

エッセイ自動採点における文法特徴と学習者レベルの関係

土肥 康輔 須藤 克仁 中村 哲
奈良先端科学技術大学院大学

{doi.kosuke.de8, sudoh, s-nakamura}@is.naist.jp

概要

本研究では、エッセイ自動採点において文法特徴を考慮することの効果について検証する。エッセイの分散表現に加えて、学習者が正しく使っている文法項目の情報や、文法誤りの情報をモデルの入力に用いることで、モデル性能が向上するかを検証した。実験の結果、正しく使っている文法項目の情報は、モデル性能の向上につながらなかった。また、文法誤りの情報は、学習者のレベルが高いデータセットにおいては効果がない一方で、様々なレベルの学習者が含まれているデータセットにおいては効果があることが示された。

1 はじめに

近年の外国語教育では、ライティングやスピーキング能力の育成も重視されるようになっており、学校現場や入試で両能力のテストが実施されることも増えている。しかし、その評価には、大きな人的・時間的コストがかかるという課題がある。採点自体に時間がかかることに加えて、評価の信頼性を保つためには、採点者の訓練が必要である。このような課題を背景に、自動採点研究が行われている。

ライティングやスピーキングといった学習者のパフォーマンスを評価する方法は、総合的評価と分析的評価に大別される。前者は学習者の作文や発話に対して1つのスコアを付与するのに対して、後者は文法や内容等の複数の観点に対してそれぞれスコアを付与する。ただし、総合的評価においても、それらの観点での出来栄が基準となっていることが一般的である。例えばTOEFL iBTの基準(表1)では、内容、構成、文法、語彙等が言及されている。特に文法、語彙については、テスト問題のトピックとは関係なく、ある習熟度レベルの学習者に特徴的な項目があることが明らかになっている[1]。人間の採点者は、そのような項目を探しながら学習者のパフォーマンスを評価しているとも言われており[1]、

表1 TOEFL iBT ライティング基準 (スコア 5)

An essay at this level largely accomplishes all of the following:

- Effectively addresses the topic and task
- Is well organized and well developed, using clearly appropriate explanations, exemplifications and/or details
- Displays unity, progression and coherence
- Displays consistent facility in the use of language, demonstrating syntactic variety, appropriate word choice and idiomaticity, though it may have minor lexical or grammatical errors

自動採点モデルにおいて文法特徴を明示的に利用することで、様々なテスト問題に対して頑健なモデルを構築することが期待できる。そこで本研究は、エッセイ自動採点(AES)において文法特徴を考慮することの効果について検証する。

2 関連研究

AESは、学習者が書いたエッセイにスコアやレベル等の評価値を付与するタスクである。AESモデルでは、特徴量エンジニアリングによって得られた特徴量が広く用いられていた([2], [3]を参照)。例えばe-rater[4]は、文法誤りや語彙の複雑さの指標を含む12種類の特徴量を用いている。[5]は様々な言語特徴を用いて学習したモデルの検証を行い、文法の複雑さや誤りに関する指標に高い重みが割り振られていることを明らかにした。[6]は文法誤り検出タスクとのマルチタスク学習を行うことで、AESのパフォーマンスが向上することを示した。

近年では、特徴量エンジニアリングを不要とする、深層学習に基づく手法が主流となってきている。RNNやBi-LSTMを用いたモデル[7, 8]や、BERT[9]等の事前学習済み言語モデルに基づくモデル[10, 11]が提案されている。しかし[2]は、これらの深層学習に基づくアプローチと特徴量エンジニアリングに基づくアプローチは相補的な関係にあり、両者を組み合わせることでモデル性能が向上する可能性を指摘した。実際に、エッセイの分散表現と人手で作成された特徴量を組み合わせたモデル[12, 13]が提案されている。[14]は、BERTから得ら

れるエッセイの分散表現と文法誤りの情報を組み合わせたモデルを提案した。しかし、文法誤り情報の追加は、話し言葉自動採点でモデル性能を向上させた一方で、AES では効果がなかった。

3 文法特徴利用の効果検証

本研究では、[14]の手法をベースに、BERTに基づくAESモデルにおいて、文法特徴を用いることの効果について検証する。[14]では、文法特徴として文法誤り情報が用いられていたが、本研究では、CEFRの基準特性研究[1]を参考に、学習者が正しく使っている文法項目の情報も用いる。CEFR[15]は言語能力を評価する国際指標で、習熟度をA1(初級)～C2(上級)の6段階に区分する。基準特性はCEFRの各レベルで特徴的な言語項目群であり、あるレベル以上の学習者が正しく使うことができるpositive linguistic features (PF)と、あるレベルの学習者が間違いやすいnegative linguistic features (NF)がある。以下の節で、検証に用いるPFとNFについて具体的に説明する。

3.1 Positive Linguistic Features

PFの抽出には、CEFR-J Grammar Profile 文法項目頻度分析プログラム[16, 17]を用いる。同プログラムは、501種類の文法項目のテキスト中での頻度を、正規表現に基づいて算出できる。しかし、そのままではテキスト中で使われていない文法項目が多く、スパースなベクトルとなってしまったため、100語あたりの延べ使用数、及び文法項目の異なり数に基づき集計したものを特徴量として用いる(表2上段)。また、4.1節で後述するデータの訓練セットにおいて、エッセイの評価値と相関の高かった文法項目(100語あたりの延べ使用数=6項目、使用の有無=25項目)のベクトルも使用する(表2中段)。

3.2 Negative Linguistic Features

NFには、100語あたりの文法誤り頻度を用いる。具体的には、ERRANT[18]によって付与される、24種類の誤りタグに基づき、特徴量を作成する(表2下段)。タグのグループ化は、EFCamDat[19]に付与されている誤りタグに対して、[14]が行ったグループ分けを参照し、できる限り対応が取れる形で行った(付録A)。テキスト中の文法誤り数の推定には、エッセイを入力とする回帰モデル(4.2.1節を参照)を用いる。

表2 使用する文法特徴量

文法特徴量	説明
g item sum	延べ使用数の合計
g item 9 lev	延べ使用数, CEFR-J レベルごとに合計
g type sum	異なり数の合計
g type 9 lev	異なり数, CEFR-J レベルごとに合計
g item 6	延べ使用数, 評価値と相関の高い6項目
g type 25	使用の有無, 評価値と相関の高い25項目
err sum	誤り数の合計
err 7 gr	誤りタグを7グループに分類し, 集計
err 24 tag	24種類の誤りタグ, 集計・グループ化なし

4 実験

実験では、3節で述べた文法特徴を用いることで、BERTに基づく採点モデルの性能が向上するかを検証する。

4.1 データセット

本研究では、EFCamDat Cleaned Subcorpus[20]、FCE[21]、W&I+L[22]の3種類のデータを用いた。EFCamDatとW&I+Lには初級～上級の様々なレベルの作文が含まれている一方で、FCEには比較的高いレベルの作文のみが含まれている。モデルの学習に使用したデータの詳細を表3に示す。

EFCamDatには、習熟度レベル1～15(CEFR A1～C1に対応)の作文が収録されている。本研究ではそのうち、文法誤りタグが付与されている498,207件のデータを用いた。データは、train, dev, testに分割し、trainはさらに、文法誤り数推定モデルの学習用と、採点モデルの学習用に分割した。採点モデルの学習データには、全習熟度レベルを含むtrain_scと、FCEとの比較を意図したtrain_b2(レベル9～13)を準備した。devとtestは、文法誤り数推定モデルと採点モデルの評価に共通して用いた。

FCEには、CEFR B2レベルの試験において学習者が書いたエッセイが収録されている。文法誤り情報に加えて、0～5.3点の範囲のスコアが付与されている。本研究では、BEA 2019の文法誤り訂正 Shared Taskで配布されたバージョンを、採点モデルの構築・評価に用いた。

W&I+Lには、学習者と母語話者によって書かれたエッセイが収録されている。文法誤り情報に加えて、学習者に対してはCEFR A1.i～C2.iiの12段階、母語話者に対してはNのラベルが付与されている。本研究では、学習者によって書かれたエッセイのみを、採点モデルの構築・評価に用いた。W&I+Lは、

表3 学習に使用したデータ

Dataset	Essays	Usage
EFCamDat	train_gr	458,913 文法誤り数推定モデル
	train_sc	3,000 採点モデル
	train_b2	2,999 採点モデル
FCE	train	2,108 採点モデル
W&I+L	train_0.9	2,700 採点モデル

test が公開されていないため、dev を評価に用い、train を分割して学習・開発データとして用いた。

4.2 実験設定

文法誤り数推定モデル、及び採点モデルの両方で、BERT を利用してエッセイの分散表現を取得する。BERT は HuggingFace 社が公開している bert-base-uncased を用い、学習データの最大入力長は 512 トークンとした。両モデルとも、BERT のパラメータは固定して学習を行った。

4.2.1 文法誤り数推定モデル

文法誤り数推定モデルは、エッセイの分散表現を入力とし、エッセイに含まれる文法誤り数を予測する。文法誤り数は、0 から 1 の範囲に正規化した。エッセイの分散表現を入力する線形層の隠れ層の数は 3、各層のノード数は 256 とした。活性化関数には relu を用いた。損失関数は平均二乗誤差 (MSE) を用い、最適化アルゴリズムは Adam、学習率 $8e-6$ 、バッチサイズ 16 とし、60 エポック学習を行った。

4.3 採点モデル

採点モデルは、エッセイの分散表現と文法特徴量を入力とし、エッセイのスコアを予測する。エッセイの分散表現と文法特徴のベクトルを結合したものを線形層への入力とした。EFCamDat と W&I+L では、エッセイに付与されている評価値は「レベル」であるが、これを連続値とみなすことで、回帰モデルを学習した。エッセイのスコアは -1 から 1 の範囲に正規化した。線形層の隠れ層の数は 5、各層のノード数は 768, 1024, 512, 256, 64 とした。活性化関数には relu を用い、dropout = 0.2 に設定した。損失関数は MSE を用い、最適化アルゴリズムは Adam、学習率 $1e-5$ 、バッチサイズ 16 とし、120 エポック学習を行った。

ベースラインとしては、エッセイの分散表現のみを入力とするモデルを準備した。エポック数を 60 とした以外は、文法特徴量を使う場合と同一の設定とした。

表4 文法誤り数推定モデルの結果 (ピアソンの積率相関係数)。err 7 gr と err 24 tag の値は各相関係数の平均値。

Features	PCC
err sum	0.6772
err 7 gr	0.4606
err 24 tag	0.2771

4.4 実験結果

4.4.1 文法誤り数推定モデル

表 4 は、推定された文法誤り数と正解データとの相関係数を示している。文法誤りの総数の推定では、比較的強い正の相関 (err sum = 0.6772) が見られた一方で、推定する誤りカテゴリ・タグの数が増加すると ($1 \rightarrow 7 \rightarrow 24$)、相関係数は低くなっていった。これは、カテゴリ・タグが細分化されることで、各文法誤りの頻度が低くなるのが影響していると考えられる。7つの誤りカテゴリを予測するモデル (err 7 gr) について、各誤りカテゴリの頻度と相関係数の関係を確認したところ、誤り数が最も多い W_lang (スペルや正書法の誤り) で相関係数が最も高く (0.6192)、誤り数が少なくなるにつれて相関係数が低くなっていくことが確認された。相関係数が 0.3978 と最も低くなった AG (数の一致に関する誤り) は、EFCamDat dev で 1,614 例しかないのに対して、W_lang は 13,285 例見つかった。また、頻度が低い誤りカテゴリは、多くの作文で当該カテゴリの誤り数が 0 となっていた。

4.4.2 採点モデル

表 5 は、推定されたエッセイのスコアと正解スコアとの相関係数を示している。実験は異なるシード値によって 3 回行い、それらの結果の平均値を表に掲載した。

PF を用いても、BERT に基づく AES モデルの性能は向上しなかった。PF を用いたモデルとベースラインの相関係数は同等であることがほとんどだが、g item sum と g item 9 lev (文法の延べ使用数) は、高いレベルの学習者のデータ (EFCamDat_b2, FCE) で用いると、相関係数が悪化した。延べ使用数は、同一の文法事項を繰り返し使っている場合でも、その数値は大きくなる。このような作文は、低いレベル帯では問題にならないものの、高いレベル帯ではネガティブな印象になると考えられる。実際、FCE train において、文法の延べ使用数と異なり数の差

表 5 採点モデルの結果 (ピアソンの積率相関係数)

Model	EFCamDat.sc	EFCamDat.b2	FCE	W&I+L
baseline	0.9098	0.7981	0.6645	0.8168
g item sum	0.8973	0.7625	0.6395	0.8141
g item 9 lev	0.9012	0.7698	0.6404	0.8167
g type sum	0.9116	0.7827	0.6618	0.8146
g type 9 lev	0.9128	0.7948	0.6621	0.8212
g item 6	0.9106	0.7998	0.6610	0.8162
g type 25	0.9136	0.8025	0.6571	0.8229
err sum	0.9089	0.7860	0.6598	0.8185

(平均=62.91) が 40 以下の作文のスコア平均は 3.08 だったのに対して、差が 80 以上の作文のスコア平均は 3.79 であった。延べ使用数に基づく特徴量はこの差を表現できていないと考えられる。これに対して、g type 25 (文法の異なり数) は、FCE 以外のデータセットにおいて、ごく僅かな差であるものの、異なるシード値での 3 回の実験全てでベースラインの性能を上回った。本研究の手法でははっきりとした PF の効果は確認できなかったため、より効果的な PF の活用方法を検討することが今後の課題である。

NF については、文法誤り数推定モデルの性能が比較的高かった err sum のみ、同モデルによる推定値を用いた実験を行った。推定された文法誤りの総数を用いた場合、採点モデルの性能は向上しなかった。しかし W&I+L においては、ごく僅かな差であるものの、異なるシード値での 3 回の実験全てでベースラインの性能を上回った。文法誤り数推定モデル (err sum) が推定する誤り数と真の誤り数の相関係数は 0.6772 (表 5) であるため、NF を用いること自体の効果がないのか、文法誤り数の推定精度が低いことが問題なのかが不明である。

4.4.3 オラクルデータによる検証

NF の効果の有無を明らかにするため、オラクルデータとして、データセットに付与されている文法誤り情報を用いた検証を行った。表 6 に、エッセイの予測スコアと正解スコアとの相関係数を示す。

学習者の習熟度が高いデータ (EFCamDat.b2, FCE) においては、NF を追加してもモデル性能の向上は見られなかった。高いレベル (少なくとも本実験で対象となっていた CEFR B2 付近) の学習者は、文法をある程度使いこなせることが当たり前であり、文法誤りという観点からはあまり差がないことが示唆される。一方で、様々なレベルの作文を含む EFCamDat.sc と W&I+L においては、相反する結果となった。EFCamDat.sc では、NF を追加してもモ

表 6 オラクルデータの結果 (ピアソンの積率相関係数)

Model	EFCamDat.sc	EFCamDat.b2	FCE	W&I+L
baseline	0.9098	0.7981	0.6645	0.8168
err sum	0.9073	0.7922	0.6576	0.8673
err 7 gr	0.9078	0.8070	0.6619	0.8783
err 24 cat	0.9083	0.8038	0.6497	0.8810

デル性能は向上しなかったが、W&I+L では相関係数が 0.05~0.07 程度上昇した。W&I+L での予測スコアと正解スコアの散布図を付録 B に示す。err sum を用いることで、正解スコアとの差異が大きい点 (特に、正解スコアより予測スコアが高い点) がベースラインよりも減少していた。また、W&I+L において異なる 3 つの NF の効果を比較すると、誤り情報を集計するよりもタグレベルの情報を維持する方が相関係数が高くなった。

学習者のレベルが様々である場合、レベルによって文法の出来栄が異なっていると想定できるため、W&I+L の結果の方が直観に合致している。EFCamDat で NF の効果がなかった原因として考えられるものの 1 つは、EFCamDat に付与されている文法誤り情報の正確さである。FCE や W&I+L のアノテーションが少数の専門家によってなされたのに対して、EFCamDat のアノテーションは、オンライン英会話学校のライティング課題に対して教師が文法誤りを訂正した結果に基づいている。そのため、訂正の一貫性や網羅性に問題があると言われており、これが結果に影響を与えた可能性がある。

2 つ目に、EFCamDat のエッセイは W&I+L と比べて文法誤りの数が少ない傾向があること (100 語あたり 8.93 個 vs. 10.83 個; 付録 C) が考えられる。これは、EFCamDat のエッセイが短いこと (平均 120.98 語 vs. 平均 200.95 語)、上述のアノテーションの網羅性の問題が関連していると考えられる。高レベル帯の文法誤り数は両データセットでほぼ同じなのに対して、文法特徴が効果的な低~中レベルの作文ではその差異が大きくなっていた。

5 おわりに

本研究では、AES で文法特徴を考慮することの効果について検証した。PF は、本研究の活用手法ではモデル性能の向上に寄与しなかったが、NF は幅広いレベルの学習者を採点する場合には有効であることが示唆された。PF を活用する手法や、高レベルの学習者に対して有効な言語特徴を検討することが今後の課題である。

謝辞

本研究の一部は JSPS 科研費 JP21H05054, JST 科学技術イノベーション創出に向けた大学フェローシップ創設事業 JPMJFS2137 の助成を受けたものである。

参考文献

- [1] John A Hawkins and Luna Filipović. **Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework**. Cambridge University Press, 2012.
- [2] Zixuan Ke and Vincent Ng. Automated essay scoring: A survey of the state of the art. In **Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence**, 2019.
- [3] Masaki Uto. A review of deep-neural automated essay scoring models. **Behaviormetrika**, Vol. 48, No. 2, pp. 459–484, 2021.
- [4] Jill Burstein, Martin Chodorow, and Claudia Leacock. Automated essay evaluation: The criterion online writing service. **AI Magazine**, Vol. 25, No. 3, pp. 27–36, 2004.
- [5] Sowmya Vajjala. Automated assessment of Non-Native learner essays: Investigating the role of linguistic features. **International Journal of Artificial Intelligence in Education**, Vol. 28, No. 1, pp. 79–105, 2018.
- [6] Ronan Cummins and Marek Rei. Neural multi-task learning in automated assessment. **arXiv**, Vol. arXiv: 1801.06830, , 2018.
- [7] Kaveh Taghipour and Hwee Tou Ng. A neural approach to automated essay scoring. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 1882–1891, 2016.
- [8] Dimitrios Alikaniotis, Helen Yannakoudakis, and Marek Rei. Automatic text scoring using neural networks. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 715–725, 2016.
- [9] 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, Volume 1 (Long and Short Papers)**, pp. 4171–4186, 2019.
- [10] Farah Nadeem, Huy Nguyen, Yang Liu, and Mari Ostendorf. Automated essay scoring with discourse-aware neural models. In **Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications**, pp. 484–493, 2019.
- [11] Ruosong Yang, Jiannong Cao, Zhiyuan Wen, Youzheng Wu, and Xiaodong He. Enhancing automated essay scoring performance via fine-tuning pre-trained language models with combination of regression and ranking. In **Findings of the Association for Computational Linguistics: EMNLP 2020**, pp. 1560–1569, 2020.
- [12] Tirthankar Dasgupta, Abir Naskar, Lipika Dey, and Rupsa Saha. Augmenting textual qualitative features in deep convolution recurrent neural network for automatic essay scoring. In **Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications**, pp. 93–102, 2018.
- [13] Masaki Uto, Yikuan Xie, and Maomi Ueno. Neural automated essay scoring incorporating handcrafted features. In **Proceedings of the 28th International Conference on Computational Linguistics**, pp. 6077–6088, 2020.
- [14] Stefano Bannò and Marco Matassoni. Cross-corpora experiments of automatic proficiency assessment and error detection for spoken English. In **Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)**, pp. 82–91, 2022.
- [15] Council of Europe. **Common European Framework of Reference for Languages: Learning, Teaching, Assessment**. Cambridge University Press, 2001.
- [16] 投野 由紀夫 (編). 平成 24 年度～平成 27 年度科学研究費補助金 (基盤研究 (A)) 研究課題番号 24242017 研究成果報告書 学習者コーパスによる英語 CEFR レベル基準特性の特定と活用に関する総合的研究. 2016.
- [17] 石井 康毅. CEFR-J Grammar Profile 文法項目頻度分析プログラム. <http://www.cefr-j.org/download.html> (2022 年 10 月 20 日ダウンロード), 2020.
- [18] Christopher Bryant, Mariano Felice, and Ted Briscoe. Automatic annotation and evaluation of error types for grammatical error correction. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 793–805, 2017.
- [19] Jeroen Geertzen, Theodora Alexopoulou, and Anna Korhonen. Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge open language database (EFCamDat). In **Selected Proceedings of the 2012 Second Language Research Forum: Building Bridges between Disciplines**, pp. 240–254, 2014.
- [20] Itamar Shatz. Refining and modifying the EFCAMDAT: Lessons from creating a new corpus from an existing large-scale English learner language database. **International Journal of Learner Corpus Research**, Vol. 6, No. 2, pp. 220–236, 2020.
- [21] Helen Yannakoudakis, Ted Briscoe, and Ben Medlock. A new dataset and method for automatically grading ESOL texts. In **Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies**, pp. 180–189, 2011.
- [22] Christopher Bryant, Mariano Felice, Øistein E. Andersen, and Ted Briscoe. The BEA-2019 shared task on grammatical error correction. In **Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications**, pp. 52–75, 2019.

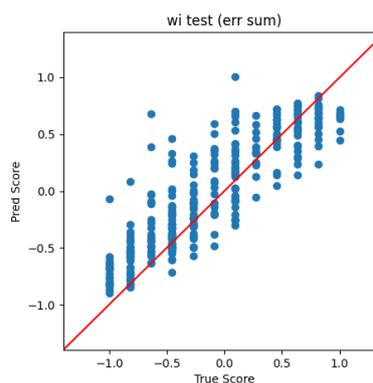
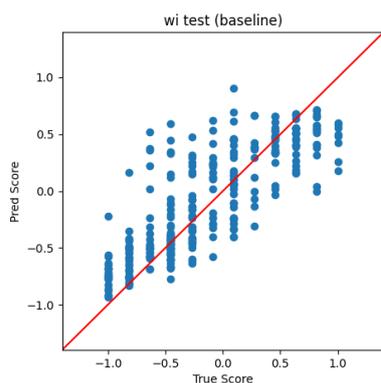
A ERRANT の誤りタグのグループ化

Lexical は同じ品詞の別の語に訂正する誤り, form は語の選択は正しいが活用形が間違っている誤り, VT は動詞の時制の誤り, PAP は前置詞・冠詞・所有格の誤り, AG は数の一致に関する誤り, W_lang はスペル・正書法の誤り, Others は語順及びその他の誤りを表す. VT, PAP, AG は [14] に基づいている.

Group	Errant tag	Meaning
Lexical	ADJ	Adjective
	ADV	Adverb
	CONJ	Conjunction
	NOUN	Noun
	PRON	Pronoun
	VERB	Verb
Form	ADJ:FORM	Adjective Form
	CONTR	Contraction
	MORPH	Morphology
	VERB:FORM	Verb Form
	VERB:INFL	Verb Inflection
VT	VERB:TENSE	Verb Tense

Group	Errant tag	Meaning
PAP	DET	Determiner
	NOUN:POSS	Noun Possessive
	PART	Particle
	PREP	Preposition
AG	NOUN:INFL	Noun Inflection
	NOUN:NUM	Noun Number
	VERB:SVA	Subject-Verb Agreement
W_lang	ORTH	Orthography
	PUNCT	Punctuation
	SPELL	Spelling
Others	OTHER	Other
	WO	Word Order

B W&I+L での baseline と err sum の結果



C EFCamDat と W&I+L の語数と 100 語あたりの文法誤り数

表 7 EFCamDat

CEFR	Level	# Tokens	# Errors
A1	1	45.01	12.59
	2	58.00	10.82
	3	62.32	10.30
A2	4	72.19	10.39
	5	88.15	9.43
	6	90.72	9.09
B1	7	110.69	8.83
	8	118.81	7.60
	9	130.60	7.45
B2	10	142.10	6.89
	11	173.19	6.34
	12	171.18	6.05
C1	13	179.77	5.34
	14	188.63	4.42
	15	183.36	3.42

表 8 W&I+L

CEFR	Level	# Tokens	# Errors
A1	1	65.44	19.84
	2	100.45	16.73
A2	3	146.30	16.41
	4	164.50	14.91
B1	5	203.62	13.52
	6	205.87	11.59
B2	7	218.71	9.55
	8	234.97	8.00
C1	9	260.15	6.78
	10	263.88	5.56
C2	11	273.99	4.43
	12	273.51	2.63