

# 訂正難易度を考慮した文法誤り訂正のための性能評価尺度



五藤 巧<sup>†</sup>, 永田 亮<sup>† ‡</sup>, 三田雅人<sup>‡</sup>, 塙 一晃<sup>‡</sup> <sup>† 甲南大学</sup> <sup>‡ 理研</sup>

## 1. 目的



学習者の英文: It has been more easy  
訂正後の英文: It was easier

訂正 難

訂正 易



訂正難易度が異なる  
問題が混在している!

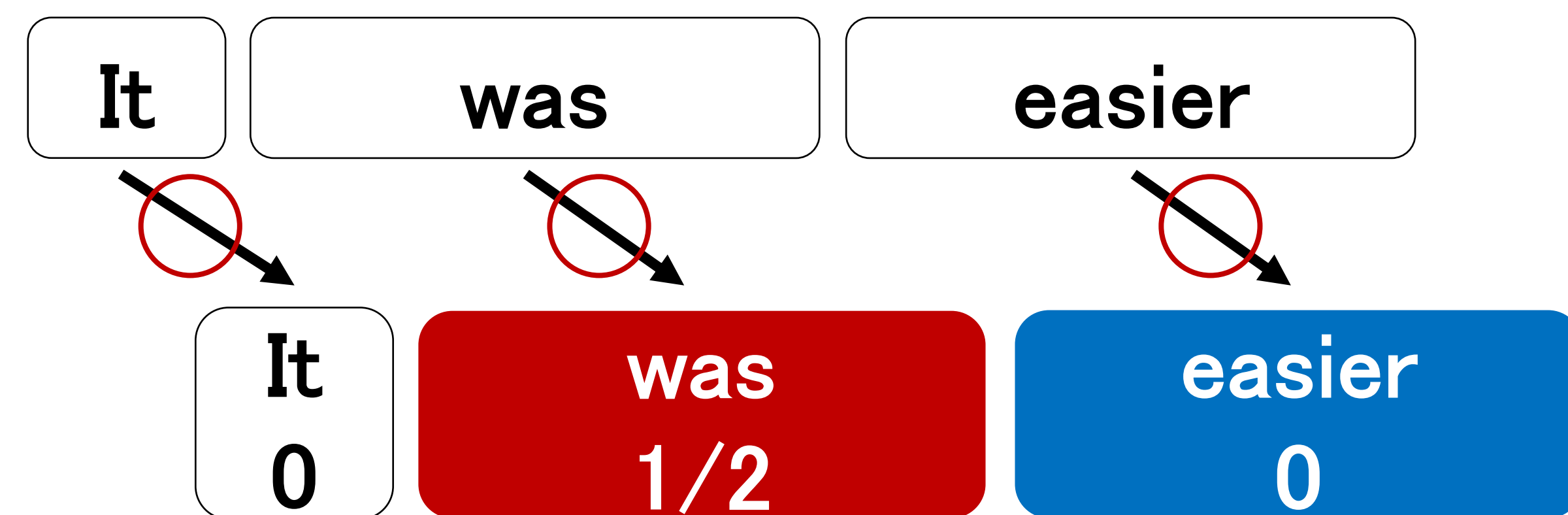
訂正の難しさを考慮した  
評価尺度が欲しい!!

## 2. 基本アイデア

訂正成功率を元に訂正難易度( $w$ )を定義:  $w = 1 - \text{訂正成功率}$

例: 2システムの場合

システム1の出力



正解文  
訂正難易度 $w$

システム2の出力



$$\text{was} \quad 1 - \frac{1}{2} = \frac{1}{2} \quad \text{訂正成功率} \quad w$$

$$\text{easier} \quad 1 - \frac{2}{2} = 0 \quad \text{訂正成功率} \quad w$$

## 3. 提案評価尺度

$$\text{Recall} = \frac{\sum_{i \in E} w_i l_i}{\sum_{i \in E} w_i}$$

$$\text{Precision} = \frac{\sum_{i \in E} w_i l_i}{\sum_{i \in C} w_i}$$

$l = 1$ (訂正に成功) or  $0$ (失敗)  $E = \text{全誤りの集合}$   $C = \text{訂正箇所の集合}$

## 4. 実験結果と考察

複数システムによるコーパス横断評価

訂正難易度を考慮した  $F_{0.5}$

CoNLL-2013	CoNLL-2014	KJ	ICNALE
Transformer 18.68	Transformer 15.17	Transformer 18.33	Transformer 18.17
SMT 15.16	SMT 13.66	CNN 17.39	LSTM 15.16
CNN 12.32	LSTM 11.01	LSTM 16.88	CNN 14.56
LSTM 11.94	CNN 9.75	SMT 8.51	SMT 12.88

従来( $M^2$ )の  $F_{0.5}$

CoNLL-2013	CoNLL-2014	KJ	ICNALE
Transformer 36.20	Transformer 48.62	LSTM 45.64	LSTM 43.02
LSTM 33.76	LSTM 48.57	CNN 45.40	CNN 40.78
CNN 33.67	SMT 46.80	Transformer 42.80	Transformer 37.72
SMT 32.30	CNN 46.16	SMT 32.04	SMT 32.91

訂正難易度の可視化の例

This **had** caused panic among the people who **had** flooded **the** local police department with ...

Personally I am **study** in overseas **busy** study life keeps me away from contact my old **friend**.

(ともにCoNLL-2014より一部抜粋)

誤りの種類別の訂正難易度の例

誤りの種類	平均 $w$	標準偏差
形容詞	0.982	0.074
動詞	0.891	0.254
時制	0.876	0.213
冠詞	0.747	0.292
動詞の活用	0.590	0.393
名詞の数	0.539	0.340
綴り誤り	0.533	0.342

## 5. まとめ

- ・訂正難易度を考慮した性能評価尺度を提案
  - ・得られた訂正難易度は直感に一致
  - ・システム間の差が明確に, 順位が安定
- (本研究にかかるツールを公開)

