

自閉スペクトラム症の眼球運動による 言語モデルのファインチューニング

前田ありさ 大関洋平

東京大学

{alyssa-maeda,oseki}@g.ecc.u-tokyo.ac.jp

概要

自閉スペクトラム症 (ASD) は、定型発達の人とは異なるコミュニケーションの取り方や行動パターンで特徴付けられる発達障害であり、文処理に困難が伴うことが多い。本研究では、ASD の文処理を理解し支援するために、ASD の人の文処理データを用いて言語モデルをファインチューニングするアプローチを提案する。このアプローチにより、ASD の人々の文処理の特徴を反映したモデルの構築が行われ、ASD の統語論処理能力に関する理解の深化や、ASD の人々の読解ニーズに応じた支援ツールの実現に向けた可能性が示唆された。

1 はじめに

自閉スペクトラム症 (ASD) は、定型発達の人とは異なるコミュニケーションの取り方や、反復的・こだわりの強い行動を特徴とする発達障害の一つである [1]。初期症状には言語獲得の遅れなどがあり、ASD が言語やコミュニケーションと密接に関係していると考えられている [1]。これらに関連し、ASD の人は文章を読むことを苦手とすることがある [2, 3, 4]。

そのため、ASD の人々の文処理を支援するツールの開発が重要とされる。しかし、ASD の人々の言語処理を支援する研究は限られており、現在、この分野で知られているツールは 2 つのみである。1 つ目は会話の文脈を基に GPT-4 を使用してメッセージの意味やトーンを説明し、曖昧な表現を明確化することで ASD のユーザーに向けて会話言語を簡略化するものである [5]。2 つ目は長文や難解な文を分割し、まれに使われる単語の定義や同義語を提供し、慣用句を説明し、文書の主旨を要約する文書要約ソフトである [6, 7]。これらのツールは、ASD の人々の読解支援に貢献しているが、主に語用論的課題に

焦点を当てており、ASD の読者が直面する他の課題には十分対応していない。そのため、より幅広い課題に対応できるツールの開発が求められている。

近年、視線データや EEG データといった認知データを活用し、人間の言語処理を反映する言語モデルの開発が進んでいる。このアプローチにより、読解時間予測などの上流タスクから感情分析などの下流タスクまで、幅広い分野でモデル性能が向上している [8, 9, 10, 11]。本研究では ASD の視線データを用いて言語モデルをファインチューニングし、ASD 特有の文処理に対応できるシステムを目指す。このシステムの実現により、ASD の言語処理能力の研究を促進し、さらに ASD の文処理を支援する新たなツールの開発につながることが期待される。

2 実験

2.1 実験設定

本研究では、事前学習済みの GPT-2 small を ASD-eye-tracking-reading dataset [12] で学習させた。このデータセットは、計 20 の文書 (合計 216 文) で構成され、ASD の成人および定型発達の成人が文章読解課題に取り組む際の視線データを含んでいる。更に、本研究ではさまざまな学習条件を設定した上で GPT-2 のファインチューニングを行った。それぞれの学習条件については以下の通りである¹⁾。

まず、各学習条件において、言語モデルを ASD データまたは定型発達データで学習させた。その後、ASD データで学習させたモデルと定型発達データで学習させたモデルを比較し、それぞれのタスクにおける性能が予想通りであるかを検証した。また、すべてのモデルを事前学習済みの GPT-2 と比較し、ファインチューニングがこれらのタスクにおけ

1) これらの学習条件に加え、予備実験ではトークン単位での学習と単語単位での学習も行なったが、トークン単位での学習は期待していた結果が得られなかった。

る性能向上に寄与したかどうかを評価した。

次に考慮されたものは学習に用いる時間データである。ASD-eye-tracking-reading dataset には、各単語に対する視認時間の合計 (TOTAL) および平均凝視時間 (AFD) の情報が含まれている。これらの生の視認時間データに加え、これらの視認時間を対数変換したデータを用いてファインチューニングを行なった。具体的には、TOTAL、 $\log(\text{TOTAL})$ 、 $\log_2(\text{TOTAL})$ 、 $\log_{10}(\text{TOTAL})$ 、AFD、 $\log(\text{AFD})$ 、 $\log_2(\text{AFD})$ 、および $\log_{10}(\text{AFD})$ を用いて GPT-2 の学習を行なった。

最後に 2 つの異なる研究に基づく学習手法を採用した。1 つ目の手法は染谷&大関 [8] によるもので、ファインチューニングにモデルのサプライズ予測を実際の読解時間に一致させるために L1Loss を使用するものである。2 つ目の手法は Kieglend ら [9] によるもので、モデルのサプライズ出力を基に人間の読解時間を直接予測する線形回帰器のパラメータを暗黙的に調整し、モデルの予測値と実際の読解時間との平均二乗誤差 (MSE) を最小化するという内容である。本研究では、前者の手法を「L1Loss を用いた学習法」、後者の手法を「MSE を用いた学習法」として言及する。ただし、これら 2 つの手法を区別する要素はこれだけに限らないことを理解しつつ、簡潔さのためこのように表現する。

このように 32 の学習条件でファインチューニングを行ない、3 つのタスクで評価することにより、ファインチューニングしたモデルがどれほど ASD の文処理の特徴を捉えられるのか、ASD の文処理への理解や支援にどのように役立てるのかを検討した。またその他全条件に共通のハイパーパラメータについては、付録 A に示した。

2.2 タスク①：読み時間予測

本研究では、ASD-eye-tracking-reading データセットを 8:1:1 の割合で学習用、検証用、テスト用データに分割した。また、このデータセットは比較的小規模であったため、モデルの性能をより信頼性高く評価するために 10 分割交差検証を実施し、モデルの汎化性能を評価した。

評価指標 それぞれのモデルがどれほど ASD の文処理の特徴を捉えられたかについては心理学的予測精度 (Psychometric Predictive Power: PPP) を用いる。本実験では先行研究 [13] の手法に基づき、まずベースラインモデルとして単語の長さや頻度など

学習条件		$\Delta \text{LogLik} (\uparrow)$	
時間データ	手法	ASD	NT
TOTAL	L1Loss [8]	4.212	4.111
$\log(\text{TOTAL})$	L1Loss [8]	7.415	7.409
$\log_2(\text{TOTAL})$	L1Loss [8]	7.479	7.472
$\log_{10}(\text{TOTAL})$	L1Loss [8]	7.309	7.295
AFD	L1Loss [8]	4.082	4.520
$\log(\text{AFD})$	L1Loss [8]	7.405	7.392
$\log_2(\text{AFD})$	L1Loss [8]	7.442	7.458
$\log_{10}(\text{AFD})$	L1Loss [8]	7.289	7.295
TOTAL	MSE [9]	7.300	7.439
$\log(\text{TOTAL})$	MSE [9]	7.406	7.379
$\log_2(\text{TOTAL})$	MSE [9]	7.397	7.379
$\log_{10}(\text{TOTAL})$	MSE [9]	7.406	7.427
AFD	MSE [9]	7.284	7.127
$\log(\text{AFD})$	MSE [9]	7.406	7.380
$\log_2(\text{AFD})$	MSE [9]	7.393	7.381
$\log_{10}(\text{AFD})$	MSE [9]	7.406	7.426
pretrained GPT-2		7.399	

表 1 TOTAL 読み時間を予測した際の PPP (ΔLogLik)。

の特徴量で読み時間のモデリングを行った。その後、各条件で学習された言語モデルから得られたサプライズを追加し、モデリング性能の向上度 (ΔLogLik) を評価した。ベースライン回帰モデルには、以下の重回帰モデルを用いる：

$$\begin{aligned} \text{TOTAL} \sim & \text{length}(w_i) + \text{freq}(w_i) \\ & + \text{length}(w_{i-1}) + \text{freq}(w_{i-1}) \\ & + \text{length}(w_{i-2}) + \text{freq}(w_{i-2}) \end{aligned}$$

PPP が高いほど、ASD の読み行動をより正確に予測できるとされ、ASD モデルは定型発達モデルや事前学習済み GPT-2 を上回る場合に優れた予測モデルと見なされる。

結果と考察 TOTAL 及び AFD の読み時間予測の結果はそれぞれ表 1 と表 2 で記されている。NT は定型発達を意味する。また $\log(\text{TOTAL})$ 及び $\log(\text{AFD})$ の読み時間予測の結果は付録 B に載せている (それぞれ表 7 と表 8)。黄色い線が引かれているものは、ASD の読み時間予測において、ASD モデルの PPP が定型発達モデル及び事前学習された GPT-2 の両方を上回ったことを示している。これらのモデルは ASD の文処理の特徴を正確に捉えたモデルとして、他のタスクでの性能をも検証した。

2.3 タスク②：文法処理

評価指標 読み時間予測で ASD の読み時間を正確に予測できたモデルは、更に BLiMP[14] で評価された。BLiMP は 12 種類の言語現象を対象に、文法的に正しい文と誤った文のペアを用いて、言語モデ

学習条件		Δ LogLik (\uparrow)	
時間データ	手法	ASD	NT
TOTAL	L1Loss [8]	<u>2.384</u>	2.087
log(TOTAL)	L1Loss [8]	<u>3.559</u>	3.554
log ₂ (TOTAL)	L1Loss [8]	<u>3.481</u>	3.478
log ₁₀ (TOTAL)	L1Loss [8]	<u>3.582</u>	3.574
AFD	L1Loss [8]	<u>1.793</u>	1.742
log(AFD)	L1Loss [8]	<u>3.568</u>	3.567
log ₂ (AFD)	L1Loss [8]	<u>3.525</u>	3.516
log ₁₀ (AFD)	L1Loss [8]	<u>3.583</u>	3.581
TOTAL	MSE [9]	3.478	<u>3.527</u>
log(TOTAL)	MSE [9]	3.556	<u>3.599</u>
log ₂ (TOTAL)	MSE [9]	3.551	<u>3.598</u>
log ₁₀ (TOTAL)	MSE [9]	3.556	<u>4.222</u>
AFD	MSE [9]	<u>3.539</u>	3.365
log(AFD)	MSE [9]	<u>3.552</u>	3.593
log ₂ (AFD)	MSE [9]	3.544	3.593
log ₁₀ (AFD)	MSE [9]	3.552	3.567
pretrained GPT-2		3.569	

表2 AFD 読み時間を予測した際の PPP (Δ LogLik)。

ルの文法知識を評価するベンチマークである。このベンチマークを用いて本研究では ASD の人々が文法処理を困難とするという実世界の研究を再現できるかを検証する。この結果を再現できたモデルが ASD の読み行動を最も正確に予測すると仮定し、ASD で観察される困難な言語現象とモデルの予測が一致するかを確認した。

結果と考察 読み時間予測で ASD の読み時間を正確に予測できたモデルを BLiMP で評価した結果が表3で記されている。ASD の人は定型発達の人より文法処理を苦手としていること [15, 16, 17, 18, 19] から、表3で黄色い線で引かれているものは、ASD モデルの BLiMP のスコアが定型発達モデルのスコアを下回ったものを指す。これらのモデルは ASD の文法処理を捉えたものとして更にそれぞれの言語現象でのスコアを調査する。

ASD の文法処理を捉えたモデルの言語現象別の結果は表4に示されている。青い線で示された現象は、ASD の文法処理に関する先行研究と一致するスコアが得られた言語現象を表している。具体的には、ASD の人が定型発達の人と比べて ARG STR[19] や BINDING[20]、D-N AGR[21]、S-V AGR[22] を苦手とし、また定型発達の人と同程度に CTRL/RAIS[23] を処理することができるという先行研究の結果と一致する結果が得られた。一方で ASD の文法処理に関する既存の研究と矛盾する結果が得られた言語現象もあり、それらは灰色の線で示されている。具体的には、ASD の人が Wh 疑問文 [15] や再帰代名詞 [20] の処理に困難を示すという先行研究に対し、本研究

ではそれらの現象を再現する結果が得られなかった。線が引かれていない言語現象については、ASD の文法処理に関する先行研究がまだ存在しない現象を指し、これらは ASD の文処理研究における新たな方向性として検討されるべきである。

これらの結果から、ASD の文処理データで言語モデルをファインチューニングし、文法知識を評価することで、ASD の人々の文法処理の特徴の大部分を再現できたことが示唆される。また、これまで先行研究で扱われていない言語現象について、新たな視点を提供し、ASD の文処理研究におけるさらなる発展の方向性を示唆している。

学習条件		Overall BLiMP	
時間データ	手法	ASD	NT
log(TOTAL)	L1Loss [8]	<u>79.41</u>	79.40
log ₂ (TOTAL)	L1Loss [8]	79.44	79.44
log ₁₀ (TOTAL)	L1Loss [8]	<u>79.37</u>	<u>79.42</u>
log(AFD)	L1Loss [8]	<u>79.41</u>	79.40
log ₁₀ (AFD)	L1Loss [8]	<u>79.37</u>	<u>79.42</u>
log(TOTAL)	MSE [9]	<u>79.19</u>	79.16
log ₂ (TOTAL)	MSE [9]	<u>79.19</u>	79.00
log ₁₀ (TOTAL)	MSE [9]	<u>79.19</u>	79.16
log(AFD)	MSE [9]	<u>79.19</u>	79.16
log ₁₀ (AFD)	MSE [9]	<u>79.19</u>	79.15
pretrained GPT-2		79.19	

表3 BLiMP の各言語現象のスコアのマクロ平均をとった総合的なスコア。

言語現象	log ₁₀ (TOTAL)		log ₁₀ (AFD)	
	ASD	NT	ASD	NT
ANAPHOR AGR	<u>97.35</u>	97.25	<u>97.35</u>	97.25
ARG STR	75.27	<u>75.33</u>	75.26	<u>75.34</u>
BINDING	79.09	<u>79.16</u>	79.09	<u>79.16</u>
CTRL/RAIS	80.60	80.60	80.58	80.58
D-N AGR	93.75	<u>93.80</u>	93.75	<u>93.81</u>
ELLIPSIS	<u>87.25</u>	87.15	<u>87.20</u>	87.10
FILLER GAP	81.44	81.44	81.40	81.40
IRREGULAR FORMS	<u>93.15</u>	93.00	<u>93.15</u>	93.10
ISLAND EFFECTS	60.00	<u>60.03</u>	<u>60.06</u>	60.00
NPI LICENSING	68.74	<u>68.84</u>	68.69	<u>68.84</u>
QUANTIFIERS	79.40	<u>79.60</u>	79.50	<u>79.65</u>
S-V AGR	<u>88.28</u>	<u>88.37</u>	<u>88.27</u>	<u>88.35</u>

表4 表3において ASD の文法処理を捉えたモデルを、BLiMP の各言語現象で評価したスコア。

2.4 タスク③：文章平易化

読み時間予測で ASD の読み時間を正確に予測できたモデルは、更に文章平易化タスクで追加学習が

行われた。文章平易化タスクの学習手法については Agarwal[24] の研究を参考にした。本研究において文章平易化タスクが選ばれた理由は、ASD の文処理データでファインチューニングを行なった言語モデルが ASD の読者の認知特性に適応し、生成文の読みやすさを向上させることが期待できるためである。なぜなら、文処理データでファインチューニングを行ったモデルは、ASD の人が読み滞る単語（すなわち、読み時間が長い単語）に対して高いサプライズを予測するように調整されている。この予測結果は、長い読み時間を要する単語が生成されにくくなることを意味し、その結果、ASD の読者にとって理解しやすい簡素化されたテキストが生成されることが考えられる。

評価指標 追加学習されたモデルは Turk Corpus[25] のテストデータセットに基づいて文を平易化するタスクを実行し、その出力は SARI と FKGL スコア [26] で評価された。SARI はモデルの出力文を元の文及びその元の文を平易化した参考文と比較し、単語の追加、保持、削除の適切さを評価する指標である。一方、FKGL は文中の単語数や単語ごとの音節数に基づいて計算され、スコアが低いほど低学年向けで読みやすい文を示す。

結果と考察 文章平易化における結果は表 5 で記されている。黄色い線が引かれているものは、SARI と FKGL に基づき、ASD モデルが定型発達モデル及び認知データでファインチューニングされていない GPT-2 より読みやすい文章を出力したことを示している。このように ASD の視線データで言語モデルをファインチューニングし、更に文章平易化タスクで追加学習を行うことで、本研究は ASD 向けの文章平易化ツールの実現に向けた第一歩となることを示している。一方で、文章平易化タスクの評価に使用した指標やデータセットは ASD に特化したものではないため、出力が ASD の読解に適しているかを検証することが今後の課題である。

3 おわりに

本研究では、ASD の文処理データでファインチューニングされた言語モデルが、ASD の文処理の特徴を正確に捉え ASD の人にとって読みやすい文章を生成する可能性を検討した。その結果、ASD 特有の認知データを活用することで、ASD 特有の文処理への理解や支援に向けた大きな可能性が示唆された。

学習条件		SARI (↑)		FKGL (↓)	
時間データ	手法	ASD	NT	ASD	NT
log(TOTAL)	L1Loss [8]	37.19	37.64	9.808	9.364
log ₂ (TOTAL)	L1Loss [8]	36.34	37.80	9.518	7.975
log ₁₀ (TOTAL)	L1Loss [8]	38.16	37.61	7.952	10.79
log(AFD)	L1Loss [8]	36.94	37.03	9.151	10.52
log ₁₀ (AFD)	L1Loss [8]	37.00	37.54	11.85	8.316
log(TOTAL)	MSE [9]	38.12	37.82	12.05	15.00
log ₂ (TOTAL)	MSE [9]	37.63	38.24	14.13	7.362
log ₁₀ (TOTAL)	MSE [9]	38.35	36.78	4.742	15.83
log(AFD)	MSE [9]	28.39	37.04	13.96	9.356
log ₁₀ (AFD)	MSE [9]	37.83	37.22	8.007	9.996
pretrained GPT-2		37.28		8.189	

表 5 文章平易化タスクにおける SARI と FKGL のスコア。

謝辞

文処理データのファインチューニングにあたり、染谷大河氏が提供して下さったコードを大いに参考にさせていただきました。この場を借りて深く感謝申し上げます。また本研究は、JST さきがけ JPMJPR21C2 および JSPS 科研費 24H00087/23H05493 の支援を受けたものです。

参考文献

- [1] American Psychiatric Association. **Diagnostic and statistical manual of mental disorders**. American Psychiatric Association Publishing, 5th ed, text rev. edition, 2022.
- [2] Gwyn W Senokossoff. Developing reading comprehension skills in high-functioning children with autism spectrum disorder: A review of the research, 1990–2012. **Reading & Writing Quarterly**, 32(3):223–246, 2016.
- [3] Ryan P Grimm, Emily J Solari, Nancy S McIntyre, Matthew Zajic, and Peter C Mundy. Comparing growth in linguistic comprehension and reading comprehension in school-aged children with autism versus typically developing children. **Autism Research**, 11(4):624–635, 2018.
- [4] Catherine RG Jones, Francesca Happé, Hannah Golden, Anita JS Marsden, Jenifer Tregay, Emily Simonoff, Andrew Pickles, Gillian Baird, and Tony Charman. Reading and arithmetic in adolescents with autism spectrum disorders: peaks and dips in attainment. **Neuropsychology**, 23(6):718, 2009.
- [5] Rukhshan Haroon and Fahad Dogar. Twips: A large language model powered texting application to simplify conversational nuances for autistic users. In **Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility**, ASSETS '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [6] Arlinda Cerga-Pashoja, Jorge Gaete, Antoneta Shishkova, and Vesna Jordanova. Improving reading in adolescents and adults with high-functioning autism through an assistive technology tool: a cross-over multinational study. **Frontiers in psychiatry**, 10:546, 2019.
- [7] Eduard Barbu, M. Teresa Martín-Valdivia, Eugenio Martínez-Cámara, and L. Alfonso Ureña-López. Language technologies applied to document simplification for helping autistic people. **Expert Systems with Applications**, 42(12):5076–5086, 2015.
- [8] 染谷大河, 大関洋平. 認知ファインチューニング：眼球運動

による大規模言語モデルのファインチューニング. In **言語処理学会第30回年次大会 発表論文集**, pages 1902–1907. 言語処理学会, 2024.

- [9] Samuel Kieglend, Ethan Gottlieb Wilcox, Afra Amini, David Robert Reich, and Ryan Cotterell. Reverse-engineering the reader, 2024.
- [10] Nora Hollenstein, Maria Barrett, Marius Troendle, Francesco Bigi-olli, Nicolas Langer, and Ce Zhang. Advancing nlp with cognitive language processing signals, 2019.
- [11] Abhijit Mishra, Diptesh Kanojia, Seema Nagar, Kuntal Dey, and Pushpak Bhattacharyya. Leveraging cognitive features for sentiment analysis. In Stefan Riezler and Yoav Goldberg, editors, **Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning**, pages 156–166, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [12] Victoria Yaneva. **Assessing text and web accessibility for people with autism spectrum disorder**. PhD thesis, University of Wolverhampton, 2016.
- [13] Tatsuki Kuribayashi, Yohei Oseki, and Timothy Baldwin. Psychometric predictive power of large language models, 2023.
- [14] Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohananey, Wei Peng, Sheng-Fu Wang, and Samuel R. Bowman. Blimp: The benchmark of linguistic minimal pairs for english, 2023.
- [15] Anthony Goodwin, Deborah Fein, and Letitia R. Naigles. Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. **Autism Research**, 5(2):109–123, 2012.
- [16] Stephanie Durrleman, Loyse Hippolyte, Sandrine Zufferey, Katia Iglesias, and Nouchine Hadjikhani. Complex syntax in autism spectrum disorders: A study of relative clauses. **International Journal of Language & Communication Disorders**, 50(2):260–267, 2015.
- [17] Stephanie Durrleman and Sandrine Zufferey. **Investigating complex syntax in autism**, pages 405–415. 01 2013.
- [18] Helen Tager-Flusberg. Sentence comprehension in autistic children. **Applied psycholinguistics**, 2(1):5–24, 1981.
- [19] Stephanie Durrleman, Hélène Delage, Philippe Prévost, and Laurice Tuller. The comprehension of passives in autism spectrum disorder. **Glossa: a journal of general linguistics**, 2(1), 2017.
- [20] Alexandra Perovic, Nadezhda Modyanova, and Kenneth Wexler. Comparison of grammar in neurodevelopmental disorders: The case of binding in williams syndrome and autism with and without language impairment. **Language Acquisition**, 20:133–154, 04 2013.
- [21] Stephanie Durrleman and Sandrine Zufferey. The nature of syntactic impairment in autism. **Rivista di Grammatica Generativa**, 34:57–86, 2009.
- [22] Inge-Marie Eigsti and Loisa Bennetto. Grammaticality judgments in autism spectrum disorders: Deviance or delay. **Journal of Child Language**, 19:1–23, 01 2009.
- [23] Vikki Janke and Alexandra Perovic. Intact grammar in hfa? evidence from control and binding. **Lingua**, 164:68–86, 2015.
- [24] Aman Agarwal. Explain to me like i am five – sentence simplification using transformers, 2022.
- [25] Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. Optimizing statistical machine translation for text simplification. **Transactions of the Association for Computational Linguistics**, 4:401–415, 2016.
- [26] Fernando Alva-Manchego, Louis Martin, Carolina Scarton, and Lucia Specia. EASSE: Easier automatic sentence simplification evaluation. In Sebastian Padó and Ruihong Huang, editors, **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations**, pages 49–54,

Hong Kong, China, November 2019. Association for Computational Linguistics.

A ハイパーパラメータ

最適化手法	Adam
学習率	1e-6
エポック数	100
バッチサイズ	16
学習率スケーラ	LinearLR

表 6 文処理データを使用したファインチューニングについて、全学習条件に共通するハイパーパラメーター。

B 読み時間予測の結果

B.1 log(TOTAL) 時間の予測

学習条件		$\Delta \text{LogLik} (\uparrow)$	
時間データ	手法	ASD	NT
TOTAL	L1Loss [8]	2.939	3.077
log(TOTAL)	L1Loss [8]	3.661	3.658
log ₂ (TOTAL)	L1Loss [8]	3.587	3.585
log ₁₀ (TOTAL)	L1Loss [8]	3.693	3.679
AFD	L1Loss [8]	2.905	3.110
log(AFD)	L1Loss [8]	3.677	3.668
log ₂ (AFD)	L1Loss [8]	3.616	3.620
log ₁₀ (AFD)	L1Loss [8]	3.687	3.682
TOTAL	MSE [9]	3.931	3.907
og(TOTAL)	MSE [9]	3.964	3.923
log ₂ (TOTAL)	MSE [9]	3.963	3.923
log ₁₀ (TOTAL)	MSE [9]	3.964	3.962
AFD	MSE [9]	3.955	3.597
log(AFD)	MSE [9]	3.963	3.896
log ₂ (AFD)	MSE [9]	3.960	3.896
log ₁₀ (AFD)	MSE [9]	3.963	3.950
pretrained GPT-2		3.961	

表 7 log(TOTAL) 読み時間を予測した際の PPP (ΔLogLik)。

B.2 log(AFD) 時間の予測

学習条件		$\Delta \text{LogLik} (\uparrow)$	
時間データ	手法	ASD	NT
TOTAL	L1Loss [8]	2.016	1.925
log(TOTAL)	L1Loss [8]	2.563	2.559
log ₂ (TOTAL)	L1Loss [8]	2.478	2.478
log ₁₀ (TOTAL)	L1Loss [8]	2.599	2.592
AFD	L1Loss [8]	1.907	1.982
log(AFD)	L1Loss [8]	2.574	2.572
log ₂ (AFD)	L1Loss [8]	2.520	2.523
log ₁₀ (AFD)	L1Loss [8]	2.601	2.595
TOTAL	MSE [9]	2.708	2.691
log(TOTAL)	MSE [9]	2.734	2.727
log ₂ (TOTAL)	MSE [9]	2.733	2.727
log ₁₀ (TOTAL)	MSE [9]	2.734	2.738
AFD	MSE [9]	2.714	2.358
log(AFD)	MSE [9]	2.735	2.704
log ₂ (AFD)	MSE [9]	2.731	2.705
log ₁₀ (AFD)	MSE [9]	2.735	2.725
pretrained GPT-2		2.744	

表 8 log(AFD) 読み時間を予測した際の PPP (ΔLogLik)。