

2020.11.27 教育アセスメント×言語処理シンポジウム

ライティング学習支援のための 文法誤り訂正技術の現状と今後の展望

理研AIP 自然言語理解チーム
東北大学情報科学研究科 博士課程
三田雅人

文法誤り訂正とは

入力文に含まれる文法誤りを自動訂正するタスク



The Machine is design to help people.



訂正システム



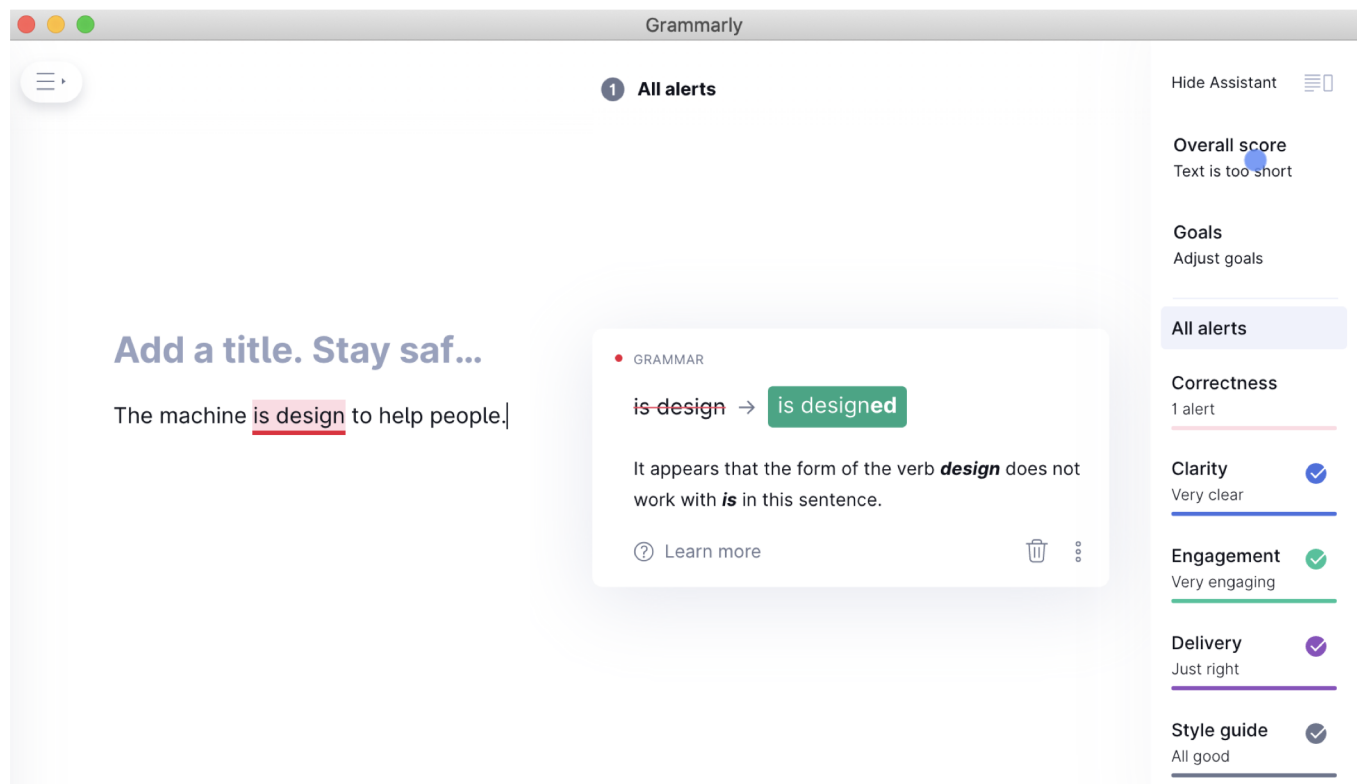
The Machine is design^{ed} to help people.

主流なアプローチ：

誤りが含まれる文から誤りが含まれない文への「翻訳」

すでに実用化されています

→ Grammarly¹, Ginger²... など.



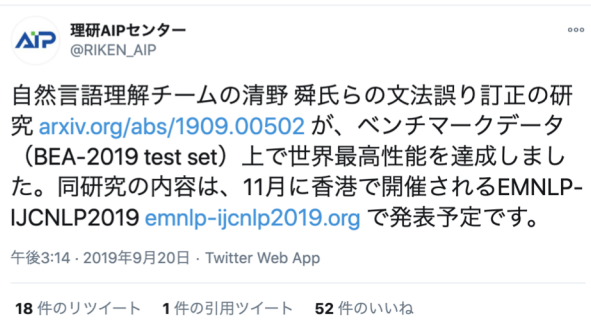
1. <https://app.grammarly.com/>
2. <http://www.getginger.jp/>

世界トップクラスの技術を持っています

- **トップカンファレンス**に数多くの論文が採択！
 - NAACL 2019, EMNLP 2019, ACL 2020, EMNLP 2020
- 国際コンペ (BEA-2019 Shared Task※) で**世界2位**！ ※ Unrestricted Track

| Rank ▲ | User ▲ | Team Name ▲ | TP ▲ | FP ▲ | FN ▲ | P ▲ | R ▲ | F0.5 ▲ |
|--------|----------------|----------------------------------|------|------|------|-------|-------|--------------|
| 1 | goo2go | LAIX | 2618 | 960 | 2671 | 73.17 | 49.50 | 66.78 |
| 2 | tomoyamizumoto | AIP-Tohoku | 2589 | 1078 | 2484 | 70.60 | 51.03 | 65.57 |
| 3 | arahusky | UFAL, Charles University, Prague | 2812 | 1313 | 2469 | 68.17 | 53.25 | 64.55 |
| 4 | hsamswcc | BLCU | 3051 | 2007 | 2357 | 60.32 | 56.42 | 59.50 |
| 5 | gurunathp | Aparecium | 1585 | 1077 | 2787 | 59.54 | 36.25 | 52.76 |
| 6 | mengyang | Buffalo | 699 | 374 | 3265 | 65.14 | 17.63 | 42.33 |
| 7 | nihalnayak | Ramaiah | 1161 | 8062 | 3480 | 12.59 | 25.02 | 13.98 |

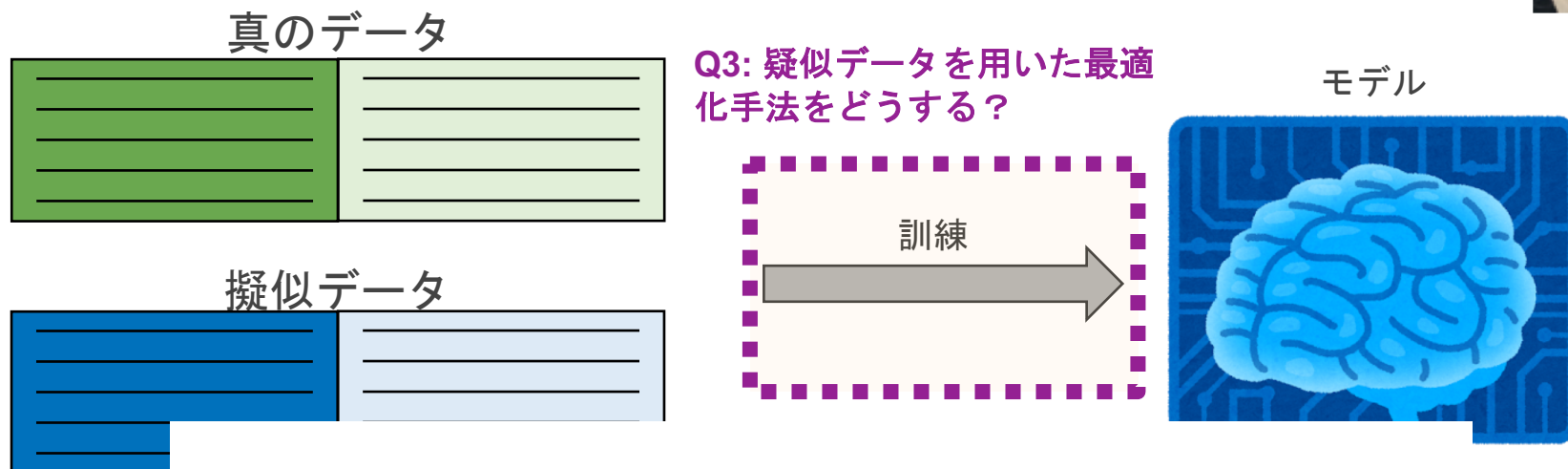
- ベンチマーク (BEA-2019 test set) 上で**世界最高性能**を達成！



擬似データの活用

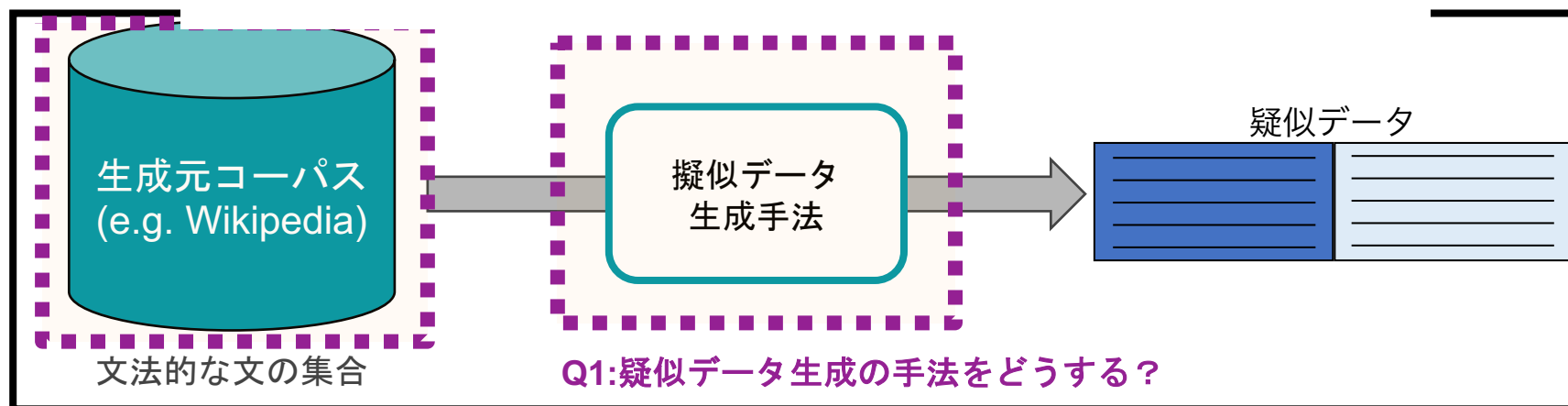
[Kiyono+, EMNLP 2019]

<https://github.com/butsugiri/gec-pseudodata>



擬似データを活用する場合の効果的な設定の探索

Q2: 生成元コー



データノイズ除去 [Mita+, EMNLP 2020]



- 人手で作成された学習者コーパス（真のデータ）にも，誤訂正や訂正漏れに起因した“ノイズ”が無視できない量含まれることを指摘
 - 訂正者のケアレスミスやスキル不足，データ収集元の性質（例. 訂正するのはあくまでオプション）などの要因

1 : Errors are inappropriately edited

Source: *I want to **discuss about** the education.*

Target: *I want to **discuss of** the education.*

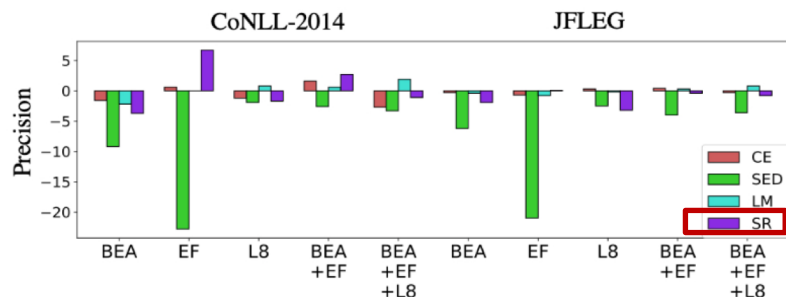
2 : Errors are left uncorrected

Source: *We **discuss about** our sales target.*

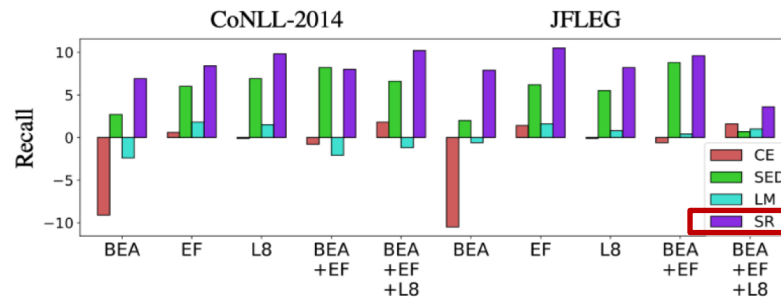
Target: *We **discuss about** our sales target.*



ノイズ除去することで適合率を維持しつつカバー率が大幅改善！



(a) Precision
適合率



(b) Recall
カバー率

現在の到達点

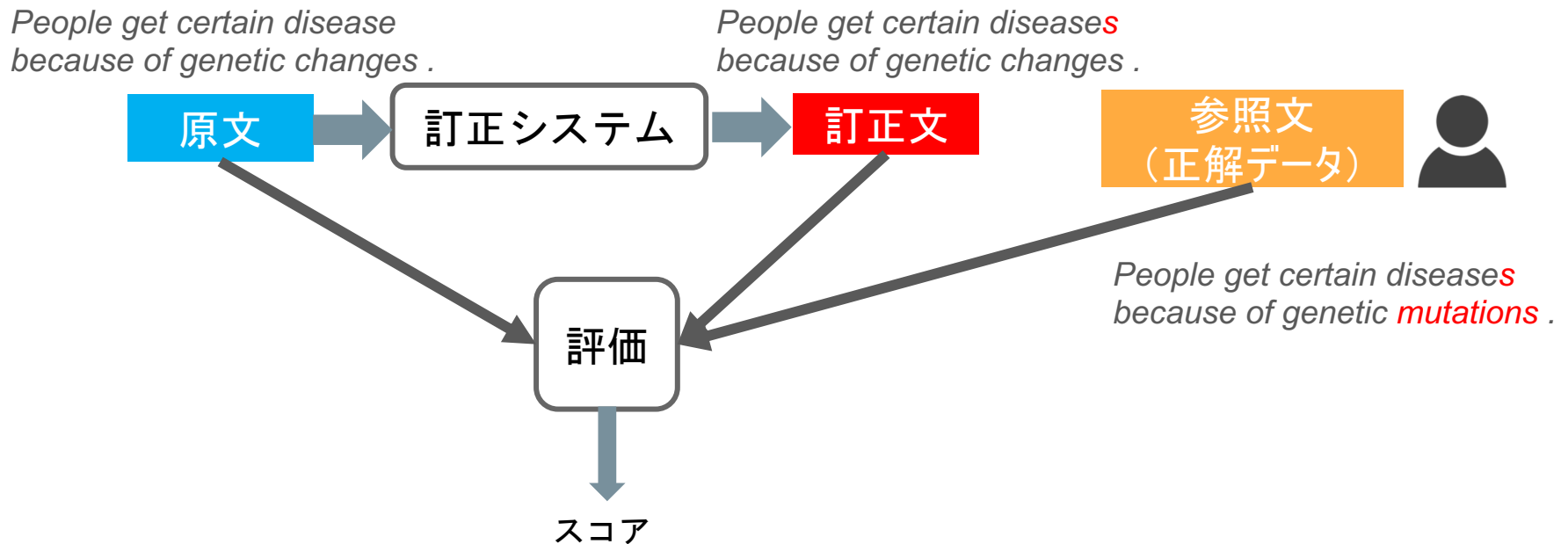
CoNLL-10 における性能

| | 適合率 | カバー率 | $F_{0.5}$ 値 |
|------|-------|-------|-------------|
| Ours | 89.38 | 53.36 | 78.75 |

- 全体の約53%の誤りに対して約90%の精度で訂正可能

評価の方法

現在主流の評価方法（参照有り評価）：



評価の限界: 正解データに含まれない「正解」を扱えない

- (i) 訂正文 も正しいが参照文の訂正とマッチしない場合（適合率↓）
- (ii) 原文 も誤りではないが参照文では訂正している場合（カバー率↓）

現在の到達点

CoNLL-10 における性能

| | 適合率 | カバー率 | $F_{0.5}$ 値 |
|----------------------------|-------|-------|-------------|
| Ours | 89.38 | 53.36 | 78.75 |
| Human (native experts) [1] | - | - | 72.58 |

➤ 実際には見かけよりもかなり良い数値

[1] Bryant and Ng, 2015. How Far are We from Fully Automatic High Quality Grammatical Error Correction?

学習支援としての文法誤り訂正

学習支援として期待できる利点：

- 直後フィードバックが可能
 - 人間ができるのはせいぜい遅延フィードバック
 - インタラクティブ学習が可能
- 書き直しの動機付け
 - 書直しを行うことは学習効果を高めることに有効

しかし、言語処理的研究の文脈ではほとんど研究がない...（そもそも性能が追い付かずできなかった）

→ 性能が向上してきているのでいよいよ研究ができる下地
ができてきた 例) フィードバックのストラテジーと教育効果

今後の展望 - 文法誤り訂正を超えて-



初期段階の下書きから流暢な文を生成

誤りを含む不完全な文

Our model () to the SemEval-2007 task and (*) good results.*

流暢性のある完全な文

We applied our model to the Semeval-2007 task and obtained good results.

We demonstrate that our model is able to achieve good results on the Semeval-2007 task and outperforms other baselines.

Experimental results showed that our model achieves comparable performance on the Semeval-2007 task and yields good results.

デモシステム公開 [Ito, Kuribayashi+, EMNLP 2020]



Langsmith

<https://editor.langsmith.co.jp/>

Generating Fluent Sentences from Early-Stage Drafts for Academic Writing

1 * Introduction

2 There are several stages in writing, including drafting, reviewing, editing, proofreading. Studies on existing writing assistance, such as Grammatical Error Correction (GEC), have focused only the early stages of writing. On the other hand, few studies have focused on supporting the early revision phase, which requires

On the other hand, only few studies have focused on supporting the early revision stage which requires the correction and extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision stage which requires the correction and extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision stage which requires the correction and extensive rewriting of the information contained in the texts.

On the other hand, few studies have focused on supporting the early revision phase which requires a correction and an extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision phases which requires the correction and extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision stages which requires the correction and extensive rewriting of the information contained in the text.

On the other hand, few studies have focused on supporting the early revision stage which requires the correction and extensive rewriting of the information contained within a given text.

In contrast, few studies have focused on supporting the early revision phase which requires the correction and extensive rewriting of the information contained in the text.



Langsmith



New document



New document with example

Generating Fluent Sentences from Early-Stage Drafts for Academic Writing



B I H2 H3

論述リライト [三田ら, YANS 2019 奨励賞]



ライティング・プロセス
(人間)

Scope

自然言語処理



文章

Revision

“For example, we show actual responses generated by a vanilla seq2seq-based model trained on Twitter conversations in Table1. *As can be seen, the* ~~The~~ responses have inconsistent style as if the model had multiple personalities.”

論述リライト



Editing

フレーズ / 文

“The language model allows *emulation of* ~~to emulate~~ the noise generated.”



Proofreading

単語

“We present results *of* ~~on~~ a quantitative analysis.”

文法誤り訂正
(Grammatical
Error Correction)

論述リライトの実例

readability (読みやすさ)

original: In this research area, methods to automatically generate image descriptions (captions), that is, image captioning, have attracted a great deal of attention (Karpathy and Fei-Fei, 2015; Donahue et al., 2015; Vinyals et al., 2015; Mao et al., 2015).

revision: In this research area, image captioning methods, which automatically generate image descriptions (captions), have attracted a great deal of attention (Karpathy and Fei-Fei, 2015; Donahue et al., 2015; Vinyals et al., 2015; Mao et al., 2015).

"have" vs "has"の選択を明確にするために配置を変えましょう

論述リライトの実例

monotonicity (単調性)

original: The paper is organized as follows. Section 2 presents some previous research on distributional similarity and word sense induction. Section 3 gives an overview of our method for word sense induction and disambiguation.

Section 4 provides a quantitative evaluation and comparison to other algorithms in the framework of the SEMEVAL-2010 word sense induction and disambiguation (WSI/WSD) task.

revision: This paper is organized as follows. Section 2 presents some previous research on distributional similarity and word sense induction. Section 3 gives an overview of our method for word sense induction and disambiguation. **In Section 4,** we provide a quantitative evaluation and comparison to other algorithms in the framework of the SEMEVAL-2010 word sense induction and disambiguation (WSI/WSD) task.

“Section” が頻発してて単調
になってます

論述リライトの実例

sentence splitting (文分割)

original: The intuition is that a particular sense is associated with a particular topic, so that different senses can be discriminated through their association with particular topical dimensions; in a similar vein, a particular instance of a word can be disambiguated by determining its most important topical dimensions.

revision: The intuition is that a particular sense is associated with a particular topic, so that different senses can be discriminated through their association with particular topical dimensions. In a similar vein, a particular instance of a word can be disambiguated by determining its most important topical dimensions.

この文章が長くなってきたので、
ここで二つに分けましょう

おわりに

- 文法誤り訂正の性能は高い水準まで向上してきた
- しかし、言語処理的な文脈で学習支援を目的とした文法誤り訂正に関する研究はほとんどない
- 技術的な下地ができてようやく学習効果を期待できる取り組みができるようになってきた

ご協力いただける教育事業者・関係者および研究者の方がいましたらお声がけください！

- 訂正フィードバックのあり方の議論
- 実証実験を通じた学習・教育効果の検証
- データ作成，モデルの開発など