L-VEIGe: 誤答画像生成による英語語彙学習支援システム
－誤答生成画像に対する定量評価モデルの提案－

L-VEIGe: Development of L2 Vocabulary Learning Support System with Suggestibility of Error by Image Generation
- A Proposal for a Quantitative Evaluation Model for Images Generated by Error-

杉田 一樹[*1], 谷 文[*2], 太田 光一[*2] 長谷川 忍[*2]
Kazuki SUGITA[*1], Wen GU[*2], Koichi OTA[*2], Shinobu HASEGAWA[*2]
[*1] 北陸先端科学技術大学院大学 先端科学技術研究科
[*1]Graduate School of Advanced Science and Technology, JAIST
[*2] 北陸先端科学技術大学院大学 遠隔教育研究イノベーションセンター
[*2]Center for Innovative Distance Education and Research, JAIST
Email: [*1]s2320028@jaist.ac.jp, [*2]{wgu, ota, hasegawa}@jaist.ac.jp

あらまし：誤答画像生成による英語語彙学習支援システム L-VEIGe (Learning-Vocabulary Error Image Generation)では，学習者の繰り返しの誤答を防ぐ上で有効であることを示した．一方，誤答画像に於いて学習者が誤答を理解するに足る情報が欠損する，特徴消失問題が存在する．本研究では，誤答画像の定量評価を目的と定め，誤りの可視化に於いて重要とされる認知的忠実性について，想起性，印象性と定義し，その定量評価モデルの構築を行った．
キーワード：英語語彙学習支援システム，第二言語習得，誤答画像の定量評価，

1. はじめに

英語語彙学習では，コンテクストから語彙を学習する方法が認知されており，画像に含まれる情報をコンテクストとみなす学習方法が有効とされている(1)．一方，学習者の繰り返しの誤りが，自然言語として定着する Fossilization の問題が存在する(2)．しかし，英語語彙学習に於いて，学習者の誤りに注目したものは少ない．誤りからの学習を英語語彙学習に適用した，誤答画像生成を活用した語彙学習支援システム L-VEIGe では，学習者の効果的な誤りへの内省を実現させ，繰り返しの誤りを防ぐうえで有効であることを示した(3)．一方，L-VEIGe では，学習者が誤答を理解するに足る情報の欠損である，特徴消失問題が存在する．本研究では，特徴消失問題への対処として，誤答生成画像に対する定量評価方法の策定を目的と定める．

2. 誤答生成画像に対する定量評価

2.1 誤答生成画像に対する評価軸

学習者の誤りに注目した研究として，EBS（Error-Based Simulation）が存在する(4)．EBS では，学習者自身が誤りを発見し，秩序の乱れを解消すべきものと捉えた時に生起する，認知的葛藤の必要性を説いており，認知的葛藤を生起させる上で，可視化表現に認知的忠実性を有する必要があるとされている(5)．生成画像が有するべき認知的忠実性を本研究では，誤答画像が学習者の誤りの理解に足る情報を有する尺度としての「想起性」，誤答画像が学習者に誤りへの気づきを促す尺度としての「印象性」を定義した．本稿では，それら二つの軸である評価値を定量的に評価するモデルを提案する．

2.2 想起性評価モデル

想起性評価モデルは，先行研究の L-VEIGe：誤答画像生成を活用した語彙学習支援システム(3)の拡張を行う．学習者が誤答を理解するに足る情報として定義した想起性をさらに，文章全体の理解に必要な，(1)コンテクストの類似度，誤答そのものの理解に必要な，(2)誤り概念の類似度と定義する．これら二つの値を評価するモデルを構築する．画像から生成されたキャプションに対する定量評価方法として先行研究では CLIP スコアが存在する(6)．CLIP スコアでは，文全体と画像に対してそれぞれベクトル化し，その類似度を評価するためのモデルであるため，生成画像と文の全体の類似度の評価が可能であるが，(2)のような特定の誤答に関する特徴を評価することが困難である．本研究では，こうした問題を解決するため，上記に定めた 2 つの評価値の評価モデルとして，a. イメージキャプショニングによる評価モデル，b. VQA による評価モデルの二つのモデルによるアンサンブルモデルを構築した．a. イメージキャプショニングによる評価モデルは先行研究(3)のものを利用し，b. VQA による評価モデルでは，画像とプロンプトを入力とし，評価値としての出力を得るためのモデルとして，GPT-4V(7)を適用した．それぞれ得られた評価値をもとに後述するモデルの学習を行う（図１）．

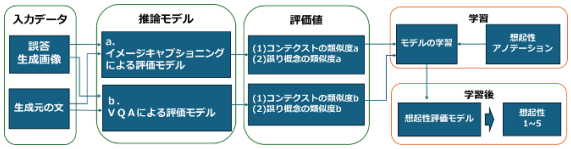

図１．想起性評価モデル

2.3 印象性評価モデル

誤答画像が学習者に誤りへの気づきを促すに足る情報の尺度としての印象性をさらに，Ⅰ．誤り概念表層比率，Ⅱ．誤り概念色差と定義する．Ⅰについては，学習者の誤りへの認知に於いて，画像に含まれる誤答概念の比率が影響するという仮定に基づき定義し，Ⅱについては，画像の誤答箇所と周囲の色差が記憶定着に寄与する前提に基づき定義した．これら二つの値を評価するモデルを構築した．Ⅰ．誤り概念表層比率では，画像とプロンプトを入力とし，プロンプトに該当する画像の特定の箇所を抽出するモデルである，CLIPSeg[8]を適用した．学習者の誤答概念を画像内から抽出し，ピクセル単位での画像内の表層比率をもとに，誤り概念表層比率を導出する．Ⅱ．誤り概念色差については，生成画像に対して特定の箇所の色彩変化を施した際に用いた彩度の変化度を値として用いる．彩度の変化度は，python ライブラリである，Pillow の Image Enhance.Color メソッドの factor の値を用いる．本稿では画像生成プロセスについては焦点を当てないため，画像生成プロセスの詳細について割愛する．得られた 2 つの特徴量をもとに印象性評価モデルの学習を行う（図２）．

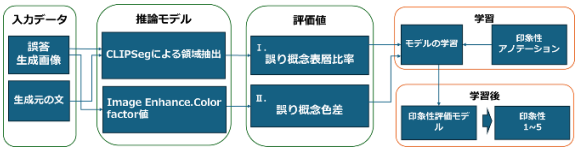

図２．印象性評価モデル

3. モデルの学習・評価

想起性評価モデルと，印象性評価モデルについて，アノテーターによるラベル付けを行い，モデルの学習を行う．ここでは，L-VEIGe の改良として，想起性，印象性，それぞれに対するアノテーション環境を構築した．想起性アノテーション環境では，学習者の選択に応じた生成画像に対する評価モデルによる出力値をもとに，「条件１__高，条件２__高」，「条件１__高，条件２__低」，「条件１__低，条件２__高」，「条件１__低，条件２__低」の４つの条件の画像を提示し，それぞれに「生成画像から誤答文を想起できる」という観点で５段階評価によるアノテーションを行う．この時，条件１はコンテクストの類似度，条件２は誤り概念の類似度をさす．図３は想起性の各条件に合致する例である．実際のアノテーションでは，生成画像の評価の観点から，条件間の差は出題により異なり，図は条件を明示的に示すための極端な例である．印象性アノテーション環境も同様，４つの条件の画像を提示し，それぞれに，「生成画像が誤りへの気づきを促した」という観点で５段階評価によるアノテーションを行う．この時，条件１は誤り概念表層比率をさし，条件２は誤り概念色差をさす．図４は印象性の各条件に合致する画像の例である．これらの値をもとに推論モデルの学習を行い，MSE をもとに評価を行う．モデルの評価値，アノテーション環境については，当日詳細を報告する．

図３．想起性アノテーション用画像例

図４．印象性アノテーション用画像例

4. まとめ

本研究では，誤答画像生成による語彙学習支援システム「L-VEIGe」の誤答生成画像が有する特徴消失問題を課題提起し，生成画像に対する定量評価モデルについて提案した．定量的な評価が実現するとき，適切な誤答生成画像の選定が実現する．今後の展望としては，本研究で定義した誤答画像が有する認知的忠実性としての，想起性，印象性に対する学習への効果の評価を行い，誤答生成画像に対する定量評価値の信頼性評価を行う．

謝辞

本研究は，JST 次世代研究者挑戦的研究プログラム JPMJSP2102 の支援を受けたものです．

参考文献

(1) Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. "Supporting visual and verbal learning preferences in a second-language multimedia learning environment." Journal of Educational Psychology, 90(1), 25-36. (1998)
(2) Vigil, N. and Oller, J. Rule fossilization: A tentative model. Language Learning, 26:281–295. (2006)
(3) 杉田 一樹，太田 光一，谷 文，長谷川 忍 : "L-VEIGe: 誤答画像生成を活用した語彙学習支援システム -画像の示唆性に対する指標の検討-"，教育システム情報学会研究報告 Vol.38(2)，pp.118-124, (2023)
(4) 平嶋宗，堀口知也: "「誤りからの学習」を指向した誤り可視化の試み"，教育システム情報学会誌，Vol.21, No.3, pp. 178-186 (2004)
(5) 平嶋 宗，"メタ認知を活性化する学習支援システム"，人工知能学会全国大会論文集, JSAI03 巻, 17 回 (2003)
(6) Hessel, Jack, Holtzman, Ari, Forbes, Maxwell, Bras, Ronan Le and Choi, Yejin. "CLIPScore: A Reference-free Evaluation Metric for Image Captioning.." Paper presented at the meeting of the EMNLP (1), (2021)
(7) Z. Yang et al., “The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision).” arXiv, Oct. 11, 2023. doi: 10.48550/arXiv.2309.17421.
(8) T. Lüddecke and A. S. Ecker, "Image Segmentation Using Text andImage Prompts," in Conf on Computer Vision and PatternRecognition, New Orleans, (2022)