

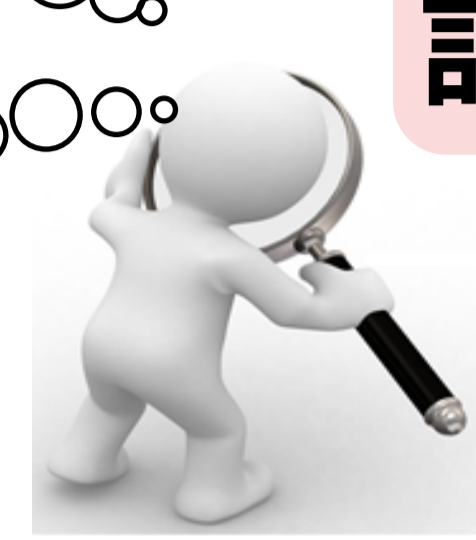
文法誤り訂正における問題の難しさを考慮した性能評価尺度の提案

五藤 巧[†], 永田 亮^{†‡}, 三田雅人^{‡‡}, 埜 一晃[‡], 水本智也^{‡‡}
[†]甲南大学 [‡]理研AIP ^{‡‡}東北大学 ^{‡‡‡}現所属はフューチャー株式会社

目的

訂正難度が異なる問題が混在している!

にも関わらず、一律に評価...



学習者の英文: It **has been** **more easy**

訂正 難

訂正 易

訂正後の英文 It **was** **easier**

訂正の難しさを考慮した評価尺度が欲しい!!



基本アイデア

人の試験の場合



正答率

低い → 難

訂正に成功したシステム数に基づいた重み

$$\text{重み: } w = 1 - \frac{\text{訂正に成功したシステム数}}{\text{全システム数}}$$

正答率

例: It **was** **easier**

	was	easier
システムA	○	○
システムB	×	○
システムC	×	○
正答率	1/3	1
w_i	2/3	0

提案評価尺度

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

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

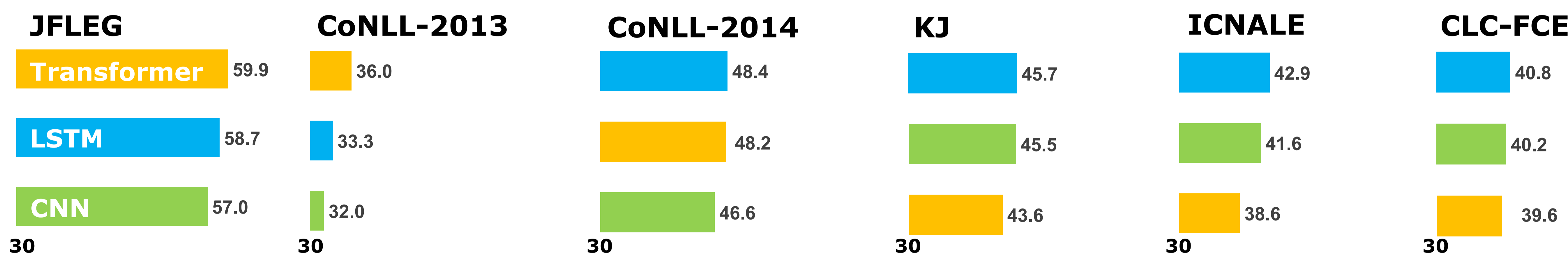
$l = 1$ (訂正に成功) or 0 (失敗)

$E =$ 全誤りの集合

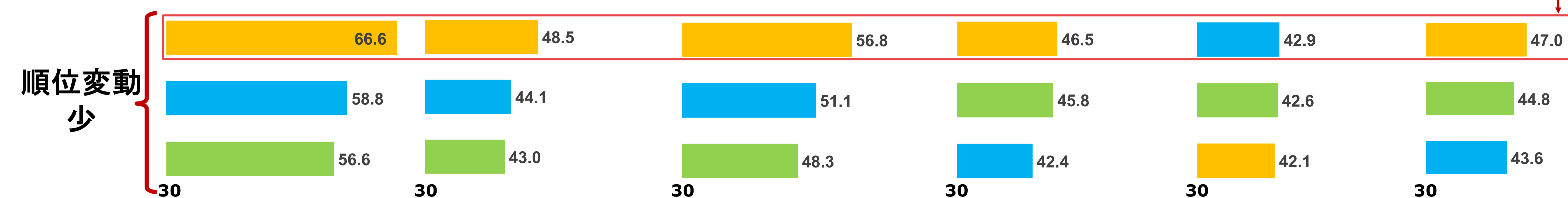
$C =$ 訂正箇所集合

実験結果と考察

従来 $F_{0.5}$



難易度を考慮した $F_{0.5}$



※現在は通常のprecisionを用いて $F_{0.5}$ を計算

Transformerは他のシステムと異なる訂正傾向を持つ?

まとめと課題

- 訂正の難しさを考慮した評価尺度を提案
- 以下の2点を確認
 - 差が明確化
 - 順位変動少 の傾向

- 重み付きPrecisionへの拡張
- システム数を増やして調査
- システムが得意・不得意とする訂正の分析
- 英文中における、訂正難易度の可視化へ応用