

令和 5 年 6 月 14 日現在

機関番号：32638

研究種目：若手研究

研究期間：2018～2022

課題番号：18K12434

研究課題名（和文）語彙学習における連想ネットワークの動的な連携と可視化

研究課題名（英文）Dynamic Linkage and Visualization of Associative Networks in Vocabulary Learning

研究代表者

寺岡 丈博（Teraoka, Takehiro）

拓殖大学・工学部・准教授

研究者番号：30617329

交付決定額（研究期間全体）：（直接経費） 2,500,000円

研究成果の概要（和文）：本研究では、日本語学習者が日本語母語話者の言葉の連想に沿って語彙を学習できることを目指して連想語彙学習システムを開発し、その評価を行った。自動で生成した問題文と回答選択肢に対して日本語教師にレーティングしてもらい、約8割が学習支援の教材として適していることがわかった。また、日本語学習者が本システムを一定期間使用して得られた学習結果を分析したところ、連想の可視化による学習の効果を確認することができた。

研究成果の学術的意義や社会的意義

海外における日本語学習の需要が高い中、自然言語処理技術が日本語学習の支援ツールや教材メディアとして応用されてきている。しかし、他の主要言語と比べると言語・教育研究の成果に基づいた辞書・コーパスや教材の開発は遅れていると考えられている。今日の日本語学習に関する辞書やコーパス、学習支援システムでは、単語の共起情報を用いたものが多く見られる。

本研究では、従来の共起情報ではなく人間が単語について連想した情報（連想情報）を語彙学習に用いている。研究成果の「学習支援の教材」と「学習の効果」に関する評価を通して、学習に連想情報がある程度有効であることを示すことができた。

研究成果の概要（英文）：In this study, we developed a support system for Japanese language learners to learn vocabulary according to association of words by Japanese native speakers, and evaluated it. The generated question sentences and answer choices were rated by Japanese language teachers, and about 80% of them were found to be suitable as learning support materials. We then asked Japanese language learners to use our system for a certain period of time and analyzed their learning data. As a result, we confirmed the effectiveness of learning by visualization of word associations.

研究分野：自然言語処理

キーワード：語彙学習 学習支援 連想情報 可視化

1. 研究開始当初の背景

海外の日本語学習者数は、2012年には約399万人まで増加していたが、2015年は約365万人まで減少している。主な理由としては、インドネシアや韓国の「中等教育における第二外国語の学習者減少」や中国の「就職を意識した日本語離れ」などが挙げられる。一方で、オーストラリアやタイ、ベトナム、フィリピンでは学習者数が大幅に増加しており、中国とインドネシアも含めて日本語教育機関数が増加していることから、日本語学習の需要は依然として高い。

このような動向の中、自然言語処理技術が日本語学習者用の支援ツールや教材メディアに応用されてきており、「辞書・コーパス」と「学習支援システム」に大別することができる。しかし、他の主要言語と比べると言語研究・教育研究の成果に基づいた辞書やコーパス、教材の開発は遅れていると考えられている。

従来の日本語学習に関するコーパスや支援システムでは、単語の共起情報を用いたものが多く見られる。日本語母語話者(データによっては日本語学習者)によって実際に話された言葉や書かれた文章における共起の頻度がまとめられており、話し言葉や書き言葉としてのコロケーションとみなすことができる。このような用例データは、学習者の言語習得状況を調べる上で有用であるが、一方で学習者の言語習得自体に最大の効果をもたらしているかは議論の余地がある。

2. 研究の目的

本研究の目的は、従来使用されてきた共起情報ではなく、動詞連想概念辞書の連想情報を語彙学習に応用することである。動詞概念辞書は、実験参加者延べ約3,200人分の連想実験データからまとめられており、動詞と深層格に対する連想語の頻度や順位の情報などを含んでいる。研究開始当初までに、この辞書を構成する連想情報と、新聞コーパス等から得られる共起情報について性質と応用の両面で比較分析を行い、それぞれの特徴や性質が異なることを確認している。

本研究では、研究開始までに試作した動詞連想概念辞書の連想情報を用いた日本語語彙学習システム(連想語彙学習システム)について開発を進め、完成させることを目指す。次に、名詞に対する連想情報から構築されている名詞連想概念辞書も本システムで動詞連想概念辞書と併用させ、名詞由来と動詞由来の連想ネットワークを動的に連携させる。そして最後に、日本語学習者が語彙のより深い理解ができるよう連想の可視化に取り組む。

以上のようにして、日本語学習者が日本語母語話者の単語(動詞・名詞)に対する連想に沿って語彙を学習できるようになることを図る。

3. 研究の方法

(1) 連想語彙学習システムの開発

試作段階から機能を拡張し、「学習」と「復習」のフェーズを完成させる。前者は、学習者が開始時に選択した「既知の動詞」と選択しなかった「未知の動詞」に対し、問題セット(問題文・回答選択肢)を自動生成し、学習者に呈示する。後者の「復習」は、学習者が「学習」で回答した問題に対して正解するまでランダムで再呈示する。これらの機能により、学習者が事前にチェックした「既知/未知の動詞」に対して正解するまでの過程をデータとして収集できるようにすることを目指す。

(2) 連想ネットワークの連携と可視化

動詞連想概念辞書に加え、名詞連想概念辞書の連想情報を問題セットの作成時に利用する。問題セットは問題文と回答選択肢から構成されるため、2つの連想概念辞書における連想ネットワークを連携させることで、問題文を構成する文節内の単語を広げることが期待できる。また、回答選択肢についても、特に不正解語を用意する際に別の単語との連想関係を考慮して生成できる。また、学習者が問題を回答した後に表示される解答の他にも連想されやすい単語を呈示することで連想関係の可視化を図る。

(3) 連想語彙学習システムの評価

本システムについては、「学習支援の教材」と「学習の効果」に関して評価を行う。前者は、自動で生成した問題セットの問題文と回答選択肢に対して、日本語教師がレーティングし、適切かどうかを判断する。後者の「学習の効果」については、実際に日本語学習者に本システムを一定期間利用してもらい、なお、日本語学習者は「学習」と「復習」の両方のフェーズを実施してもらい、「既知/未知の動詞」に関して「学習」フェーズで回答を誤った問題について「復習」フェーズで正解するまでにどれくらいの回答数を要したか等について調べる。

4. 研究成果

(1) 連想語彙学習システムの開発

生成された問題セット(問題文と回答選択肢)が「学習」と「復習」のフェーズで当初の目的通りに呈示され、学習者の回答結果が各フェーズで収集できるようになった。問題文は、述語の動詞の他に、その動詞と連想関係のある名詞(動詞連想概念辞書の刺激語(動詞))に対する連想語)と格助詞を組み合わせた2文節から構成される。これらの2文節のうちの1つは回答を選択する該当部分となる。回答選択肢は4語あり、正解語1語(上位で連想されやすい語(連想語))と不正解語3語から成る。回答結果を収集することで、「既知/未知」の動詞に関する問題文に対し正解語あるいはどの不正解語が選択されたか、統計的な分析が可能となっている。そのため、今後は学習者のデータ数を増やしていくことで、機械学習によるモデル化が可能となり、個人個人に適した出題に繋がっていくことが考えられる。

(2) 連想ネットワークの連携と可視化

動詞連想概念辞書の連想語は、日本語教育語彙表における初級前半から中級前半までの単語の約7割を占めていることから、日本語能力試験N4・N3相当の日本語学習者が本システムによる学習に適している可能性が高い。つまり、生成される問題セットは初学者が対象になることを考慮し、不正解語の3語は問題文に対して明らかに間違いである必要がある。そのため、不正解語は問題文の述語とある程度の関係性があるが、構文として成り立たない(あるいは意味が通らない)ような語とした。そこで、動詞と名詞の連想ネットワークを連携させ、正解語とは異なる深層格における連想語を不正解語として抽出している。また、学習者の回答後に正解語の次に連想されやすい単語を幾つか表示し、動詞とその連想関係を可視化することで学習者が語彙をより深く理解できるようにした。

(3) 連想語彙学習システムの評価

学習支援の教材としての評価として、生成した語彙問題の問題文ならびに回答選択肢がそれぞれ学習コンテンツとしてどれくらい適しているか、専門家(国内・海外の大学機関で日本語教育に携わっている日本語教師)がレーティングを実施した。その結果、「問題文と正解語」と「不正解語の3語」はそれぞれ77%、84%の割合で、問題セット全体(問題文と回答選択肢)では70%の割合で適していることがわかった。全体的には曖昧性が生じない回答選択肢を約8割の精度で生成できており、連想ネットワークの連携の効果を確認した。その一方で、日本語として正しくない表現や特定の場面を表す問題文に対して曖昧性がある不正解語の呈示が一部見られていたため、これらの改善が今後の課題として挙げられる。

学習の効果については、日本語学習者26名に本システムを一定期間使用してもらい、学習結果を分析した。日本語能力試験N4・N3・N3-N2相当のクラスに所属している学習者は「学習」フェーズで問題セット全体の約2割を間違えていたが、N2相当以上のクラスに所属している学習者は明らかに誤りが少なく、問題セット全体の1割も満たなかった。また、「復習」フェーズを「学習」の翌日以降に学習者に取り組んでもらった結果、初回に間違えた問題に対し、ほとんどの学習者が1回目で正解していた。これは、「学習」で間違えた問題ごとに動詞と深層格に対し、正解語の他に連想されやすい語を幾つか表示していたことで学習者は語彙の理解を深めたことが考えられ、可視化の効果が示唆されている。

以上より、語彙学習において連想ネットワークの連携と可視化について一定の効果を確認することができたといえる。

5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 9件 / うち国際共著 0件 / うちオープンアクセス 5件）

1. 著者名 Yamasaki Shota, Teraoka Takehiro	4. 巻 13323
2. 論文標題 Focus Estimation Using Associative Information to Support Understanding of Art	5. 発行年 2022年
3. 雑誌名 Design, User Experience, and Usability: Design Thinking and Practice in Contemporary and Emerging Technologies	6. 最初と最後の頁 279 ~ 288
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-031-05906-3_21	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Takehiro Teraoka, Tetsuo Yamashita	4. 巻
2. 論文標題 Construction of Associative Vocabulary Learning System for Japanese Learners	5. 発行年 2020年
3. 雑誌名 Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation (PACLIC 34)	6. 最初と最後の頁 294 ~ 301
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 -

1. 著者名 ITO Kiyooki, TERAOKA Takehiro	4. 巻
2. 論文標題 Deep Case Estimation and Japanese Anaphora Resolution with a Verb-Associative Concept Dictionary	5. 発行年 2020年
3. 雑誌名 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)	6. 最初と最後の頁 79 ~ 84
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TAAI51410.2020.00023	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 -

1. 著者名 Funato Hirofumi, Teraoka Takehiro	4. 巻
2. 論文標題 Disambiguation of Japanese Idiomatic Expressions Using Bias in Occurrence of Verbs	5. 発行年 2020年
3. 雑誌名 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)	6. 最初と最後の頁 114 ~ 118
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TAAI51410.2020.00029	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 -

1. 著者名 Ishimoto Yuichi、Teraoka Takehiro、Enomoto Mika	4. 巻
2. 論文標題 An Investigation of Prosodic Features Related to Next Speaker Selection in Spontaneous Japanese Conversation	5. 発行年 2019年
3. 雑誌名 2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)	6. 最初と最後の頁 1~5
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/O-COCOSDA46868.2019.9041205	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Teraoka Takehiro	4. 巻
2. 論文標題 Analysis of Associative Information for Second Language Learning of Japanese	5. 発行年 2018年
3. 雑誌名 Proceedings of The 4th Asia Pacific Corpus Linguistics Conference (APCLC 2018)	6. 最初と最後の頁 434 ~ 439
掲載論文のDOI (デジタルオブジェクト識別子) なし	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18件 (うち招待講演 0件 / うち国際学会 5件)

1. 発表者名 塚田晃平, 寺岡丈博
2. 発表標題 状況と発話の対極関係に基づいた皮肉表現の検出
3. 学会等名 情報処理学会第85回全国大会
4. 発表年 2023年

1. 発表者名 千野愛実花, 寺岡丈博
2. 発表標題 非タスク指向型対話システムにおける関連性を考慮した質問生成
3. 学会等名 非タスク指向型対話システムにおける関連性を考慮した質問生成
4. 発表年 2023年

1. 発表者名 Chino Amika、Teraoka Takehiro
2. 発表標題 Relevance-aware Question Generation in Non-task-oriented Dialogue Systems
3. 学会等名 25th International Conference on Human-Computer Interaction (HCI International 2023) (国際学会)
4. 発表年 2023年

1. 発表者名 伊藤清晃, 寺岡丈博
2. 発表標題 日本語照応解析における深層格推定に基づいた先行詞の同定
3. 学会等名 言語処理学会第28回年次大会
4. 発表年 2022年

1. 発表者名 山崎翔太, 寺岡丈博
2. 発表標題 連想情報と時系列を考慮した文章の焦点推定による対話破綻検出
3. 学会等名 言語処理学会第28回年次大会
4. 発表年 2022年

1. 発表者名 山崎翔太、寺岡丈博
2. 発表標題 対話破綻の特徴に応じた回避手法の提案
3. 学会等名 情報処理学会第82回全国大会
4. 発表年 2020年

1. 発表者名 船藤裕文、寺岡丈博
2. 発表標題 動詞出現頻度の偏りを用いた慣用表現の曖昧性解消
3. 学会等名 情報処理学会第82回全国大会
4. 発表年 2020年

1. 発表者名 伊藤清晃、寺岡丈博
2. 発表標題 述語の連想情報を用いた日本語ゼロ照応解析
3. 学会等名 情報処理学会第82回全国大会
4. 発表年 2020年

1. 発表者名 石本祐一、寺岡丈博、榎本美香
2. 発表標題 三人会話の次話者選択に関わる言語・音響特徴の分析
3. 学会等名 日本音響学会2020年春季研究発表会
4. 発表年 2020年

1. 発表者名 Ishimoto Yuichi、Teraoka Takehiro、Enomoto Mika
2. 発表標題 A Prediction Model for End-of-Utterance Based on Prosodic Features and Phrase-Dependency in Spontaneous Japanese
3. 学会等名 APSIPA Annual Summit and Conference 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 長谷川凱, 寺岡丈博
2. 発表標題 言葉の連想に基づいたなぞなぞ文の自動生成
3. 学会等名 情報処理学会第81回全国大会
4. 発表年 2019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

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