

文法誤りの訂正難易度の判別における人とシステムの一一致調査

五藤 巧[†], 永田 亮[‡], 三田雅人[§] [†]奈良先端大 [‡]甲南大 [§]理研 [¶]東北大

1. 背景と目的

背景

- Gotou 2020+ が訂正難易度を考慮した評価尺度を提案した
- 人手の訂正難易度との一致は報告されていない

明らかにしたいこと

- 自動評価の訂正難易度はどれくらい人と一致するのか？
また、どのような場面で使うべきか？

2. 訂正難易度の自動計算

- より多くのシステムが解けた誤りほど、訂正が易しいとみなす
- 既存システムの出力を用いてN段階で誤りの訂正難易度を計算

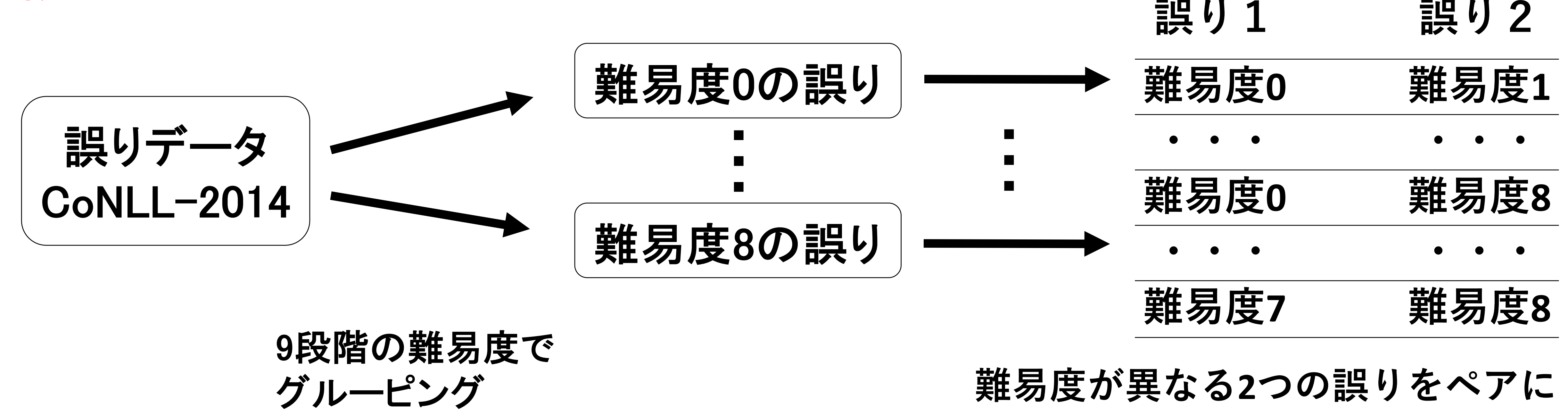
3. 調査方法

誤りペアのうち難しいほうを判別

誤り 1	判別結果	誤り 2
... has → have ...	>	... aet → eat ...
... eat → ate ...	=	... do → did ...
we → We ...	<	He → His ...

- 提示されたペアを人(2人)が評価
- 判別結果の一致率を算出

誤りペアの作成

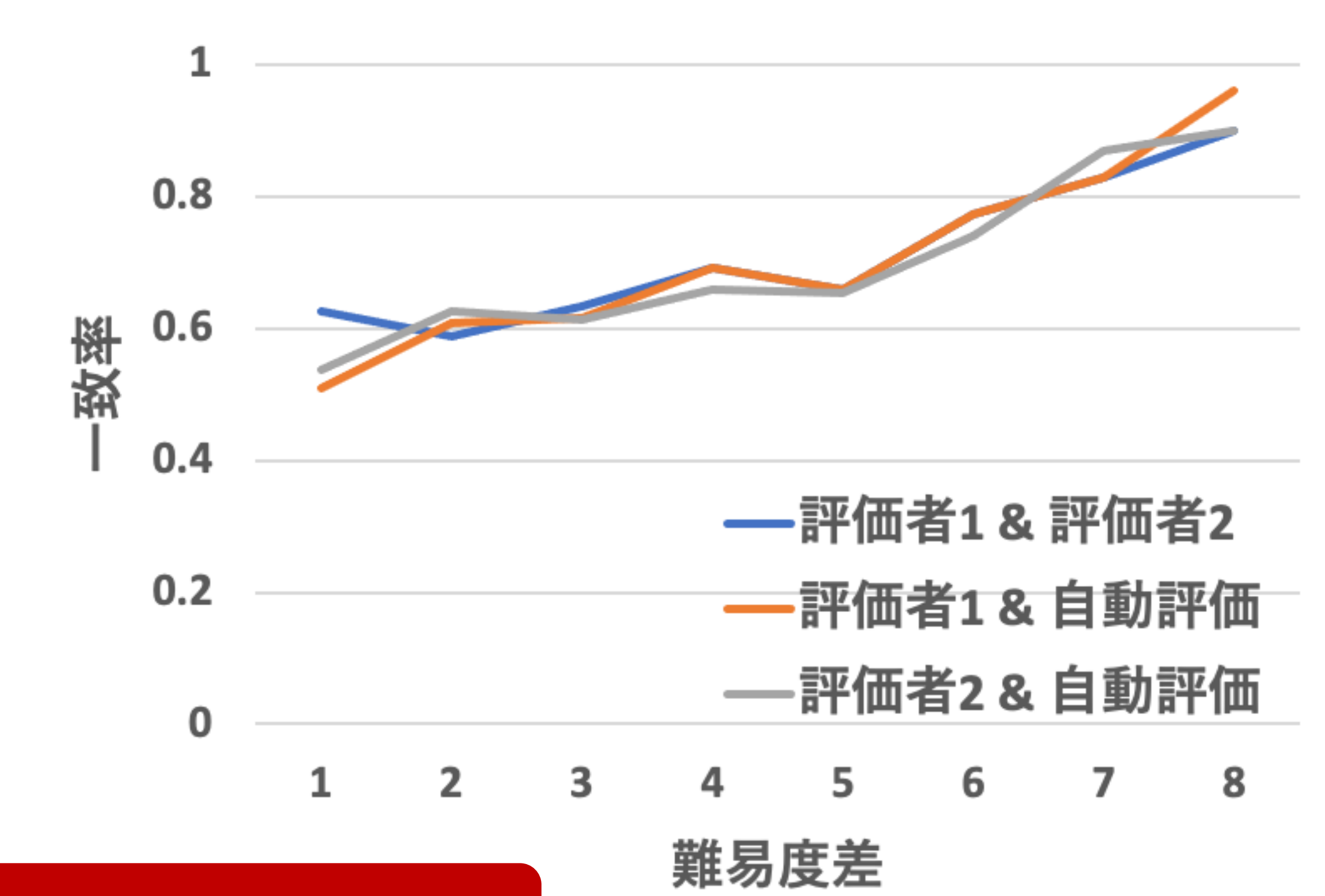


4. 調査結果

全体的な一致率

比較	一致率
評価者1 & 評価者2	66.39
評価者1 & 自動評価	64.72
評価者2 & 自動評価	64.28

難易度差ごとの一致率



人の評価と自動評価の一致を示唆

誤りタイプの一一致率

比較 \ 誤りタイプ	スペル	分割・統合
評価者1 & 評価者2	0.77	0.30
評価者1 & 自動評価	0.54	0.45
評価者2 & 自動評価	0.61	0.70

- スペル誤りの不一致の例
... the importance of the face-to-face human contacts has **decreased** → **decreased**.
- 分割・統合誤りの不一致な実例
... to make informed decisions and **tosimilarly** → **to similarly** undergo genetic...

- スペルミスや統合による誤りは未知語になるため、自動訂正が困難。
- 未知語になる語の候補は訓練データ依存。一方、人はそれを考慮することが難しいため、不一致の原因に。
- 自動評価の難易度は、少なくともシステムの解ける/解けないの尺度を表す。
→ Shared Taskにおいて定性的な観点から改善点を分析することに有効。

5. まとめ

- 訂正難易度についてシステムと人的一致を調査。
- 人と自動評価は一致することを示唆。
- 自動評価の訂正難易度は、Shared Taskにおける分析で有効。