

モバイル学習者の読書行動に基づく未知単語推定 Unknown Word Estimation Based on Mobile Learners' Reading Behavior

東村 理功¹⁾ バーゴ アンドリユー¹⁾ 岩田 基¹⁾ 黄瀬 浩一¹⁾
Riku Higashimura Andrew Vargo Motoi Iwata Koichi Kise

1 はじめに

語彙の習得は語学学習の基本である。スマートフォンの普及により、語学学習者は日々のスケジュールの中でより柔軟に学習できるようになっている [1, 2, 3, 4]。新たな語彙を習得するためには、学習者個人が意味を知らない単語 (以下、未知単語) を何らかの方法で語学学習の対象にしなければならぬ [5]。読書時に未知単語を記録する方法が考えられるが、大変手間であり、読書自体の妨げにもなる。そのため、学習者が読んだ文書中から自動で未知単語を記録することができれば便利である。そこで、文書中の未知単語を推定する指標の一つとして、単語の難しさを表す出現頻度を用いる方法がある [6]。一般に、出現頻度が低い単語ほど、その単語が利用される機会が少ないため、どの学習者にとっても未知単語である可能性が高い。一方で、単語の難しさは学習者の語彙知識に依存する。例えば、多くの学習者は、ある分野では語彙知識が豊富であっても、他の分野では語彙知識が不足していることがある。それゆえに、出現頻度の高い簡単な単語が未知単語である場合、学習の必要性が高いにも関わらず、出現頻度のみに基づいた学習では学習漏れの可能性がある。したがって、一人一人の学習者に応じた未知単語を推定する必要がある。学習者に依存した読書行動データの一つに読書時間がある。読書時間は、学習者の英文読解能力の指標として有効であることが報告されている [7, 8, 9, 10]。

本稿では、学習者に依存しないテキスト特徴量と、学習者に依存した読書行動データの両方を用いた未知単語推定手法を提案する [11, 12]。ここで、読書行動データとして、スマートフォンのセンサデータで得られる、読書時間、スマートフォンの 3 軸加速度 (2Hz)、読み返し回数を用いる。これらを取り入れることによって、学習者個人の英文読解能力を反映させる。実験の結果、実験参加者 30 名のうち 28 名に対して、読書行動データが未知単語推定に寄与することを確認した。以降、2 章で関連研究、3 章で提案手法、4 章で実験、5 章でまとめと今後の課題について述べる。なお、本研究は大阪府立大学大学院工学研究科倫理委員会の承認を得ていることを付記しておく。

2 関連研究

2.1 文章難易度に関する定量的なテキスト特徴量

学習者の語彙知識と文章の難しさは強く関係していると報告されている [13]。これまで、定量的なテキスト特徴量から文章の難しさをモデル化して予測する手法が提案されている [14, 15, 16]。Brown らは、定量的なテキスト特徴量として、文に含まれる機能語の占める割合、7 文字以上の単語の占める割合などといった文構造の特性を用いている [15]。ここで機能語とは、前置詞や接続詞などのような文法的な関係を表す単語である。Senter らは、信頼性が高いとされている文章の難しさの指標

として、Automated Readability Index (ARI) を提案している [16]。ARI はコンピュータで自動処理することを目的に提案された指標であり、計算対象の文章における平均単語長 α と平均文長 β を用いて (1) 式で算出される。

$$ARI = (4.71 \times \alpha) + (0.5 \times \beta) - 21.43 \quad (1)$$

また、Flesch らは、単語の音節数を用いた文章の難しさの指標として、Flesch Reading Ease (FRS) を提案している [14]。FRS の計算結果は 0-100 までの範囲で示され、60-70 を標準として、得点が高くなるほど読みやすい文章であることを示し、反対に得点が低くなるほど難しい文章であることを示す。計算対象の文章における平均文長 β と平均音節長 γ を用いて (2) 式で算出される。

$$FRS = 206.835 - (1.015 \times \beta) + (84.6 \times \gamma) \quad (2)$$

ここで (1) 式と (2) 式内の平均単語長 α は 1 単語当たりの平均文字数、平均文長 β は 1 文あたりの平均単語数、平均音節長 γ は 1 単語当たりの平均音節数を表す。これらの指標により、単語の出現頻度のみを利用した場合よりも精度が高く、文章の難しさに関するテキスト特徴量を得ることができる。以上のように、これらのテキスト特徴量は、単語の出現頻度に加えて、学習者に依存しない特徴量として未知単語推定に有効であるといえる。

2.2 視点情報に基づいた未知単語推定

学習者の行動データから、学習者個人の理解度や読解力について多くのことを知ることができる。そのような行動データの一つが視線情報である。視線情報はアイトラッカを用いて測定できる。この 10 年間で、視点情報に基づいて未知単語を検出する研究が行われている [17, 18, 19, 20]。Garain らは、PC で読書中の視点情報を用いて、文書中の単語の内、学習者が難しいと感じた単語を推定する手法を提案した [20]。この研究では、5 名の視点情報を記録し、主観的高難易度単語の推定精度を検証している。検証の結果、全単語を対象とした場合、再現率が最大るとき、再現率 99.0%、適合率は 33.4%であった。また、固有名詞を学習および評価時に除いた場合は、再現率が最大るとき、再現率 99.0%、適合率は 39.6%であった。

視線情報をモバイル端末上で用いることを考えると、アイトラッカは構造的にモバイル端末への接着が難しく、利便性にも欠けるため、そのままモバイル端末上で利用することはできない。さらに、スマートフォンのフロントカメラを用いたアイトラッキング手法では精度が不十分である [21]。そのため、モバイル端末で未知単語推定を行うには、他の方法を開発する必要がある。

2.3 英文読解能力と読書時間の関係

学習者の英文読解能力を推定する他の指標として、読書時間がある。文章を読む時間から英文読解能力を推定する研究として、長沼らの研究 [7] がある。この研究では、大学 1・2 年生 239 名に対して 2000 年度に実施され

1) 大阪公立大学大学院 情報学研究科

表 1 推定に用いる特徴量

No	特徴量
1-2	単語数, 文字数
3	平均単語長
4	Automated Readability Index
5	Flesch Reading Ease
6-7	7 文字以上の単語の { 数, 占める割合 }
8-9	内容語の { 数, 占める割合 }
10-11	機能語の { 数, 占める割合 }
12	音節数
13-14	{ 単音節, 多音節 } の単語の数
15	単語の出現頻度
16-18	出現頻度の { 最小値, 最大値, 平均値 }
19	総読書時間
20-21	{ 1 単語, 1 文字 } あたりの読書時間
22-24	スマートフォンの { x, y, z } 軸加速度の最小値
25-27	スマートフォンの { x, y, z } 軸加速度の最大値
28-30	スマートフォンの { x, y, z } 軸加速度の平均値
31-33	スマートフォンの { x, y, z } 軸加速度の分散
34	読み返し回数

たベネッセの第 3 回「英語コミュニケーション能力テスト¹⁾」のリーディングスコアと読書時間の相関を検証した。その結果、文章の読みやすさの言語的要因である「難易度」「トピック」は読む速度に影響を与えると報告している。その内、難易度については、「語彙のレベル」「文構造の複雑さ」「文の長さ」などの基準がある。これにより、読書時間が学習者の文章読解能力を表す指標として有効であることを示した。

3 提案手法

3.1 特徴量取得

図 1(b) に、本研究で実装したスマートフォン用アプリケーションの、読書画面のスクリーンショットを示す。このアプリケーションは、スマートフォンの画面上で文中の数単語を 1 行ずつ表示する。これにより、各行をどのように読んでいるかを解析して、各単語が未知単語かどうかを識別できるようにしている。

最終的に得られる特徴量は、大きく分けてテキスト特徴量と読書行動データの 2 つである。表 1 に、未知単語の識別に用いる特徴量を示す。テキスト特徴量としては、単語数 (No.1)、文字数 (No.2)、平均単語長 (No.3)、2.1 節で説明した特徴量 (No.4-11)、単語の音節に関する特徴量 (No.12-14)、単語の出現頻度の統計値 (No.15-18) がある。読書行動データとしては、総読書時間 (No.19)、単語や文字ごとの読書時間 (No.20-21)、スマートフォンの 3 軸加速度 (2Hz) の統計値 (No.22-33)、読み返し回数 (No.34) がある。なお、冠詞などの stop words や固有名詞はこれらの特徴量を計算する前に取り除いている。

3.2 特徴量選択

推定に不要な特徴量が含まれている可能性があるため PR 曲線の AUC が高くなるように Sequential Backward Floating Selection を用いて推定精度の高い特徴量を選択する。

3.3 識別

選択された特徴量をもとにサポートベクターマシン (SVM) を用いて未知単語であるか否かを識別する。

1) <https://www.benesse.co.jp/gtec/>

カーネルは RBF を用いる。学習はユーザ依存で行う。

4 実験

実験はグループ P とグループ S の 2 回に分けて実施した。実験場所は、両グループともに読書の妨げにならない場所とした。このとき、読書中のスマートフォンの持ち方については指定していない。

まず、グループ P の実験参加者は日本人大学生、及び大学院生 22 名 (男性 11 名, 女性 11 名) であった。実験参加者の年齢は、平均 22.3 歳、標準偏差は 1.51 歳であった。実験の総時間は 4 時間であり、そのうち 3 時間を使って文書を読んでもらった。残りの 1 時間は実験説明と実験後のアンケートに使った。データを全て取得した実験参加者に対し、1 時間あたり 1,000 円、総額 4,000 円の謝礼を支払った。

次に、グループ S の実験参加者は日本人大学生 1 年生 (男性 6 人, 女性 2 人) であった。実験の総時間は 3 時間であり、そのうち 2 時間 30 分を使って文書を読んでもらった。残りの 30 分は実験説明と実験後のアンケートに使った。

4.1 テキスト特徴量と読書行動データの取得

図 1 に読書手順を示す。文書は Newsela データセット [22] を使用した。Newsela データセットは 11 段階の難易度に分かれている英語教材である。そのため、学習者は自身の英語能力を踏まえて文書の難易度を変更することができる。

まず、図 1(a) に示す文書選択画面より、実験参加者が興味のあるタイトルの文書を選択してもらった。これは、文書内容への関心が理解度や学習に結びつくことが報告されているためである [23, 24]。また、未知単語が適度に含まれた文書を読んでもらうため、未知単語が 15 個前後含まれると推定された文書を選択してもらった。

次に、文書を選択すると、図 1(b) に示す読書画面に選択した文書が表示される。この読書画面において 1 行ずつ読んでもらい、1 行ごとに 3.1 節で述べた特徴量を取得した。実験参加者には、読み返しを含め自由に読んでもらい、「意味を理解しながら一通り読むように」と指示した。

そして、文書を読み終えたあと、図 1(c) に示す画面で、意味を知らなかった単語を未知単語として選択してもらった。グループ P の実験参加者にはさらに、内容を理解しながら文書を読んでいることを確かめるため、文書の要約と難易度 (1: とても簡単~5: とても難しい) を口頭で答えてもらった。グループ S の実験参加者にはこの口頭での確認を実施しなかった。

これらの作業を繰り返し行い、グループ P では 9 文書分、グループ S では 5 文書分のデータを取得した。

実験終了後、実験参加者の性格特性を調べる Big Five テスト [25]、学習スタイル調査 (ILS) [26]、本実験に関するアンケートを行った。本実験に関するアンケートでは、英文を読むことは得意であるかについて調査した。

4.2 評価方法

ユーザ依存で交差検証を行い、PR 曲線の AUC である AUPR を評価指標として計算した。この AUPR は、学習者が読んだ全単語に対して、未知単語であるか否かの識別の際に求められる分類スコアをもとに計算した。得られた分類スコアをもとに閾値を変化させることで PR 曲線を求めることができる。全ての学習者に対して共通す

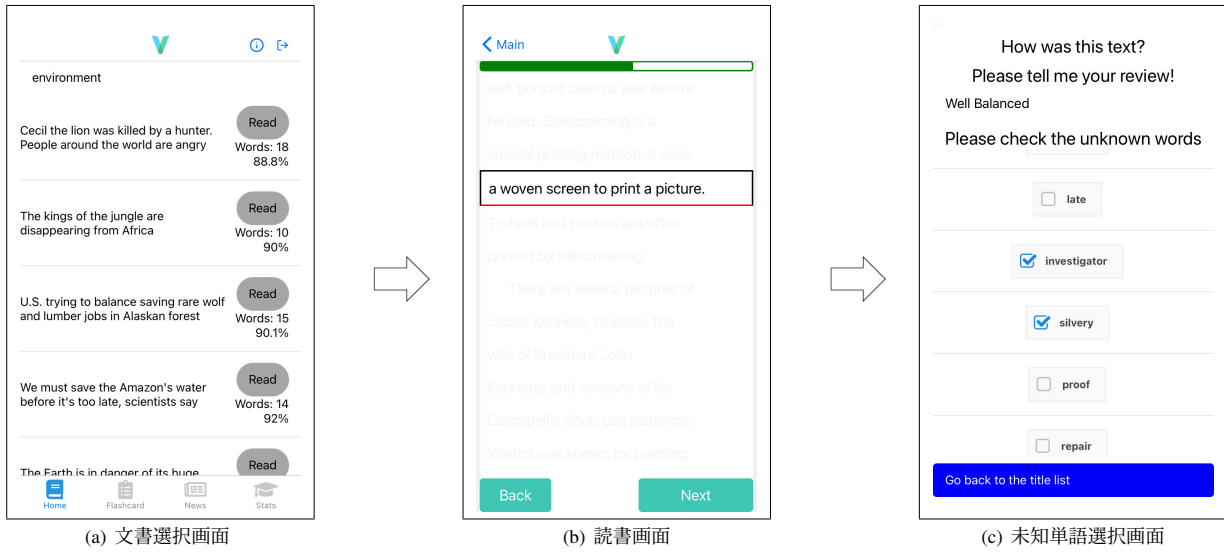


図 1 読書の流れ

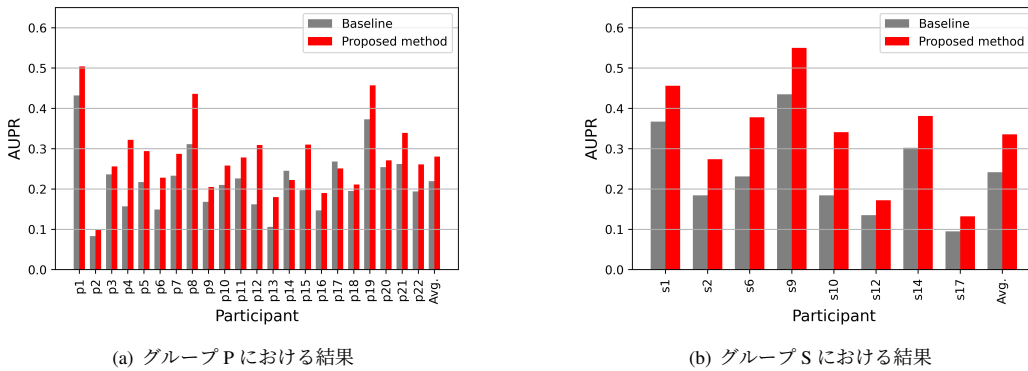


図 2 提案手法とベースラインの AUPR

客観的な指標であるテキスト特徴量のみで学習した SVM をベースラインとし、提案手法と比較した。SVM のカーネルには RBF を用いた。パラメータは $C = 1.0$, $\gamma = 1/34$ とした。 γ の分母の値は特徴数である。また、推定精度の検証には、読んでもらった文書のうち 1 文書をテストデータとし、それ以外の文書を学習データとする Leave-One-Document-Out 交差検証を用いた。

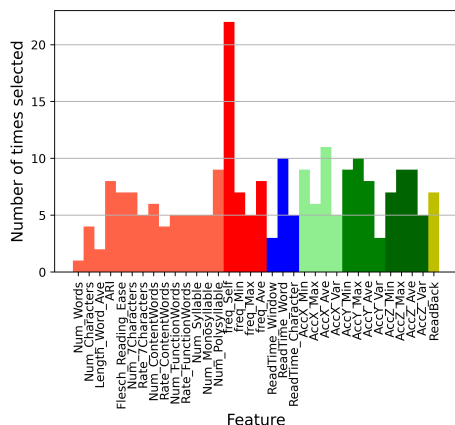
4.3 結果と考察

図 2 に、提案手法およびベースラインの AUPR を比較した結果を示す。図 2 より、実験参加者 30 名のうち 28 名に対して、提案手法がベースラインを上回る結果となった。提案手法とベースラインの間に有意差があるかどうかを 2 標本 t 検定 (データが正規分布であることを確認後) で判定した。その結果、グループ P では $p = 6.84 \times 10^{-6}$, グループ S では $p = 5.75 \times 10^{-4}$ より、統計的に有意な差があることが確認された。

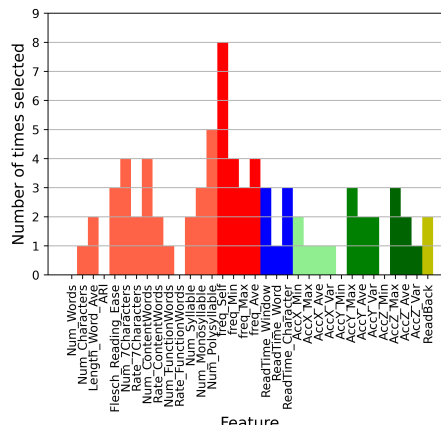
図 3 に、特徴量が選択された回数を示す。図 3 の横軸は推定に用いた特徴量を表し、縦軸はその特徴量が選択された回数を表す。図 3 より、全ての実験参加者において単語の出現頻度が選択された。このことから、単語の出現頻度は未知単語推定に最も寄与する特徴量であり、ARI のようなテキスト特徴量と比較して、多くの人に有効な特徴量であることを確認した。また、スマートフォンの 3 軸加速度 (2Hz) は、英文読解能力の指標となる読

書時間と同程度に未知単語推定に寄与していた。このことから、スマートフォンの持ち方が未知単語推定に寄与していることが示唆された。

どのような学習者に提案手法が有効であるかを実験後のアンケート等により調査した。各実験参加者における提案手法の AUPR からベースラインの AUPR を引いた値が 0.05 未満、0.05 以上で 2 つのグループに分ける。これら 2 つのグループはそれぞれ、提案手法が有効でなかった (AUPR の差が 0.00 未満の場合) または有効であるか否か不確実 (AUPR の差が 0.00-0.05 の場合) な実験参加者のグループ、提案手法が有効であった実験参加者のグループを表す。まず、Big Five テストと ILS の結果から調査したが、どのような学習者に提案手法が有効であるかを判別することはできなかった。次に、実験後のアンケートにより英文を読むことが得意か否かを調査した。実験後のアンケートから、AUPR の差が 0.05 未満のグループ 11 名の内 9 名が、英文を読むことが得意であるという回答が得られた。残りの 2 名のうち、1 名は「どちらとも言えない」という回答であった。一方で、AUPR の差が 0.05 以上のグループ 19 名の内 9 名が、英文を読むことが苦手であるという回答が得られた。残りの 10 名のうち、3 名は「どちらとも言えない」という回答であった。これより、文脈から未知単語の意味を推測する能力が高い学習者は、既知単語 (意味を知っている



(a) グループ P における結果



(b) グループ S における結果

図 3 推定に用いた特徴量の選択された回数

単語) と未知単語を読む際の振る舞いの差が小さく、有効な読書行動データが得られなかったと考えられる。

5 おわりに

本研究では、学習者に依存しないテキスト特徴量と学習者に依存した読書行動データの両方を用いて文中の未知単語を推定する新しい英語学習アプリケーションを開発した。その結果、提案手法はテキスト特徴量のみを用いるよりも、多くの学習者にとって未知単語の識別に有効であることが確認された。さらに、アイトラッカを使う必要がないため、モバイル端末での未知単語推定を可能としている。

今後の課題として、スマートフォンで取得できる読書行動データをさらに追加することが挙げられる。本研究では、スマートフォンのメモリ処理の関係で、スマートフォンの 3 軸加速度のサンプリング周波数は 2Hz であった。したがって、より高いサンプリング周波数の 3 軸加速度を利用できれば、精度の改善を期待できる。また、スクロールで読み進める機能を追加することでより英文を読みやすい UI を作成することも考えている。

謝辞

本研究の一部は JST CREST (JPMJCR16E1), JST Trilateral AI Research (JPMJCR20G3), JSPS 科研費基盤 (B) (20H04213), JSPS 国際共同研究強化 (B) (20KK0235), 阪大 Society5.0 グランドチャレンジの補助による。

参考文献

- [1] Zhenzhen Chen, Weichao Chen, Jiyou Jia, and Huili An. The effects of using mobile devices on language learning: a meta-analysis. *Educational Technology Research and Development*, Vol. 68, No. 4, pp. 1769–1789, 2020.
- [2] Helen Crompton and Diane Burke. The use of mobile learning in higher education: A systematic review. *Computers & Education*, Vol. 123, pp. 53–64, 2018.
- [3] Mostafa Al-Emran, Hatem M Elsherif, and Khaled Shaalan. Investigating attitudes towards the use of mobile learning in higher education. *Computers in Human Behavior*, Vol. 56, pp. 93–102, 2016.
- [4] Yao-Ting Sung, Kuo-En Chang, and Tzu-Chien Liu. The effects of integrating mobile devices with teaching and learning on students' learning performance: A metaanalysis and research synthesis. *Computers & Education*, Vol. 94, pp. 252–275, 2016.
- [5] James Coady and Thomas Huckin. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge University Press, 1997.
- [6] Elfrieda H. Hiebert, Judith A. Scott, Ruben Castaneda, and Alexandra Spichtig. An analysis of the features of words that influence vocabulary difficulty. Vol. 9, p. 8, 2019. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.
- [7] 長沼君主, 和田朋子. 速読力とテキストの読みやすさによる英語リーディング能力の測定. *日本語テスト学会研究紀要*, Vol. 5, pp. 34–45, 2002.
- [8] Keith Rayner, Sara C Sereno, Robin K Morris, A Rene Schmauder, and Charles Clifton Jr. Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, Vol. 4, No. 3-4, pp.

SI21–SI49, 1989.

- [9] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, Vol. 124, No. 3, p. 372, 1998.
- [10] Katsunori Kotani, Takehiko Yoshimi, Takeshi Kutsumi, Ichiko Sata, and Hitoshi Isahara. Efl learner reading time model for evaluating reading proficiency. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 655–664. Springer, 2008.
- [11] 東村理功, 岩田基, 黄瀬浩一. スマートフォンを用いた読書活動データからの未知単語推定. *電子情報通信学会総合大会*, 2022.
- [12] Riku Higashimura, Andrew Vargo, Motoi Iwata, and Koichi Kise. Helping Mobile Learners Know Unknown Words through their Reading Behavior. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022.
- [13] Steven A Stahl. Vocabulary and readability: How knowing word meanings affects comprehension. *Topics in language disorders*, Vol. 23, No. 3, pp. 241–247, 2003.
- [14] Rudolph Flesch. A new readability yardstick. *Journal of applied psychology*, Vol. 32, No. 3, p. 221, 1948.
- [15] James Dean Brown. An EFL readability index. *University of Hawai'i Working Papers in English as a Second Language*, Vol. 15, No. 2, pp. 85–119, 1997.
- [16] RJ Senter and Edgar A Smith. Automated readability index. Technical report, CINCINNATI UNIV OH, 1967.
- [17] Katsuya Fujii and Jun Rekimoto. Subme: An interactive subtitle system with english skill estimation using eye tracking. In *Proceedings of the 10th Augmented Human International Conference 2019*, pp. 1–9, 2019.
- [18] Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. Towards inferring language expertise using eye tracking. In *CHI' 13 Extended Abstracts on Human Factors in Computing Systems*, pp. 217–222, 2013.
- [19] Rui Hiraoka, Hiroki Tanaka, Sakriani Sakti, Graham Neubig, and Satoshi Nakamura. Personalized unknown word detection in non-native language reading using eye gaze. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 66–70, 2016.
- [20] Utpal Garain, Onkar Pandit, Olivier Augereau, Ayano Okoso, and Koichi Kise. Identification of reader specific difficult words by analyzing eye gaze and document content. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 1, pp. 1346–1351. IEEE, 2017.
- [21] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2176–2184, 2016.
- [22] Wei Xu, Chris Callison-Burch, and Courtney Napoles. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, Vol. 3, pp. 283–297, 2015.
- [23] Judith M. Harackiewicz, Jessi L. Smith, and Stacy J. Priniski. Interest matters: The importance of promoting interest in education. Vol. 3, pp. 220–227, 2016. Publisher: SAGE Publications.
- [24] Suzanne Hidi. Interest, reading, and learning: Theoretical and practical considerations. *Educational psychology review*, Vol. 13, No. 3, pp. 191–209, 2001.
- [25] Lewis R Goldberg. The development of markers for the big-five factor structure. *Psychological assessment*, Vol. 4, No. 1, p. 26, 1992.
- [26] Richard M Felder, Linda K Silverman, et al. Learning and teaching styles in engineering education. *Engineering education*, Vol. 78, No. 7, pp. 674–681, 1988.