

英単語並べ替え問題に解答する際に発生する 迷いの解答履歴データ分析

坂野 僚亮† 宮崎 佳典‡

静岡大学大学院総合科学技術研究科情報学専攻† 静岡大学大学院情報学領域‡

1. はじめに

英単語並べ替え問題は、与えられた英単語を並べ替えることで提示された日本語文に即した英文を作成する問題である。この問題の解答過程のデータを取得するため、コンピュータ上で解答できるシステムが開発された[1]。同システムではマウスのドラッグ&ドロップ (D&D) 操作によって単語を並べ替えることで、その軌跡データを取得できる。本研究ではこの軌跡データから、学習者の問題解答に際した「迷い」の有無を正しく判別できるか、という点に着目している。迷いの判定には、機械学習(教師あり学習)を用いている。

[1]では、学習者の迷いと英文における意味的なまとまりであるチャンクとの関連性についての分析が行われた。後述する「グループ化機能」や「単語退避レジスタ」の利用に際して、その内容とチャンクとの関係によって、解答データにおいて機械学習の特徴量と迷いとの違いが見られ、チャンクと迷いと関連性が示された。[2]では機械学習の特徴量の精査という観点で分析を行い、新たに特徴量を追加し予測精度の向上を実現した。F値は約83%であった。道本ら[3]は、学習者のチャンク、コロケーションの習得における学習方法と知識の関係、さらにその習得にあたっての動機付けについての検討をした。

本研究では、解答中に学習者が利用できる機能である単語退避レジスタについて、その利用の分析と、これに関連した新たな機械学習の特徴量の策定を行った。

2. システムの概要

学習者が解答に使用する問題解答インタフェースの概要図を図1に示す。問題解答画面には問題提示欄、解答欄、単語退避レジスタが用意されている。学習者は日本語を参考にして問題提示欄の単語をD&D操作によって移動して並べ替え、決定ボタンを押下することで問題に解答する。この単語の移動に際して、学習者は複数単語を矩形選択した後にまとめて移動させる「グループ化機能」を使用できる。また単語退避レジスタには、単語を一時的に退避しておくことができる。1問の解答終了後に、迷った単語、その1問の解答を通じた迷い度の選択を行う。迷い度と

は、最終解答に至るまでにその問題に対してどの程度迷ったかを示す指標であり、「ほとんど迷わなかった」、「少し迷った」、「かなり迷った」、加えて「誤って決定ボタンを押した」の4つからなる。「誤って決定ボタンを押した」は、解答データからノイズとして除去する。また特定の箇所ではなく問題全体に対して迷った場合、「全体的に迷った」という項目を選択することができる。取得した軌跡データを利用して、機械学習を行う。



図1 インタフェース概要図

機械学習では軌跡データを迷いあり(選択迷い度が「かなり迷った」)、迷いなし(選択迷い度が「ほとんど迷わなかった」)の2つのラベルで分類し、ラベルと特徴量の組であるデータを用いて分類器を構築する。分類のアルゴリズムには、ランダムフォレストを用いた。実装にはPythonのライブラリであるscikit-learnを使用している。

3. データ取得実験概要

軌跡データを収集した実験では、学習者にシステムを用いて英単語並べ替え問題を解答していただいた。本稿では某A大学の英語系科目を履修している学生22名を対象とした本問30問からなる660件のデータ、B大学の48名の学生を対象とした本問110問からなる5280件のデータを扱う。問題は大学の英語教員の監修によって作成された。またA大学ではチャンク情報についても同様に作成された。機械学習に使用している特徴量は以下の通りである：

- ・ 解答時間
- ・ 総移動距離
- ・ 第一ドラッグ前時間
- ・ 第一ドラッグ後時間
- ・ 平均移動速度
- ・ 最高移動速度
- ・ 最大静止時間
- ・ D&D回数
- ・ 最大D&D時間
- ・ 最大D&D間時間
- ・ 合計D&D間時間
- ・ Uターン回数 (X軸)
- ・ Uターン回数 (Y軸)

第一ドラッグ前時間、第一ドラッグ後時間は、学習者が問題解答を開始してから最初にドラッグ操作をするまで

Data Analysis of Learners' Study Logs to Ascertain Hesitation in Solving Word-Reordering Problems

† Ryosuke Banno, Department of Informatics, Graduate School of Integrated Science and Technology, Shizuoka University

‡ Yoshinori Miyazaki, College of Informatics, Shizuoka University

の時間、及びその後の時間である。D&D 時間は、D&D 操作にかかった時間である。D&D 間時間は、一度 D&D 操作が完了してから次に D&D 操作が行われるまでの間の時間である。Uターン回数は、X軸（水平）、Y軸（垂直）方向にそれぞれマウスの進行方向が逆転した回数である。

機械学習では迷いなし、迷いありのデータを少ない方のデータ数に合わせて同数ずつランダムに抽出し(A 大学では 230 件, B 大学では 1718 件ずつ)、分類する 10 交差検定を行った(ベースライン 50%)。[2]で得られた A 大学のデータでの予測精度の評価指標を表 1 に示す。全体の正解率は 83.3%であり、全ての評価指標においてベースラインを大きく上回り、バランスの良い結果を得られた。

表 1 [2]機械学習による分類結果(A 大)

ラベル	適合率	再現率	F 値
迷いなし	85.9%	81.3%	83.1%
迷いあり	82.5%	85.7%	83.7%

4. 単語退避レジスタの使用に関する検討

学習者が解答中に利用できる機能として、単語退避レジスタがある。本稿では解答欄以外への意図的な移動は、迷いに関連するものではないかと考え、新たにレジスタの使用について回数値を計算し、機械学習の特徴量として使用することを考えた。本稿では D&D 操作における移動元、移動先、また 3 つのレジスタのいずれを使用したかを問わず、レジスタを使用した移動回数全てを合計し、その値を機械学習の特徴量として追加した。また、解答中にレジスタを一度でも使用したかどうかという指標として Boolean 型の特徴量も採用した。これらの特徴量を追加した結果、B 大学のデータでの機械学習による分類結果に関し、精度の向上が見られた。結果を表 2 に示す。

表 2 レジスタ使用特徴量追加後の分類結果(B 大)

特徴量	ラベル	適合率	再現率	F 値
追加前	迷いなし	74.7%	75.8%	75.2%
	迷いあり	75.3%	74.0%	74.5%
追加後	迷いなし	75.5%	76.7%	75.8%
	迷いあり	76.1%	74.9%	75.3%

B 大学では特徴量の追加前後各 10 回の評価指標に対して行った T 検定では、全体の正解率、迷いなしの適合率、F 値、迷いありの再現率、F 値の 5 つにおいて 5%で統計的有意差が得られた。対して A 大学のデータでは変化は見られなかった。B 大学では A 大学に比して、全体の割合としての 1 問当たりのレジスタ使用が約 3 倍あったため、迷い予測への貢献が現れやすかったと推測される。

レジスタ使用に伴って行われると考えられるグループ化についても分析を行った。グループ化において作成された単語群と教員によって作成されたチャンクとの共通項がどの程度あるのか、また同じく教員により設定された

21 個の文法項目別にどの程度共通項が存在するのかを調べた。B 大学については問題作成者によってチャンクの設定がされていないため、本稿では A 大学のみを対象とした。共通項の有無については、チャンクをもとに英語の意味的關係を考慮して手動で判断した。共通項を多く持つ文法項目の上位の一部を表 3 に示す。

表 3 チャンク共通項数上位の文法項目 (A 大)

文法項目	共通回数	問題数	割合
助動詞	18	44	0.409
前置詞(句)	15	44	0.341
補語/二重目的語	5	22	0.227
否定	34	154	0.221

さらにここで示した 4 つの文法項目を含む問題の解答データのみを抽出して、レジスタ使用に関する特徴量の追加前後での機械学習を行った。結果を表 4 に示す。

表 4 表 3 内文法項目の問題での機械学習 (A 大)

特徴量	ラベル	適合率	再現率	F 値
追加前	迷いなし	76.8%	71.6%	72.7%
	迷いあり	74.3%	75.8%	73.8%
追加後	迷いなし	77.4%	70.7%	72.4%
	迷いあり	74.0%	76.9%	73.8%

結果としては、全体の精度が B 大学のものに近くなり、低下している。レジスタがよく使用される、チャンク作成が多く行われる文法項目については迷いの予測が難しくなり、その使用方法の分析を行うことでより詳細な学習者の迷い分析ができ、予測精度の向上ができると考えられる。

5. まとめ

本稿では英単語並べ替え問題への解答に際しての迷いの予測について、新たな機械学習の特徴量としてレジスタの使用回数を採用し、一部で機械学習の精度の向上が得られた。またレジスタの使用、チャンク作成において、文法項目との関連性について分析を行い、レジスタ使用のあるデータ及び文法項目について、全体のデータの場合との予測精度の違いが示唆された。今後は、レジスタ使用のある解答データに対してその使用方法、内容の文法項目ごとの違いの抽出に焦点を当てた問題によるデータ取得実験によって、より詳細な関連性の分析ができると考える。

参考文献

- [1] 米津康香, 宮崎佳典, ほか: 英単語並べ替え問題における解答中の動作履歴を用いた迷い検出, 外国語教育メディア学会中部支部研究大会, 2018.
- [2] 坂野僚亮, 宮崎佳典, ほか: 機械学習を用いた英単語並べ替え問題解答時の学習者の迷い予測, 電気・電子・情報関係学会東海支部連合大会, 2020.
- [3] 道本ゆう子, ほか: 「チャンク」習得における学習者へのインプットの影響, 太成学院大学紀要第 17 巻, 2015.