

令和 6 年 5 月 10 日現在

機関番号：12608

研究種目：挑戦的研究（萌芽）

研究期間：2021～2023

課題番号：21K18358

研究課題名（和文）コンピュータとの対話による言語学習者の生産語彙能力評価の自動化

研究課題名（英文）Production vocabulary assessment through dialogue systems

研究代表者

徳永 健伸（Tokunaga, Takenobu）

東京工業大学・情報理工学院・教授

研究者番号：20197875

交付決定額（研究期間全体）：（直接経費） 5,000,000円

研究成果の概要（和文）：対話におけるアライメントは、対話を円滑に進めると考えられている。我々は、単語が同調する語彙アライメントに加え、使用される語彙の難易度も同調すると考えてこれを語彙レベルアライメントと名付けた。既存の日本語対話コーパスを分析することにより、この現象が自然な対話において発生することを明らかにし、これが言語学習者の生産語彙評価に利用できることを示した。そして、その実現のために語彙レベルと現在の対話のトピックを考慮して使用するべき単語リストを提案するモジュールを提案し、この単語リストを大規模言語モデルのプロンプトに組み込むことによって、LLMが生成する発話の語彙レベルを制御する手法を提案した。

研究成果の学術的意義や社会的意義

既存の語彙テストは受容語彙を評価するものが多く、生産語彙を評価するテストはほとんどない。日本の英語教育では読む/聞く能力が偏重され、書く/話す能力がないがしろにされているという最近の批判の中でも指摘されているように、言語の生産能力はコミュニケーション能力、特に情報発信力にとって重要な要素であり、生産語彙はその中核をなす言語知識である。本研究では、学習者がコンピュータとの対話をおこなう過程で自発的に発する発話を分析し、学習者の生産語彙能力を客観的に評価するシステムを実現する。これは、言語教育の現場において教師の負担を軽減し、適切な評価によって学習者の学習効率を改善する効果が期待できる。

研究成果の概要（英文）：Lexical alignment is believed to facilitate smooth dialogue. We believe alignment also happens at the vocabulary difficulty used in the utterance, which we name "lexical level alignment". By analysing an existing corpus of Japanese dialogue, we have shown that this phenomenon occurs in natural dialogue and that it can be used to assess language learners' productive vocabulary. To realise this, we proposed a module that proposes a list of words to be used in the utterances considering the lexical level and the topic of the current dialogue. Incorporating this word list into the prompts for LLMs enables us to control the lexical level of the LLM's utterances.

研究分野：自然言語処理

キーワード：対話システム 語彙レベル アライメント 言語学習 語彙テスト 生産語彙

1 研究開始当初の背景

言語学習者の生産語彙能力を評定する先行研究は、文の欠損部分を埋めさせる手法 (Laufer and Nation 1999), 母語で書かれた文を翻訳をさせる手法 (Koizumi 2003), 作文を書かせる手法 (Laufer and Nation 1995), 刺激語を与え連想語を列挙させる手法 (Meara and Fitzpatrick 2000, González and Piriz 2016) などに分類できる。しかしながら、これらの手法には、テストする対象語を柔軟に設定できなかつたり、学習者の語の使用が必ずしも自発的でないという限界がある。言語テストで使われるスピーキングテストは学習者の生産語彙能力を評定するよい手法であるが、その評価は試験者の主観による全体論的な評定となり、生産語彙に特化したミクロな評定は難しい。全体論的な評定と同時に語彙や文法などの個々の能力別のミクロな評定ができることが学習者へのフィードバックの観点からは望ましい。このように学習者の生産語彙能力を高い精度で客観的に評定する手法については十分な研究がおこなわれていない。

言語教育の観点からは、言語教育で主体性が育まれるべき「自主的発話」の自動評価方法を提案することにより、言語教育の評価方法の改善に貢献すると同時に、学習者が自分の好きな領域 (趣味・専門・日常) で主体的に話すことが、そのまま自分の言語能力の評価に反映されることにより学習者の学ぶ意欲にも貢献する。また、言語処理の観点からは、既存の言語資源や言語処理技術を言語教育に応用するアプローチから言語教育を目的とした言語資源や処理技術を作り出すアプローチへ考え方を転換する機会を与える。言語教育を言語処理の一応用分野として考えるのではなく、両者が対等な立場で研究を進める新しい研究分野を創出することが求められている。

研究代表者の徳永は言語処理の教育応用に関する研究 (Gecchele et al. 2019, Putra et al. 2020) で多くの成果をあげている。特に、挑戦の萌芽研究 (15K12166; 2015-2016) ではインターネット上のテキストから英語の受容語彙テストを自動作成する研究を実施し、国際論文誌に 4 報、国際会議に 3 報の論文を出版している (Susanti et al. 2017, 2018, 2020)。この研究を通して受容語彙能力の評定に比べて生産語彙能力の評定をおこなう研究が非常に少ないことを認識し、生産語彙能力の評定を研究課題とするに至った。また、徳永は対話システムの研究 (Fukunaga et al. 2018, Ogawa et al. 2020, Tellols et al. 2020) でも多くの成果をあげており、対話システムの技術を用いて、対話のトピックを系統的に制御して学習者の自発的な発話を引きだし、その発話を分析することによって学習者の生産語彙能力の自動評定をおこなうという着想を得た。

2 研究の目的

本研究は言語学習者の生産語彙能力をコンピュータとの対話を通して自動的に評定する手法を提案し、そのプロトタイプ・システムを実装して手法の有用性を実証することを目的とする。応用言語学では、語彙を言語表現を理解するために使う **受容語彙** と言語表現を生産するために使う **生産語彙** に区別している。母語でも読む/聞くなら理解できるが、自ら書く/話す際には使わない語があることから、受容語彙と生産語彙の区別には説得力がある。一般に受容語彙は生産語彙よりもサイズが大きいと言われている。

既存の言語テストにおける語彙問題は受容語彙の能力を評定するもの、つまり語の意味が理解できるかどうかを問うものが多く、語が使えるかどうかを評定する問題はほとんどない。一部の穴埋め問題は、文を完成するために語を選択させるという意味では生産語彙を評定しているとも言えるが、前後の文脈が言語的な制約やヒントを与えることになり、自発的に語を想起し使用できる生産語彙能力を評定しているとは言い難い。

日本の英語教育では読む/聞く能力が偏重され、書く/話す能力がないがしろにされているという最近の批判の中でも指摘されているように、言語の生産能力はコミュニケーション能力、特に情報発信力にとって重要な要素であり、生産語彙はその中核をなす言語知識である。学習者の生産語彙能力を的確に評定し、伸ばすことは学習者の情報発信力を向上させることにも資する。

このような背景から、本研究では、学習者がコンピュータとの対話をおこなう過程で自発的に発する発話を分析し、学習者の生産語彙能力を客観的に評価するシステムを実現する。そのための要素となる手法・技術は言語処理ならびに言語教育の学術的な観点からも重要であるが、言語教育の現場において教師の負担を軽減し、適切な評価によって学習者の学習効率を改善する効果が期待できる。

3 研究の方法

本研究では、言語学習者が対話システムと対話をする中で使う語彙のレベルをシステム側が制御し、学習者が使う語の語彙レベルをシステムが徐々に上げさせていった時に、どの語彙レベルまでの語が使えるかを見ることによって学習者の語彙レベルを測定することを狙っている。そのためには、二つの課題を克服する必要がある。ひとつは学習者が使う語の語彙レベルをシステムが制御すること、もうひとつはシステム自身が使う語の語彙レベルを制御できることである。

まず、学習者が使う語の語彙レベルをシステムが制御することに関して、我々は従来から心理言語学分野で知られている語彙アライメントの現象に着目した。語彙アライメントは対話者が使う語や表現が対話の進行にともなって収斂していくという現象である。我々はこのような現象が語彙レベルについても生じるのではないかと考えた。たとえば、子供や非母語話者と話す場合、自分の発話を相手が理解できるように、難しい語を使うことは避けるのが自然である。我々はこの現象を語彙レベルアライメント (Lexical Level Alignment: LLA) と名付け、既存の対話コーパスを分析することによって特定の条件で語彙レベルアライメントが生じることを明かにした。

もうひとつの課題、システムの発話に使用する語の語彙レベルの制御に関して、現在、盛んに研究されている大規模言語モデル (Large Language Model: LLM) を実装基盤として用いた。まず、既存の LLM 単体 (GPT3.5-turbo, GPT4) で語彙レベルの区別がどれくらい可能かを調査した結果、細かな語彙レベルの制御が困難であることが明らかになったので、指定された語彙レベルの単語リストを現在の対話のトピックを考慮しながら辞書から選択するアルゴリズムを開発し、このモジュールの出力を LLM へのプロンプトに組み込むことによって、発話生成における細かな語彙レベルの制御をおこなう手法を検討した。

4 研究成果

4.1 語彙レベルアライメントの分析

我々が提案する語彙レベルアライメントが実際の対話で生じることを確認するため、BTSJ 1000 人日本語自然会話コーパス (USAMI 2023) を用いた分析をおこなった。BTSJ は、母語話者、言語学習者などさまざまな背景を持つ対話者が雑談やインタビューなど様々な種類の対話をおこなった音声とその書き起しを集積したコーパスである。全部で 514 対話が収録されており、各対話の平均ターン数は 290 で、比較的長い対話となっている。この研究では雑談対話 396 対話を分析対象として選んだ。これらは対話者が知人か初対面か、母語話者か日本語学習者かによって表 1 のように区別できる。

表 1: BTSJ の対話区分

	母語話者-母語話者	母語話者-学習者
初対面同士	125	59
知人同士	141	43

語彙レベルの定義にはさまざまな方法があるが、本研究では国立国語研究所によって作成された日本語の代表的なシソーラスである分類語彙表 (National Institute for Japanese Language and Linguistics 2004) に浅原らがクラウドソーシングによって付与した親密度 (Asahara 2019) を語の語彙レベルとして使用する。また、発話の語彙レベルをその発話中に含まれる語を語彙レベル順に整列し、難しい方から Q% の位置にある語の語彙レベルとして定義した。これは発話中の難しい Q% の語が理解できなくても発話全体は理解できると仮定した定義になっている。

以上の準備のもと、対話の前半に含まれる対話参加者間の発話の語彙レベルの差と後半の語彙レベルの差の変化を調査した結果、表 2 のような結果となった。

表から初対面の母語話者と学習者の間では対話の前半と後半で発話の語彙レベルの差が縮まっていることがわかる。すなわち、語彙レベルアライメントが生じている。母語話者同士の場合は語彙レベルは飽和状態にあるかそもそも非常に高いレベルにあるので知人か初対面かを問わず、対話の前後で差がでにくいと考えられる。また、母語話者と学習者が知人同士の場合は、最初から相手の語彙レベルを知っていると思われるので、やはり対話の前後で差がでにくいと考えられる。我々の目的は言語学習者の語彙レベル評定であり、その語彙レベルは未知であるから、我々の想定する状況では語彙

表 2: 対話の前半と後半の語彙レベルの差の変化 (平均値と括弧内は p 値)

対話区分	対話数	$Q = 25$	$Q = 50$
母語話者-学習者, 初対面同士	59	-.040* (.026)	.003 (.569)
母語話者-学習者, 初対面同士	43	.005 (.601)	-.002 (.441)
母語話者-母語話者, 知人同士	125	-.013 (.103)	-.005 (.247)
母語話者-母語話者, 知人同士	141	.020 (.932)	.015 (.974)

レベルアライメントは生じると想定できる。この成果は the 27th Workshop on the Semantics and Pragmatics of Dialogue (MariLogue) において発表した (Tseng et al. 2023)。

4.2 既存 LLM の語彙レベル知識の評価

既存の LLM が語彙レベルに関する知識を有するかどうかを調査するために、分類語彙表の単語親密度に基づいて語彙レベルの異なる語対を LLM に提示し、どちらが簡単な語かを答えさせる実験をおこなった。まず、分類語彙表で定義されている 519 の意味カテゴリについて意味カテゴリ内の単語のすべての組合せで単語対を作り、単語間の語彙レベルの差で 0.2 刻みで 19 のグループに分類した。各グループから最大 10 の単語対を選び、合計で 197 単語対を実験に用いた。これらの単語対を以下のプロンプトを使って GPT 3.5 (gpt-3.5-turbo-0613-16k) と GPT 4 (gpt-4-0613) に入力した。

(System) 対話中に日本語学習者にとって分かりやすい方の単語を教えてください。単語のラベル (“A” または “B”) のみ出力すること。文脈がなくても必ず “A” か “B” を選択すること。

(User) A: 仕事をする
B: 鞆掌する

(Assistant) A

(User) A: 半作
B: ありがとう

(Assistant) B

実験の結果、GPT 4 は GPT 3.5 よりも簡単な語を正しく当てることができるが、語彙レベルの差が 0.4 ~ 0.6 あたりでは十分な性能とはいえないことがわかった。

4.3 対話システムの発話の語彙レベルの制御

LLM 単体で語彙レベルの制御をおこなうことは困難だと判断したので、指定された語彙レベルの単語リストをあらかじめ用意し、それらの単語をできるだけ使うようにプロンプトで指示することによってシステムの発話の語彙レベルを制御する手法を提案した。この際、もうひとつ考慮すべき点として対話の一貫性を保つために、現在の対話トピックに関連した単語に限定して単語リストを構成することである。これらの要件を満たすために、提案手法ではまず、(1) LLM を用いて、対話トピックを代表するキーワードを発話履歴から抽出し、(2) 分類語彙表の分類情報を参照して、対話トピックを代表するキーワードと同じ分類の単語を抽出し、その中で指定した語彙レベルより簡単な単語を選択して発話生成用単語リストを作成する。(3) 最後に作成した発話生成用単語リストと発話履歴を含めたプロンプトを LLM に提示し、それらの単語をできる限り使うように応答の発話を生成させる。

BTSJ コーパスの 368 雑談対話のうち、40 ターン未満の 8 対話を除外し、残りの 360 対話において各対話の前半の最後の 20 発話を抽出して評価データとした。この 20 発話において、次の発話が対話システムのアシスタントのターンになるように、一番目の話者をアシスタント、二番目の話者をユーザと仮定する。従って、20 発話の最後のユーザの発話に対してアシスタントが応答を生成する。評価実験では指定語彙レベルを 0.6, 0.8, 1.0, 1.2, 1.4 に設定し、発話生成用単語リストのサイズを 25, 50, 100, 200, 400, 800, 1600 に設定した。評価実験の結果、発話生成用の単語リストのサイズを 25 に設定するのが一番効果的で、この時に LLM の生成する発話を指定した語彙レベルに制御できることがわかった。この成果は言語処理学会第 30 回年次大会 (NLP2024) で発表した (Tseng et al. 2024)。また、2024 年 6 月 12-14 日に開催される 7th ESTIDIA Conference Exploring Real-life, Fictional and Virtual Dialogue: Similarities, Differences and Complementarities でも発表予定である。

参考文献

- Masayuki Asahara. Word familiarity rate estimation using a Bayesian linear mixed model. In *Proceedings of the First Workshop on Aggregating and Analysing Crowdsourced Annotations for NLP*, pages 6–14, Hong Kong, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-5902. URL <https://aclanthology.org/D19-5902>.
- Shunya Fukunaga, Hitoshi Nishikawa, Takenobu Tokunaga, Hikaru Yokono, and Tetsuro Takahashi. Interpretation of implicit conditions in database search dialogues. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, pages 477–486, 2018.
- Marcello Gecchele, Hiroaki Yamada, Takenobu Tokunaga, and Yasuyo Sawaki. Supporting content evaluation of student summaries by idea unit embedding. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2019)*, pages 343–348, 2019.
- Rafael Alejo González and Ana Maria Piquer Piriz. Measuring the productive vocabulary of secondary school CLIL students: Is Lex30 a valid test for low-level school learners? *Vigo International Journal of Applied Linguistics*, 13:31–54, 2016.
- Rie Koizumi. A productive vocabulary knowledge test for novice Japanese learners of English: Validity and its scoring methods. *JABAE/T (Japan-Britain Association for English Teaching Journal)*, 7:23–52, 2003.
- Batia Laufer and Paul Nation. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3):307–322, 1995.
- Batia Laufer and Paul Nation. A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1):33–51, 1999.
- Paul Meara and Tess Fitzpatrick. Lex30: an improved method of assessing productive vocabulary in an L2. *System*, 28(1):19–30, 2000.
- National Institute for Japanese Language and Linguistics. *Bunrui goihyo zôho kaitei-ban* (Word List by Semantic Principles Revised and Enlarged Edition). Dainippon Tosyo, Tokyo, 2004.
- Haruna Ogawa, Hitoshi Nishikawa, Takenobu Tokunaga, and Hikaru Yokono. Gamification platform for collecting task-oriented dialogue data. In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020)*, page 7084-7093, 2020.
- Jan Wira Gotama Putra, Simone Teufel, Kana Matsumura, and Takenobu Tokunaga. TIARA: A tool for annotating discourse relations and sentence reordering. In *Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 6912–6920, 2020.
- Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, and Hiroyuki Obari. Controlling item difficulty for automatic vocabulary question generation. *Research and Practice in Technology Enhanced Learning*, 12(25):1–16, 2017.
- Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, and Hiroyuki Obari. Automatic distractor generation for multiple-choice english vocabulary questions. *Research and Practice in Technology Enhanced Learning*, 13(15):1–16, 2018.
- Yuni Susanti, Takenobu Tokunaga, and Hitoshi Nishikawa. Integrating automatic question generation with computerized adaptive test. *Research and Practice in Technology Enhanced Learning*, 15(9):1–22, 2020.
- DolÇa Tellols, Takenobu Tokunaga, and Hilofumi Yamamoto. Assessing language learners’ free productive vocabulary with hidden-task-oriented dialogue systems. In *Proceedings of Workshop on User-Aware Conversational Agents (user2agent 2020)*, pages 1–6, 2020.
- YiKai Tseng, Takenobu Tokunaga, and Hikaru Yokono. Lexical level alignment in dialogue. In *Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue (MariLogue)*, pages 1–10, 2023.
- YiKai Tseng, 徳永健伸, and 横野光. 大規模言語モデルを用いた対話システムの語彙レベル制御. In *言語処理学会第 30 回年次大会 (NLP2024) 発表論文集*, pages 2732–2737 (B10–3), 2024.
- Mayumi (ed.) USAMI. Building of a Japanese 1000 person natural conversation corpus for pragmatic analyses and its multilateral studies, and ninjal institute-based projects: Multiple approaches to analyzing the communication of Japanese language learners., 2023.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Tellois Dolca, Tokunaga Takenobu, Yamamoto Hilofumi, Yokono Hikaru	4. 巻 38
2. 論文標題 Pic2PLex: A Test for Assessing Free Productive Vocabulary	5. 発行年 2023年
3. 雑誌名 Transactions of the Japanese Society for Artificial Intelligence	6. 最初と最後の頁 B~M52_1-11
掲載論文のDOI（デジタルオブジェクト識別子） 10.1527/tjsai.38-1_B-M52	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 Dolca Tellois, Takenobu Tokunaga, Hikaru Yokono
2. 発表標題 Vocabulary Volume: A New Metric for Assessing Vocabulary Knowledge
3. 学会等名 Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022)（国際学会）
4. 発表年 2022年

1. 発表者名 YiKai Tseng, Takenobu Tokunaga, Hikaru Yokono,
2. 発表標題 Lexical level alignment in dialogue
3. 学会等名 The 27th Workshop on the Semantics and Pragmatics of Dialogue (MariLogue)（国際学会）
4. 発表年 2023年

1. 発表者名 YiKai Tseng, 徳永健伸, 横野光
2. 発表標題 大規模言語モデルを用いた対話システムの語彙レベル制御
3. 学会等名 言語処理学会第30回年次大会(NLP2024)
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	山元 啓史 (Yamamoto Hilofumi) (30241756)	東京工業大学・リベラルアーツ研究教育院・教授 (12608)	
研究 分担者	横野 光 (Yokono Hikaru) (60535863)	明星大学・情報学部・准教授 (32685)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------