

# 埋め込みノルムによる連続的な文難易度制御

藤原 有希<sup>1</sup> 宮田 莉奈<sup>1</sup> 梶原 智之<sup>1,2</sup> 荒瀬 由紀<sup>3</sup>

<sup>1</sup> 愛媛大学 <sup>2</sup> 大阪大学 <sup>3</sup> 東京科学大学

{fujiwara@ai., miyata@ai., kajiwara@}cs.ehime-u.ac.jp arase@c.titech.ac.jp

## 概要

本研究では、英語教育支援のために、CEFR に準拠した文難易度の連続的な制御手法を提案する。先行研究では、入力文頭に目標難易度を指示する特殊トークンを付与して難易度を制御してきた。しかし、目標難易度トークンを個別に語彙に登録する既存手法には、(1) 難易度の順序関係を学習できない (2) 離散的な難易度しか扱えない (3) 入力文の難易度を認識しないため平易な表現も編集してしまうという課題がある。そこで我々は、目標難易度に依存しない共通の難易度埋め込みを用意し、そのノルムの大小として入力文や出力文の難易度の高低を表現する。これによって、難易度の順序関係を自然に学習でき、連続的な難易度も自然に扱えるうえに、1 モデルで難易度推定と難易度制御の両タスクに対応できる。評価実験の結果、英語における文難易度の推定と制御の両方で提案手法の有効性を確認した。

## 1 はじめに

テキスト平易化 [1] とは、文の主要な意味を保持しつつ難解な表現を平易に言い換えるタスクである。この技術は、様々な読者の読解支援 [2-4] に有用だけでなく、言語教育支援においては各学習者の言語能力に合わせた教材の提供 [5] にも貢献する。

多様な読者や学習者に対応するために、近年のテキスト平易化は、出力文の難易度制御 [6-9] に重点を置いている。これらの先行研究では、目標難易度を指示する<3>などの記号（目標難易度トークン）を入力文頭に加えることで出力文の難易度を制御している。しかし、複数の目標難易度トークンを個別に学習するため、それらの間の相互作用がなく、難易度の順序関係を十分に学習できない。また、目標難易度トークンは事前に語彙に登録するため、離散的な難易度しか扱えない。そして、これらの手法では入力文の難易度を明示的に認識しないため、平易な表現も編集してしまうという課題がある。

これらの課題に対処するために、本研究では、目標難易度に依存しない共通の難易度トークン（図 1 の [Lv]）を用意する。そして、この難易度トークンの埋め込みのノルムに難易度を反映し、難解な文ほど大きいノルムで表現する。これによって、難易度の順序関係の学習および連続値としての表現の両方を、自然に実現できる。提案手法を Transformer [10] ベースのエンコーダ・デコーダモデルに適用すると、符号化器において入力文の難易度を明示的に表現できる。そして、難易度埋め込みのノルムを（向きはそのまま）目標難易度の大きさまで縮めて復号器に渡すことで、出力文の難易度を制御できる。

提案手法の有効性を検証するために、英文の難易度推定および難易度制御の性能を評価した。ここで、文難易度の基準には、CEFR<sup>1)</sup>を採用した。実験の結果、両タスクにおいて提案手法の有効性を確認できた。本研究の貢献は、以下の3点である。

1. Transformer モデル内で難易度を明示的に表現および編集する手法を提案し、単一モデルで難易度推定と難易度制御の両方を実行可能にした。
2. ノルムでの難易度表現によって、これまで困難であった文難易度の連続的な制御を実現した。
3. 英文の難易度推定および難易度制御の両タスクにおいて、既存手法を上回る性能を達成した。

## 2 提案手法

本研究では、Transformer [10] ベースのエンコーダ・デコーダモデルに対して、文難易度を表現するための特殊トークン [Lv] を追加し、それを符号化したベクトルのノルムによって入力文の難易度を表現する。さらに、復号器に与えるベクトルのノルムの大きさに応じて出力文の難易度を制御する。ノルムとして文難易度を表現することにより、Transformer

1) Common European Framework of Reference for Languages  
英語学習者の言語能力を表す国際標準の指標で、A1, A2, B1, B2, C1, C2 の 6 段階 <https://www.coe.int/en/web/common-european-framework-reference-languages>

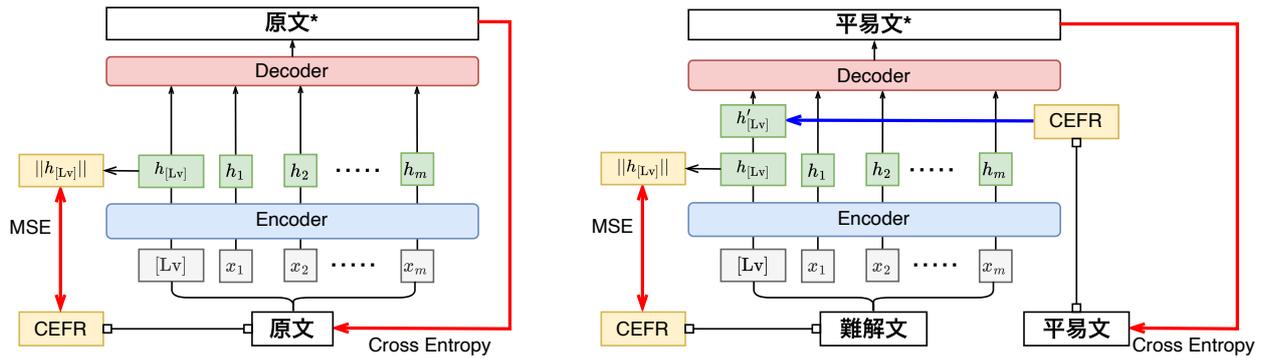


図1 提案手法の概要：難易度の表現に関する訓練（左）および難易度の制御に関する訓練（右）

モデルが明示的に難易度を学習できるとともに、連続的な難易度の制御も可能となる。提案手法では、図1に示すように、事前訓練済みTransformer [11] に対して、難易度の表現に関する訓練および難易度の制御に関する訓練の2段階の追加訓練を適用する。

## 2.1 Step1: 難易度の表現に関する訓練

ノルムによる文難易度の表現を実現するために、自己符号化の訓練の中で、入力文の難易度とベクトルのノルムの大きさを一致させる。この訓練には、文  $X_i \in \{x_{i1}, x_{i2}, \dots, x_{i|X_i|}\}$  に対して難易度  $g_i \in \{1, 2, \dots\}$  が付与されたデータセット  $\{(X_1, g_1), \dots, (X_N, g_N)\}$  を用いる。ただし、難易度  $g_i$  は値が小さいほど平易であることを意味する。

まず、符号化器に対して文  $X_i$  および難易度トークン [Lv] を入力し、ベクトル  $\mathbf{H}_i = \{\mathbf{h}_{i[Lv]}, \mathbf{h}_{i1}, \mathbf{h}_{i2}, \dots, \mathbf{h}_{i|X_i|}\}$  を得る。そして、入力文の難易度  $g_i$  と難易度埋め込みのノルム  $\hat{g}_i := \|\mathbf{h}_{i[Lv]}\|$  が一致するように、式(1)の誤差を最小化する。

$$L_g = \frac{1}{N} \sum_{i=1}^N (g_i - \hat{g}_i)^2 \quad (1)$$

また、符号化器と復号器を同時に更新するために、復号器ではベクトル  $\mathbf{H}_i$  から入力文  $X_i$  の再構成をクロスエントロピー損失  $L_c$  によって訓練する。最終的に、難易度の表現に関する1段階目の追加訓練では、重み  $\alpha$  を導入し、 $L_g$  と  $L_c$  のマルチタスク学習によって、損失  $L_1 = L_g + \alpha L_c$  を最小化する。

## 2.2 Step2: 難易度の制御に関する訓練

2.1節で訓練したモデルを、テキスト平易化パラレルコーパス上でさらに追加訓練することによって、難易度の制御能力を獲得する。ここでは、難解文  $X_i$  と平易文  $Y_i$  の文対に対して、それぞれ難易度  $g_i$  および  $t_i$  が付与されたデータセット

$\{(X_1, g_1), (Y_1, t_1), \dots, (X_N, g_N), (Y_N, t_N)\}$  を使う。

まず、2.1節と同様に、難解文  $X_i$  および難易度トークン [Lv] を符号化器に入力し、ベクトル  $\mathbf{H}_i$  を得る。なお、この段階では、難易度埋め込み  $\mathbf{h}_{i[Lv]}$  のノルムは難解文  $X_i$  の難易度を表現している。

次に、復号器に目標難易度の情報を与えるために、難易度埋め込み  $\mathbf{h}_{i[Lv]}$  を式(2)のように平易文  $Y_i$  の難易度  $t_i$  に応じて縮小する。

$$\mathbf{h}'_{i[Lv]} = \mathbf{h}_{i[Lv]} * \frac{t_i}{\|\mathbf{h}_{i[Lv]}\|} \quad (2)$$

そして、符号化器の出力に対して式(2)のノルム操作を施したベクトル  $\mathbf{H}'_i = \{\mathbf{h}'_{i[Lv]}, \mathbf{h}_{i1}, \mathbf{h}_{i2}, \dots, \mathbf{h}_{i|X_i|}\}$  を復号器に入力し、平易文  $Y_i$  を生成するためにクロスエントロピー損失  $L_s$  の最小化を訓練する。最終的に、難易度の制御に関する2段階目の追加訓練では、重み  $\beta$  を導入し、 $L_g$  と  $L_s$  のマルチタスク学習によって、損失  $L_2 = L_g + \beta L_s$  を最小化する。

## 2.3 実装の詳細

Step1の訓練では、英文に対してCEFR難易度が付与されたCEFR-SP [12]のデータセットを使用した。CEFR-SPでは、それぞれの英文に対して2人のアノテータがCEFR難易度を付与しているが、本研究では両者の平均値<sup>2)</sup>を用いた。

Step2の訓練では、英語のテキスト平易化パラレルコーパスであるNewsela-Auto [13]を使用した。Newsela-Autoは難解文に対して難易度の異なる複数の平易文が付与されているため難易度制御の実験に適しているが、提供されている難易度ラベルは文書単位かつ米国の学年に対応するものである。本研究では、文単位でCEFRの難易度を使用したいため、CEFR-SPで訓練された文難易度推定器<sup>3)</sup>を用いて各

2) 難易度はA1を1, A2を2, ..., C2を6とし、例えばA1とA2のラベルを持つ文には1.5の難易度ラベルを与えた。

3) <https://github.com/yukiari/CEFR-SP>

表1 データセットの統計情報

	訓練用	検証用	評価用
CEFR-SP (文数)	13,990	1,843	1,660
Newsela-Auto (文対数)	96,085	12,203	12,388

表2 難易度推定タスクの実験結果

	MAE	RMSE
文難易度推定器	0.454	0.573
BERT 回帰モデル	0.398	0.503
提案手法: Step1	0.423	0.540
提案手法: Step2	0.401	0.522
提案手法: Step1 → 2	<b>0.395</b>	<b>0.501</b>

文に CEFR 難易度を付与した。ここで、対応する難解文と複数の平易文の組において、難解文よりも高難度の平易文や、同じ難易度に該当する複数の平易文はノイズとなるためデータセットから除外した。

モデルには事前訓練済みの BART<sup>4)</sup> [11] を使用した。最適化には Adam [14] を使用し、学習率を  $10^{-6}$ 、バッチサイズを 32 に設定し、検証用セットにおける損失関数に関する 3 エポックの early stopping によって訓練した。損失関数の重みは  $\alpha = \beta = 1$  とした。

### 3 評価実験：難易度推定

本節では、提案手法における難易度埋め込みのノルムによる難易度表現の正確性を評価する。

**実験設定** 表 1 に示す CEFR-SP を用いて、英文の難易度推定を実験した。ここで、CEFR-SP の一部の文は Newsela-Auto と重複しているため、Newsela-Auto の訓練用セットに含まれる文は除外した。比較手法には、既存手法<sup>3)</sup> [12] および BERT<sup>5)</sup> [15] を CEFR-SP で訓練した回帰モデルを採用した。評価指標には、Mean Absolute Error (MAE) および Root Mean Squared Error (RMSE) を使用した。

**実験結果** 表 2 に実験結果を示す。提案手法は比較手法よりも低い誤差を達成し、その有効性を確認できた。提案手法はノルムとして文難易度を表現するため、深層学習の推論を伴う比較手法よりも解釈性が高い。本実験から、提案手法が解釈性と難易度推定性能を両立できていることが明らかになった。

### 4 評価実験：難易度制御

次に、提案手法による難易度制御について、参照文との類似性および制御性の観点から評価する。

4) <https://huggingface.co/facebook/bart-base>

5) <https://huggingface.co/google-bert/bert-base-uncased>

表3 難易度制御タスクの実験結果

	SARI	add	keep	del	BERTScore
入力文	8.28	0.00	24.84	0.00	0.712
BART	38.26	5.10	31.33	78.33	0.716
+grade	40.27	5.66	34.32	80.82	0.723
Llama	38.33	5.23	29.35	80.41	0.702
+grade (FSL)	40.95	5.98	29.15	<b>87.73</b>	0.713
+grade (SFT)	38.98	<b>6.34</b>	31.38	79.22	0.717
Step1	8.59	0.00	24.88	0.90	0.712
Step2	40.82	5.65	35.09	81.72	0.723
Step1 → 2	<b>42.39</b>	5.91	<b>37.39</b>	83.85	<b>0.725</b>

**実験設定** 表 1 に示す Newsela-Auto を用いて、英文の難易度制御を実験した。比較手法には、入力文を編集しないベースライン、目標難易度を与えずに BART を訓練するベースライン、目標難易度を指示する特殊トークンを入力文頭に付与する既存手法 [6] (BART+grade) に加え、大規模言語モデル<sup>6)</sup> [16] も用いた。大規模言語モデルは、訓練なしでタスクのみを指示する 0-shot 推論 (Llama), Newsela-Auto の訓練セットから無作為抽出した 3 件の事例も与える few-shot 学習 (Llama+grade (FSL)), Newsela-Auto による教師あり学習 (Llama+grade (SFT)) の 3 種類の方法で使用した。大規模言語モデルに指示や事例を与えるプロンプトは、先行研究 [17] における Prompt 2 [18] に従った。評価指標には、参照文との表層的な類似性を評価する SARI<sup>7)</sup> [20] および意味的な類似性を評価する BERTScore [21] を使用した。

**実験結果** 表 3 に実験結果を示す。まず、提案手法は既存手法や大規模言語モデルの性能を上回り、SARI と BERTScore の両評価指標において最高性能を達成した。次に、SARI の内訳 (add/keep/del) に着目すると、Llama+grade (FSL) は del が高い一方で keep においては BART ベースラインにも劣っており、入力文中の平易な表現を過剰に削除する傾向がある。一方で、提案手法は入力文における平易な表現の保持を評価する keep の指標において最良の結果を示すとともに、add や del においても良好な傾向が見られた。最後に、提案手法は Step2 の難易度の制御に関する訓練のみでも高い性能が得られているものの、Step1 の難易度の表現に関する訓練を組み合わせることによって全ての評価指標において改善が見られた。本実験から、難易度制御における提案手法の有効性が確認できた。

6) <https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct>

7) 実装は EASSE [19] <https://github.com/feralvam/easse>

表 4 提案手法による難易度制御の事例

文	CEFR 難易度
原文 Some marine biologists and ocean advocates warn that such a compassion-laden response is little more than a futile attempt to wrap a Band-Aid on an oceanwide problem.	5.0
生成文 Some marine biologists and ocean advocates say that such a compassionate response is little more than a Band-Aid on an oceanwide problem.	4.0
Some marine biologists and ocean advocates say that such a response is little more than a way to help the ocean.	3.5
Some scientists and ocean advocates say that is not the answer.	3.0
Some scientists and ocean advocates say that is not the case.	2.5

表 5 離散的な難易度制御の性能に関する分析

	自動評価		人手評価	
	Acc	$\rho$	Acc	$\rho$
BART+grade	0.250	0.847	0.380	0.906
Llama+grade (FSL)	0.186	0.685	0.333	0.812
提案手法: Step2	0.500	0.919	0.467	0.917
提案手法: Step1 → 2	<b>0.691</b>	<b>0.961</b>	<b>0.687</b>	<b>0.962</b>

表 6 連続的な難易度制御の性能に関する分析

	自動評価		人手評価	
	Acc	$\rho$	Acc	$\rho$
BART+grade	0.028	0.631	0.007	0.802
Llama+grade (FSL)	0.031	0.119	0.053	0.596
提案手法: Step2	0.165	0.806	0.133	0.852
提案手法: Step1 → 2	<b>0.412</b>	<b>0.888</b>	<b>0.387</b>	<b>0.897</b>

## 5 分析: 難易度の制御性

本節では、難易度制御モデルが、目標難易度の低下に従ってより平易な文を生成できるかどうかを、自動評価と人手評価の両方で詳細に分析する。

**実験設定** Newsela-Auto から難易度  $g_i \in \{5, 6\}$  の難解文 544 文を抽出し、それぞれを 4 種類の目標難易度に対して平易化した。そして、出力文の難易度を自動または人手で評価し、そのランキング性能を分析した。ここで、離散的な難易度制御の性能を分析する表 5 の設定では目標難易度を  $t_i \in \{4, 3, 2, 1\}$  とし、連続的な難易度制御の性能を分析する表 6 の設定では目標難易度を  $t_i \in \{4.0, 3.5, 3.0, 2.5\}$  とした。比較手法には、表 3 において高い SARI を示した BART+grade および Llama+grade (FSL) を用いた。ただし、前者は離散的な目標難易度しか入力できないため、目標難易度が非整数の際は、前後の目標難易度を表す特殊トークンの平均ベクトルで代用した。

**評価指標** 自動評価には、CEFR-SP で訓練された文難易度推定器<sup>3)</sup> [12] を使用し、各文の難易度を評価した。人手評価には、Amazon Mechanical Turk<sup>8)</sup> を通して、継続的な高評価の実績を持つ「マスター資格」保有者かつ英語母語話者である評価者を雇用<sup>9)</sup> した。難解文を 50 文無作為抽出し、3 人ずつの評価者に 4 つの平易文を難易度でランキングさせた。ラ

ンキングの性能評価には、完全一致率 (Acc) およびスピアマンの順位相関係数 ( $\rho$ ) を用いた。

**定量評価** 表 5 および表 6 に分析の結果を示す。自動評価と人手評価の両方で、提案手法が一貫して最高性能を示した。特に、表 6 の連続的な難易度制御については、ランキングの完全一致率 (Acc) において比較手法は 10%未満という低評価であった。提案手法は、ノルム操作によって連続的な難易度制御を実現できたため、大幅に性能を改善できた。

**定性評価** 表 4 に、提案手法による連続的な難易度制御の事例を示す。この例は、難易度 5 の原文に対して、目標難易度を  $\{4, 3.5, 3, 2.5\}$  と変化させながら、つまり、難易度埋め込みのノルムを目標難易度に従って縮小しながら、平易文を生成したものである。目標難易度の低下とともに、原文に対してより多くの編集を施し、より平易な語彙やより短い文長の文を生成していることが確認できる。この事例からも、提案手法が難易度埋め込みのノルム操作によって適切に文難易度を制御できることがわかる。

## 6 おわりに

本研究では、文難易度をベクトルのノルムとして表現することにより、連続的な難易度制御を実現した。英語における評価実験により、難易度推定および難易度制御の両タスクにおける提案手法の有効性を確認するとともに、これまで困難であった難易度の連続的な制御が有効に機能することを確認した。

今後は、他の言語や文体における有効性の検証や、大規模言語モデルへの応用に取り組みたい。

8) <https://www.mturk.com/>

9) アノデータには時給 12 ドルを支払った。これは、代表的なクラウドソーシングサービスのひとつである Prolific における推奨報酬額である。 <https://www.prolific.com/pricing>

## 謝辞

本研究は、JSPS 科研費（基盤研究 B，課題番号：JP25K03233）の助成を受けて実施した。

## 参考文献

- [1] Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. Data-Driven Sentence Simplification: Survey and Benchmark. *Computational Linguistics*, Vol. 46, No. 1, pp. 135–187, 2020.
- [2] John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. Practical Simplification of English Newspaper Text to Assist Aphasic Readers. In *Proceedings of the AAIL-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pp. 7–10, 1998.
- [3] Sarah E Petersen and Mari Ostendorf. Text Simplification for Language Learners: A Corpus Analysis. In *Proceedings of the Workshop on Speech and Language Technology in Education*, pp. 69–72, 2007.
- [4] Jan De Belder and Marie-Francine Moens. Text Simplification for Children. In *Proceedings of the SIGIR 2010 Workshop on Accessible Search Systems*, pp. 19–26, 2010.
- [5] Batia Laufer. How Much Lexis is Necessary for Reading Comprehension? *Vocabulary and Applied Linguistics*, pp. 126–132, 1992.
- [6] Carolina Scarton and Lucia Specia. Learning Simplifications for Specific Target Audiences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 712–718, 2018.
- [7] Daiki Nishihara, Tomoyuki Kajiwara, and Yuki Arase. Controllable Text Simplification with Lexical Constraint Loss. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 260–266, 2019.
- [8] Daiki Yanamoto, Tomoki Ikawa, Tomoyuki Kajiwara, Takashi Ninomiya, Satoru Uchida, and Yuki Arase. Controllable Text Simplification with Deep Reinforcement Learning. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, pp. 398–404, 2022.
- [9] Rina Miyata, Koki Horiguchi, Risa Kondo, Yuki Fujiwara, and Tomoyuki Kajiwara. EhiMeNLP at TSAR 2025 Shared Task: Candidate Generation via Iterative Simplification and Reranking by Readability and Semantic Similarity. In *Proceedings of the Fourth Workshop on Text Simplification, Accessibility and Readability*, pp. 217–222, 2025.
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Proceedings of the 31st Conference on Neural Information Processing Systems*, pp. 5998–6008, 2017.
- [11] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7871–7880, 2020.
- [12] Yuki Arase, Satoru Uchida, and Tomoyuki Kajiwara. CEFR-Based Sentence Difficulty Annotation and Assessment. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 6206–6219, 2022.
- [13] Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. Neural CRF Model for Sentence Alignment in Text Simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7943–7960, 2020.
- [14] Diederik P. Kingma and Jimmy Lei Ba. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference for Learning Representations*, 2015.
- [15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, 2019.
- [16] Llama Team. The Llama 3 Herd of Models. 2024.
- [17] Tannon Kew, Alison Chi, Laura Vásquez-Rodríguez, Sweta Agrawal, Dennis Aumiller, Fernando Alva-Manchego, and Matthew Shardlow. BLESS: Benchmarking Large Language Models on Sentence Simplification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 13291–13309, 2023.
- [18] Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. ASSET: A Dataset for Tuning and Evaluation of Sentence Simplification Models with Multiple Rewriting Transformations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4668–4679, 2020.
- [19] Fernando Alva-Manchego, Louis Martin, Carolina Scarton, and Lucia Specia. EASSE: Easier Automatic Sentence Simplification Evaluation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing: System Demonstrations*, pp. 49–54, 2019.
- [20] Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. Optimizing Statistical Machine Translation for Text Simplification. *TACL*, Vol. 4, pp. 401–415, 2016.
- [21] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. BERTScore: Evaluating Text Generation with BERT. In *Proceedings of the 8th International Conference on Learning Representations*, 2020.