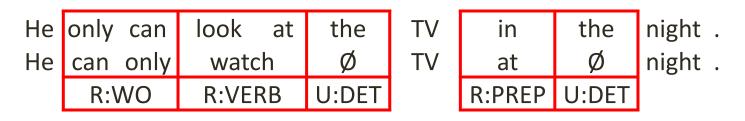
He only can look at the TV in the night. He can only watch TV at night.

#### **Overview**

Grammatical Error Correction (GEC)

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

#### Annotated Data



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	ΤV	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	Не	can	only	watch	TV	at	night	•			

• Levenshtein

Orig	He	only	can	look	at	the watch	ΤV	in	the	night	
Cor	r He	Ø	can	Ø	only	watch	ΤV	Ø	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	Не	only	can	look	at	the watch	ΤV	in	the	night	
Corr	Не	Ø	can	Ø	only	watch	ΤV	Ø	at	night	

• Damerau-Levenshtein, linguistic features, merging rules

Orig	He	only	can	look	at	the	ΤV	in	the	night	•
Corr	He	can	only	wat	ch	Ø	ΤV	at	Ø	night	

Performance on several datasets: ~85% F1

~50 rules based on automatic markup

Original	Corrected	Туре	Rule Info
the	Ø	U:DET	POS
Ø	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

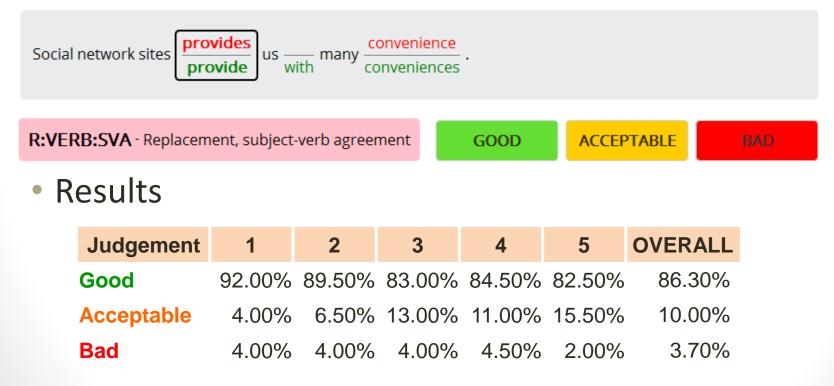
- Dataset agnostic
- Less subjective

	True			<b>Operation Tier</b>	
		Туре	Missing	Unnecessary	Replacement
		Adjective	M:ADJ	U:ADJ	R:ADJ
		Adverb	M:ADV	U:ADV	R:ADV
	Ę	Conjunction	M:CONJ	U:CONJ	R:CONJ
	Speech	Determiner	M:DET	U:DET	R:DET
		Noun	M:NOUN	U:NOUN	R:NOUN
	of	Particle	M:PART	U:PART	R:PART
ier	Part	Preposition	M:PREP	U:PREP	R:PREP
Token Tier	à	Pronoun	M:PRON	U:PRON	R:PRON
kei		Punctuation	M:PUNCT	U:PUNCT	R:PUNCT
5		Verb	M:VERB	U:VERB	R:VERB
		Contraction	M:CONTR	U:CONTR	R:CONTR
		Morphology	_	_	R:MORPH
	Other	Orthography	—	—	R:ORTH
	ot	Other	M:OTHER	U:OTHER	R:OTHER
		Spelling	_	_	R:SPELL
		Word Order	—	—	R:WO
		Adjective Form	—	—	R:ADJ:FORM
:	ler	Noun Inflection	—	—	R:NOUN:INFL
	2	Noun Number	—	—	R:NOUN:NUM
	8010	Noun Possessive	M:NOUN:POSS	U:NOUN:POSS	R:NOUN:POSS
	ouo	Verb Form	M:VERB:FORM	U:VERB:FORM	R:VERB:FORM
	Morphology Lier	Verb Inflection	—	—	R:VERB:INFL
	Σ	Verb Agreement	_	_	R:VERB:SVA
		Verb Tense	M:VERB:TENSE	U:VERB:TENSE	R:VERB:TENSE

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



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



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

Hypothesis		R	eference			
This are a grammatical sentences .	S	5 This is grammatical sentences				
	Α	2-2 Det	а			
	Α	3-4 Nn	sentence			
			Kev:	ТР	FP	FN

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

Hypothesis	Reference						
This are a grammatical sentences .		This is grammatical senter 2-2 Det a	nces .				
		3-4 Nn sentence					
<ul> <li>Our Proposal</li> </ul>		Key: TP	FP FN				
Hypothesis		Reference					
<b>S</b> This is grammatical sentences .	S	This is grammatical sente	nces .				
A 1-2 R:VERB:SVA are	Α	2-2 M:DET a					
A 2-2 M:DET a		3-4 R:NOUN:NUM sente					

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

All scores are F0.5

14

	CAMB	CUUI	AMU
Approach	Hybrid SMT	Classifiers	SMT
Official	37.33	36.79	35.01
Μ	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 scor	es are F0.5

	CAMB	CUUI	AMU		
Approach	Hybrid SMT	Classifiers	SMT		
Official	37.33	36.79	35.01		
Μ	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		
	Р	R	F0.5	Р	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?