

データマイニングの手法を用いた 英語ライティングへのアプローチ

—日本人英語学習者のエッセイ評価に影響を与える
文法的誤りパターンの検討—

東京都／早稲田大学大学総合研究センター 助手 石井 雄隆

申請時：東京都／早稲田大学大学院在籍

概要

本研究の目的は、データマイニングの手法を用いた日本人英語学習者のエッセイ評価と文法的誤りパターンの関係性についての検討である。日本人英語学習者のエッセイ評価において、文法的誤りがどの程度関係しているかというものは、まだ十分に明らかにされていない。本研究では、その関係性について検討するため、2つの調査を行った。1つは、文法的誤りを20個のカテゴリーに分類し、エッセイ評価別の共起関係についてデータマイニング手法の1つであるアソシエーション分析を用いて、日本人英語教師がライティングを評価する際に寄与する文法的誤りについて検討した。もう1つは、文法的誤りの頻度情報からエッセイ評価の予測をするため、画像処理などによく用いられる手法である最近傍法を用いてエッセイ自動評価の可能性について検討した。

1

研究の背景と目的

1.1 データマイニングと教育研究

社会の高度情報化・情報発信の低コスト化に伴い、大量のデータが常に生成されている。また記憶媒体の大容量化・通信の高速化に伴い、膨大なデータの蓄積が進み、整理されていない膨大なデータが社会のあらゆる所に存在している。

そういった大量のデータの中から規則性や関連性など意味のあるパターンを自動的に抽出する手法をデータマイニング (Adriaans & Zantinge, 1998) と呼ぶ。従来、データ解析は多変量解析などの統計的手法が主流であったが、1990年代の中頃、人工知

能の分野におけるルールの自動生成の研究が行われたことがデータマイニングの発端であると言われている。

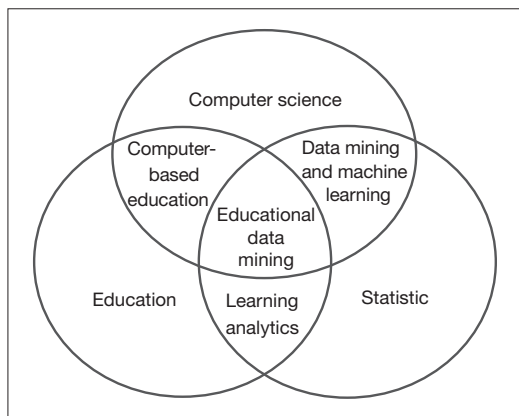
上記で述べた膨大なデータの蓄積は、高等教育の現状においても当てはまる。Romero and Ventura (2013, p.12) は、「教育機関が直面する1つの大きな課題は、急速な教育データの増加とその管理上の意思決定の質を改善するためのデータの使用である」ということを指摘している。そういった背景を踏まえ、エデュケーションナルデータマイニングや Learning Analytics and Knowledge (LAK) という分野が近年、注目を集めている。これは、「教育における問題を処理するために、教育環境から得られる特別なデータ集合に対してデータマイニング技術を適用する」(Romero & Ventura, 2013, p.12) のことを指す。

なぜ、教育データに対して、データマイニング技術を適用するかという点に関しては、教育データの特性が関係している。教育データには、下記の3つの特徴が含まれている。1点目は、異なる情報源(システム)から、学習者の活動の膨大な量のデータが提供されるということである。また2点目としては、すべての学生が活動や演習などを必ずしも完了していないため、不完全なデータが存在することが挙げられる。3点目は、データの粒度の異なる多様なデータが存在するという点である。例えば、2値データの性別や5段階のアンケート、100点満点のテストの得点などである。

なぜデータマイニングを教育研究に用いるのかという点については、3つの理由が挙げられる。1点目は、統計解析より、大量のデータが扱えるという

ことである。2点目は、実用性を重視し、データ解析の探索的側面を強調しているという点である。3点目は、欠損値などを含む不完全なデータに対し、頑健な分析ができるということである。統計解析にも欠損値を扱う方法は存在するが、データマイニング手法の方がより頑健であることが指摘されている。

図1に示すように、エデュケーションルデータマイニングは、学際的な新領域であり、教育学・統計学・コンピュータサイエンスの3領域にまたがっている。



▶ 図1：エデュケーションルデータマイニングと関連する主な領域 (Romero and Ventura (2013) に基づく)

エデュケーションルデータマイニングの研究事例を4つ紹介する。Romero, Romero, Luna and Ventura (2010) は、web ログデータにアソシエーション分析 (4.2節参照) を用いて、成績と Learning Management System (LMS) 上の活動との間の関係性と影響を評価した。また、Krüger, Merceron and Wolf (2010) は、練習問題の取り組みに関するログに同じくアソシエーション分析を用いて、学習者が学習資源をどのように使用したかを分析した。Peckham and McCalla (2012) では、ログデータに k-means 法を適用して、読解力のタスクにおける学習者の行動パターンの同定を検討した。He (2013) は、Live Video Streaming (LVS) system におけるオンラインクエスチョンやチャットメッセージに対してデータマイニングやテキストマイニングを行い、学習者の操作の状況を調査した。その結果、オンラインクエスチョンやチャットメッセージにおける学生の参加パターンについて類似性や相

違点が観察された。上記のように、さまざまな教育研究にデータマイニング技術は適用されており、外国語教育の分野においても今後こうしたアプローチは増加することが予想される。

1.2 英語ライティング研究と本調査の背景

日本における英語教育では、グローバル化が進行する国内外の情勢を受け、高度な英語力を持つ人材の育成が強調されている。それに伴い、国際社会で活躍できる人材育成のために、近年国内の大学において、ライティングセンターの設置などが進んでおり、ライティングとその指導に関する関心は高まってきている (吉田, Johnston, & Comwell, 2010)。

しかしながら、日本の英語教育において、英語ライティングは指導が十分に行われていないことが質問紙調査から報告されている (板津・保田・大井, 2013)。それには2つの大きな理由が原因として挙げられる。1つは、学習者のエッセイ評価は教師にとって大きな負担となるということ、もう1つは、ある評価者は文法的誤りに厳しいが、ある評価者は文法的誤りに寛容であるといった評定者間における評価の不一致の問題が存在するということである。

上記の問題を解決するための1つの提案として、これまでの外国語教育学、言語学、情報学の知見を統合したライティング自動評価システムの開発が考えられる。これまでに外国語教育学における英語ライティング自体の研究に関する知見だけではなく、情報学の分野で近年盛んに行われている英語学習者の誤りの自動検出 (水本・林部・小町・永田・松本, 2012) の研究、また外国語教育の分野で投野・望月 (2012) などで行われている自動エラータグ付与の研究などの知見を参考にしながら英語学習者の誤りの自動検出について検討を重ねた。これらの研究は、エッセイ自動評価の基礎的検討として位置づけられる。

一例として、石井 (2014a) では、制約に基づく文法理論を用いた日本人英語学習者の文法的誤りの自動検出に関する研究を行った。これは、語彙機能文法と呼ばれる文法理論と、それに基づく自然言語処理システムである Xerox Linguistic Environment を用いて、三人称単数現在形の s の脱落、日本語の干渉による誤り、冠詞の余剰および脱落、時制の誤りの自動検出の可能性について検討した。その結

果、日本人英語学習者が犯しやすい誤りは、規則に基づくアプローチからでも検出が可能であることが示唆された。

本研究では、エッセイ自動評価の基礎的検討として、日本人英語学習者の文法的誤りとエッセイ評価の関係性を検討するため、2つの調査を行う。1つ目は、文法的誤りを20個のカテゴリーに分類し、エッセイ評価別の共起関係についてアソシエーション分析を用いて検討し、日本人英語教師がエッセイを評価する際に寄与する文法的誤りについて検討する。2つ目は、文法的誤りの頻度情報を用いた英語学習者のライティング評価の予測について検討する。

本研究が、文法的誤りに着目するには3つの理由が存在する。1点目としては、エッセイ評価の変数としての寄与率の高さである。石井・近藤 (2014) は既存の自動採点システムをレビューする中で、e-rater を代表例として挙げている。e-rater は、Educational Testing Service (ETS) の Jill Burstein の研究グループが開発を始めたエッセイ自動採点システムである。e-rater の最初のバージョンでは、約60の作文の特徴量に基づいて、重回帰分析による評価を行っていた。その際に、すべての特徴量を用いるのではなく、トピックによって10個程度の変数を選んでいった。2004年に開発された新しいバージョンは、Burstein, Chodorow and Leacock (2004) によると、以下の12個の特徴量を用いて、作文を評価するように設計されていた。

1. 総語数に対する文法エラーの割合
2. 総語数に対する語の使用法についてのエラーの割合
3. 総語数に対する手順のエラーの割合
4. 総語数に対するスタイルについてのエラーの割合
5. 必要とされる談話要素の数
6. 談話要素における平均語数
7. 作文を6点法で採点する際に語彙の類似度が一番近い点数
8. 最高点を取った作文との語彙の類似度
9. Type-Token Ratio
10. 語彙の困難度
11. 平均単語長
12. 総語数

(石井・近藤, 2014, p. 4)

この12個の変数は、トピックによらず固定されており、e-rater は、これらを基に重回帰分析を行い、学習者の作文を自動で評価している。専門家とe-rater の評価の一致率はBurstein and Wolska (2003) では97%であり、高い精度で学習者のエッセイを評価することができる。またe-rater はCriterionというウェブベースの作文支援ツールに組み込まれている。これは、学習者の作文をその場で採点し、診断的なフィードバックを与えることができる。

上記のように、現在のe-rater は、評価に使用する変数のうち、3分の1は文法的誤りである。しかしながら、外国語として英語を学習する日本人英語学習者のライティングにおいては、文法的誤りの寄与率はそれ以上に高いのではないかと考えられることが、本研究において文法的誤りに注目する理由の1点目である。

2点目は、自然言語処理の分野で近年盛んに行われている英語学習者の文法的誤りの自動検出に関する外国語教育からの検討の必要である。例えば、英語学習者支援のための共通タスクとして、誤り検出・訂正ワークショップ (Error Detection and Correction Workshop) が2012年に行われた。これは、英語学習者の作文の誤り自動検出のワークショップで、共通のデータセットを用いて、前置詞誤り・動詞 (主語-動詞の一致) 誤りの2つのトラックに加え、誤りの種類を限定しないパイロットトラックが実施された。また国外でも、Helping Our Own (HOO) というワークショップが行われており、2011年はAssociation for Computational Linguistics (ACL) Anthology Reference Corpus を用いた論文の文法的誤り訂正のコンテストが、また2012年は、Cambridge Learner Corpus を用いて前置詞と限定詞の文法的誤り訂正を対象として行われた。

しかしながら、こうした一連のワークショップでは、学習者の誤りの検出に関する精度のみを問題としており、学習者の誤りの中で、こういったものがエッセイの評価に影響するかなどといった外国語教育研究者にとって関心のある問題については、検討がされていない。こういった状況を鑑み、外国語教育研究と自然言語処理の接点について検討するというのが、本研究のもう1つのねらいである。また、そういった知見を自然言語処理の分野にも還元することで、こういった文法的誤りを自動で検出するシ

STEMが望ましいかなどについての知見を提供することが可能となる。

本研究が文法的誤りに着目する3点目の理由は、ヨーロッパ言語共通参照枠（以下、CEFR）に代表される能力記述文における、より詳細な記述への理論的根拠である。CEFRとは、欧州評議会言語政策部門の専門家チームが30年以上の歳月をかけて制作した、行動思考アプローチと複言語・複文化主義を理念としたcan-do ディスクリプタである。CEFRの共通参照枠の構成は、縦軸と横軸から構成されている。縦軸は言語能力の記述であり、一番下のレベルが「基礎段階の言語使用者」、第2段階は「自立した言語使用者」、一番上のレベルは「言語に熟達している」レベルである。表1は、ヨーロッパ言語共通参照枠の文法的正確さについての能力記述文である。

■ 表1：ヨーロッパ言語共通参照枠における文法的正確さの能力記述文（吉島・大橋（訳・編），2004，p.124に基づく）

	文法的正確さ
C2	(例えば、これから言うことを考えているときや、他人の反応をモニターしているようなときといった) 他のことに注意を払っているときでも、複雑な言葉について常に高い文法駆動力を維持している。
C1	常に高い文法的正確さを維持する。誤りは少なく、見つけることは難しい。
B2	高い文法駆動力がある。時には「言い間違い」や、文構造での偶然起こした誤りや些細な不備が見られる場合があるが、その数は少なく、後で見直せば訂正できるものが多い。
	比較的高い文法駆動力が見られる。誤解につながるような間違いは犯さない。
B1	馴染みのある状況では、割合正確にコミュニケーションを行うことができる。多くの場合高いレベルでの駆動力があるが、母語の影響が明らかである。誤りも見られるが、本人が述べようとしていることは明らかにわかる。
	比較的予測可能な状況で、頻繁に使われる「繰り返し」やパターンのレパートリーを、割合正確に使うことができる。
A2	いくつかの単純な文法構造を正しく使うことができるが、依然として決まって犯す基本的な間違いがある—例えば、時制を混同したり、性・数・格などの一致を忘れたりする傾向がある。しかし、本人が何を言おうとしているのかはたいていの場合明らかである。
A1	学習済みのレパートリーについて、いくつかの限られた単純な文法構造や構文を使うことができる。

例えば、C1 レベルには、「誤りは少なく、見つけることは難しい」という記述があり、B1には「割合正確に使うことができる」というような記述が見られる。このように、CEFRの文法的正確さは、誤りの程度について言及しているのみである。本研究により、どの段階の学習者にはどのような誤りが多いかということがわかれば、その知見を基に、この枠組みを精緻化することが可能となる。

2 先行研究

本節では、調査に関係する先行研究について紹介する。石井（2014b）は、英語学習者が産出する作文におけるトピックの誤りの共起関係について検討した。トピックはエッセイにおける言語使用に影響を与える（Reid, 1990）という立場と、エッセイのトピックが異なっていたとしても、評価に影響を与えない（Spaan, 1993）という立場が存在する。そこで石井（2014b）は、Konan-JIEM Learner Corpus と呼ばれる日本人英語学習者が産出した英作文のエラータグ付きコーパスを用いて、作文における誤りのアソシエーション分析を行った。アソシエーション分析とは、百貨店や店舗で集めている購買履歴から顧客が購入するアイテムの組み合わせの規則を抽出するためなどに用いられているものであり、本研究では、1つのエッセイ中にどういった誤りが共起するかをトピックごとに検討した。その結果、トピックによって学習者が共起して産出しやすい誤り、またトピックの影響を受けにくい誤りが観察された。

アソシエーション分析は、投野（2013）が、学習者のCEFRレベルの推定に用いている。また小林（2012）が英語学習者の発話コーパスの分析に用いており、近年その英語教育への応用可能性が検討されつつある分析手法である。

また学習者コーパスに基づく文法的誤りを対象とした研究はこれまでもなされてきた。例えば、語彙（Milton & Freeman, 1996）、時制（Granger, 1999; Housen, 2002）、冠詞（Mason & Uzar, 2001）、コロケーション（Tono, 1996; Nesselhauf, 2005）、複数の誤り（Nicholls, 2003; Abe & Tono, 2005; Albert, Garnier, Rykner, & Saint-Dizier, 2009）などである。またKitamura（2011）は、文法的誤りとエッセイ

評価の関係性について決定木分析を用いて調査した。決定木分析とは、樹木モデルと呼ばれる非線形回帰分析、非線形判別分析の一つの方法を指し、説明変数の値を何らかの基準をもとに分岐させ、判別・予測のモデルを構築する分析手法である(金, 2007, p.229)。その結果、学習者の文法的誤りは、エッセイのトピックに関係があると結論づけている。Kitamura が対象とした誤りは、主語と動詞の一致 (*He have been living there since June.*)、動詞の形 (*I can't skiing well, but ...*)、不完全な文 (*Because people's interesting thing is not the same.*) の3点のみであった。しかしながら、学習者の誤りは多様であるので、本研究では文法的誤りのカテゴリーを20個にして、エッセイ評価別の文法的誤りのパターンについて検討する。

3 データ

本研究では、分析に Konan-JIEM Learner Corpus と呼ばれる日本人英語学習者が産出した英作文のエラータグ付きコーパスを用いる。このコーパスは、甲南大学と教育測定研究所が共同で収集し、日本人英語学習者17名が10個のトピックについて書いた170個のエッセイから構成される。トピックは、それぞれ University Life, Summer Vacation, Gardening, My Hobby, My Frightening Experience, Reading, My Home Town, Traveling, My Favorite Thing, Cooking である。また、文法的誤り情報と品詞・句情報が人手で付与されている。データは、授業内でエッセイのトピックを見た後、5分間エッセイについて考えるプランニングタイムを与えられ、35分で作文を書くという手順で収集された。研究目的のためであれば言語資源協会を通じて、購入が可能である。表2は、データの概要であり、表3は、エラータグの種類である。資料2に実際のテキスト中でどのようにエラータグが付与されたかをタグごとに記載している。NICT-JLE コーパスのエラータグガイドラインを参考に、書き言葉用に修正されたものを用いて2人のアノテーターによってエラータグが付与された (Nagata, Whittaker, & Sheinman, 2011)。ただしこのエラータグには、余剰、脱落などの誤りの詳細な情報は付与されていない。また、このコーパスは、文法

的誤りのみを対象としてタグ付けがされており、スペリングの誤りは対象とされていない。

■ 表2：データの概要

エッセイの数	170
エッセイを書いた学生の数	10
総文数	2409
総単語数	19285
異なり語数	2054

■ 表3：エラータグの種類

タグ	内容
n_num	名詞-単複エラー, 加算, 不可算エラー
n_lxc	名詞-語彙選択エラー
n_o	名詞-その他のエラー
pn	代名詞に関するエラー
v_agr	動詞-人称・数の不一致
v_tns	動詞-時制エラー
v_lxc	動詞-語彙選択エラー
v_o	動詞-その他のエラー
mo	助動詞に関するエラー
aj	形容詞に関するエラー
av	副詞(句)に関するエラー
prp	前置詞に関するエラー
at	冠詞に関するエラー
con	接続詞に関するエラー
rel	関係詞に関するエラー
itr	疑問詞に関するエラー
o_lxc	2語以上から成る成句での語彙選択ミス
ord	語順エラー
uk	特定不能なエラー, 構成上の致命的なミス
f	フラグメント(断片, 未完の文など)

4 調査 1

4.1 データの前処理

このコーパスには、エッセイの評価が付与されていないため、英語を1年以上教えた経験を持つ協力者に評価を行ってもらった。それぞれ10個のトピックのうち、2つを評価してもらい、1つのエッセイにつき2人の評価者に評価してもらった。また評価尺度には TOEFL CBT のライティング評価に用いられていた総合的評価の尺度を用い、そ

れを6段階から10段階に変更した。尺度を変更した理由としては、水本（2008）が指摘しているように、学習者の習熟度があまり高くない場合には、点数にばらつきが出にくくなり、分析の際に必要な情報である分散が小さくなると考えられたことが理由である。

これにより、本研究においては評定者のトレーニングを行うことができなかったため、一部のトピックで評定者間の信頼性を担保することができず、一定の信頼性を示した6つのトピックで調査を行った。下記が調査に用いたエッセイトピックとその信頼性係数である。

■表4：エッセイトピックごとの信頼性係数

トピック	信頼性係数
Summer Vacation	.75
Gardening	.66
My Hobby	.67
My Frightening Experience	.81
My Favorite Thing	.72
Cooking	.76

4.2 アソシエーション分析

本研究で採用する分析手法は、アソシエーション分析である。アソシエーション分析は、通常の共起分析で用いる対数尤度比や相互情報量といった指標と異なり、3つ以上の共起を扱うことが可能である。また単なる相関関係ではなく、因果関係の情報も得ることができるので、従来の手法では、観察することができなかった学習者のエッセイの特徴を測定することが可能である。頻出するアイテムの組み合わせの規則を漏れなく抽出し、その中から興味深い結果を探し出すことを主な目的とする。これを英語学習者のエッセイに適用し、1つのエッセイにどのような誤りが共起するかをエッセイ評価ごとに検討する。分析には統計解析環境であるRと分析用パッケージであるarulesを利用する。

アソシエーション分析の分析指標には確信度(confidence)、支持度(support)、リフト値(lift)が存在する。1つ目は、確信度であり、これは「条件が起きた場合に、結論が起きる割合」である。2つ目は、支持度である。これは、「条件と結論が同時に起こる場合が全トランザクションに占める割合」である。3つ目は、リフト値であり、これは「結論

が条件と無関係に起こっていないかどうか」を示す。その中で、本研究では支持度を採用する。量的指標である支持度を用いて、学習者の誤りの共起関係をエッセイ評価別に検討する。

また、分析に際し、アプリアリアルゴリズム(Agrawal & Srikant, 1994)を用いる。これは、評価指標の下限を設定し、それを下回る組み合わせに関する計算を省略するものである。本研究では、support値の下限を0.1に設定した。

4.3 調査1の結果と考察

上記6つのトピックをエッセイ評価の結果から上位・中位・下位の3グループに分け、アソシエーション分析を用いて分析を行った。結果は下記のとおりである。

調査1の結果について、習熟度別の支持度を表5に掲載した。結果から考えられるのは、初級学習者は、動詞の時制における誤りをしやすいということである。これは、裏を返せば、エッセイの評価者は、動詞の時制における誤りをしている学習者に対して敏感であるという側面も存在するということである。

同じ観点から考えると、上級学習者は、冠詞のエラーと名詞に関係したエラーが共起しやすいということが示唆された。次の2つの例文は、実際の学習者の産出例である。

I like to go to <at crr="">the</at> <n_num crr="movies">movie</n_num>.

I often go to <at crr="">the</at> <n_num crr="movies">movie</n_num>.

これらの共起パターンは、エッセイ評価にあまり影響を与えない可能性が示唆された。この理由として、冠詞を習得することは、英語教師を含む日本人英語学習者にとって習得が難しい項目であるからと考えられる。例えば、石田（2002）が高校生に質問紙調査を実施したところ、約80パーセントの学生が、冠詞を正確に使用することが難しいと感じているという結果となった。さらに石田は、日本人英語学習者にとって冠詞が難しい理由として2つ挙げている。1点目は、名詞や動詞のような内容語に加えて、冠詞や前置詞のような機能語は、抽象的な文法概念を表している。それゆえ、日本人英語学習者にとっ

■ 表 5：習熟度別の誤りの共起関係

上級

Rule	Support
{ at } => { n_num }	0.59
{ n_lxc } => { n_num }	0.56
{ pn } => { n_num }	0.52
{ n_lxc } => { at }	0.52
{ v_lxc } => { n_num }	0.48
{ at, n_lxc } => { n_num }	0.48
{ n_lxc, n_num } => { at }	0.48
{ at, n_num } => { n_lxc }	0.48
{ v_agr } => { n_num }	0.44
{ pn } => { prp }	0.44

中級

Rule	Support
{ } => { at }	0.88
{ } => { prp }	0.85
{ prp } => { at }	0.75
{ at } => { prp }	0.75
{ n_num } => { prp }	0.62
{ n_num } => { at }	0.62
{ v_lxc } => { at }	0.54
{ prp, n_num } => { at }	0.54
{ at, n_num } => { prp }	0.54
{ v_tns } => { at }	0.52

初級

Rule	Support
{ } => { prp }	0.91
{ } => { v_tns }	0.83
{ } => { at }	0.83
{ v_tns } => { prp }	0.78
{ prp } => { v_tns }	0.78
{ at } => { prp }	0.78
{ prp } => { at }	0.78
{ uk } => { at }	0.74
{ at } => { uk }	0.74
{ uk } => { prp }	0.74

て冠詞の本質的な特徴をとらえることは難しいということを挙げており、2点目としては、日本語には、冠詞に対応する語が存在しないことを理由として挙げています。

調査1では、アソシエーション分析を用いて、日本人英語学習者の文法的誤りのエッセイ評価別の共起パターンについて検討した。その結果、上級学習者および初級学習者に特徴的な文法的誤りが観察された。

5 調査 2

5.1 データの前処理

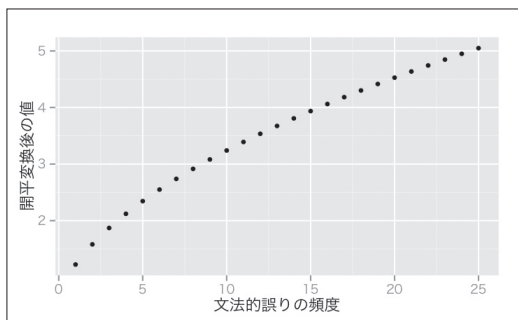
調査2では、文法的誤りの頻度情報を用いて英語学習者のライティング評価の予測を行った。調査1と同様に Konan-JIEM Learner Corpus を用いた。調査2では、調査1と同じデータに対して異なるエッセイ評価を付与した成田 (2013) の評価を用いた。成田によると、大学で英語教育に従事している教員2名 (英語母語話者) と筆者の3人で、以下のような手順で評価したと書かれている。

- (1) 各評価者は、1つのライティング・トピックに関して作成された17人のエッセイを一度に全て読み、質的に優れていると判断したエッセイのみをマーキングする。
- (2) 評価者3人による(1)の評価結果を照合し、トピックごとに優れているとみなすことができるエッセイを協議して選定する。
- (3) 10種類のライティング・トピックが課題として与えられていたため、(1)と(2)の評価作業をトピックごとに10回繰り返す。

(成田, 2013, p.7)

その後、文法的誤りの頻度をそのまま使用した場合では、精度の高い予測ができなかったため、頻度を開平変換した。開平変換は、非線形変換の一種であり、学習実験における消去抵抗やロールシャッハ得点などのあまり生起しない反応頻度などに適用される (森・吉田, 1990)。図2にあるように、開平変換を施すと小さい値同士の間隔は広くなり、大きい値同士の間隔は狭まる。スライドのグラフは横軸が変換前の頻度、縦軸が変換後の頻度となっている。変換前の小さい値同士の間隔は変換後広くなり、変換前の大きい値同士の間隔は変換後は小さくなる。例えば、頻度が5増えた場合、小さい値と大きい値で変換後の値の増え方が大きく異なる。これは、あ

る程度少ない数の間違いの変化には敏感だが、10、20と増えていくと誤りが1つ増えることに鈍感になるという人間の感覚に近いのではないかと考えられる。



▶ 図2：開平変換のイメージ

また石井 (2014b) の結果を踏まえると、トピックによって学習者が使用する文法項目 (時制など) に違いが生じるので、誤りの頻度を用いてトピックをクラスタリングした。その結果、図3のようなグループに分かれた。それゆえ、文法的誤りの頻度が類似していると考えられる、図3の左側に示した5つのトピック (Cooking, Gardening, Reading, My home town, Traveling) について書かれた英作文を分析の対象とした。

また、Konan-JIEM Learner Corpus には、20種類のエラータグが使用されているが、エッセイ評価の予測に使用できるであろう9種類のエラータグ (名詞-単複・加算, 不可算エラー, 動詞-時制エラー, 副詞 (句) に関するエラー, 名詞-その他のエラー, 動詞-語彙選択エラー, 前置詞に関するエラー, 代名詞に関するエラー, 形容詞に関するエ

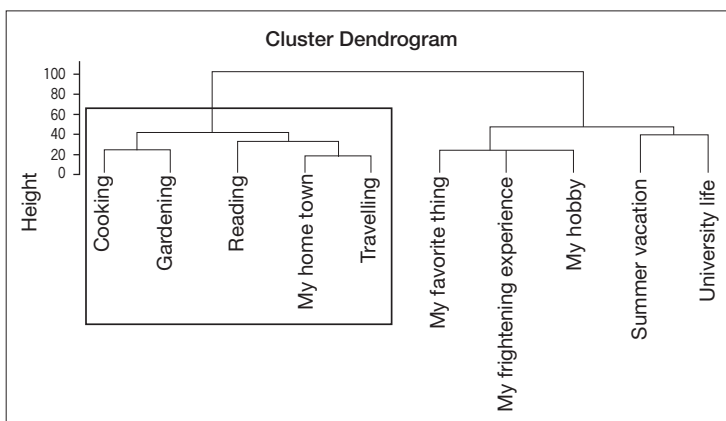
ラー, 語順エラー) を選択した。選択基準としては、エッセイ全体の誤りを見たときに、その頻度が極端に少ないもの (全エッセイ中数回の頻度)、成田 (2013) の Good, Poor で分けたとき、その両方で頻度の平均値 (変換後の値) にほとんど差がないものは使用しなかった。

5.2 最近傍法

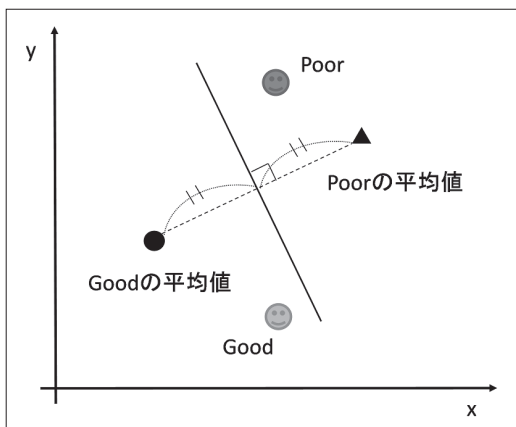
最近傍法を用いて、予測を行った。最近傍法は、画像処理や音声認識などによく用いられるパターン認識の手法である。Kondo (2010) では、アジア人英語学習者の音読自動評価システムの構築に最近傍法、K近傍法、重回帰分析の3つの手法の中から最近傍法が最も精度が高かったということを報告している。予測変数はエッセイの評価、説明変数は9種類のエラータグの頻度を開平変換したものをを用いた。

手順について、図4を基に説明する。実際は9種類のエラータグを使用した9次元になるが、9次元は図示できないので、ここでは2種類のエラーでエッセイ評価を予測する場合の例を示す。

初めに、データを学習用データ70と評価用データ20に分けた。評価用データを成田 (2013) に基づき Good と Poor に分け、それぞれのカテゴリーで対象となる2種類の誤りの平均値を算出し、プロットする。次に、2つの平均値を線で結び、この線の垂直二等分線を引く。この垂直二等分線のどちら側にプロットされるかで評価用データのエッセイ評価を決める。例えば、垂直二等分線より右側にプロットされれば、Poor と判定され、垂直二等分線より左側にプロットされれば、Good と判定される。



▶ 図3：トピックごとのクラスタ分析の結果



▶ 図 4：最近傍法のアルゴリズム

実際には、評価用データと Good および Poor の平均値との距離を測り、近い方を評価用データの評価と判定する。2変数の場合のこの距離の算出方法は、下記のとおりである。

$$D(p, X) = \sqrt{\frac{(x_i - \bar{x})^2}{sd_x} + \frac{(y_i - \bar{y})^2}{sd_y}}$$

誤りの頻度の平均値はそれぞれに異なるので、距離を標準偏差で割っている。

5.3 調査 2 の結果と考察

標準ユークリッド距離を利用した最近傍法を用いて、70の学習用データで9種類の文法的誤りの平均値を Good, Poor それぞれで計算した。評価データでは、各エッセイが文法的誤りの値と Good, Poor の平均値への距離を計算し、その距離が近い方をそのエッセイの評価とした。評価用データにおける成田 (2013) による評価と本手法で予測した評価の一致度は.57 (Cohen's Kappa) であった。

本研究では、文法的誤りに注目し、9種類のエラーによるエッセイ評価の予測を行った。その結果、文法的誤りの頻度のみを用いても、学習者のライティング評価は、高い精度で予測が可能であることが示唆された。

6 おわりに

本研究では、データマイニングの手法を用いて、日本人英語学習者のエッセイ評価に影響を与える文法的誤りパターンについて2つの調査を行った。調査1では、アソシエーション分析を用いて日本人英語学習者のエッセイ評価に影響を与える文法的誤りパターンおよびエッセイ評価に影響を与えない文法的誤りパターンについて調査した。その結果、習熟度ごとに共起しやすい誤りパターンが観察された。また調査2では、最近傍法を用いて、エラータグを用いた日本人英語学習者のエッセイ自動評価を行った。それにより、かなり高い精度で、文法的誤りの頻度情報からエッセイの自動評価が可能であることが示唆された。

本調査における今後の課題としては、学習者の対象文法項目の習得状況を観察するために、正用率と誤用率の割合なども検討したいと考えている。またタグガイドラインの見直しを検討する必要があると考える。しかし、タグの種類の詳細さとタグ付けの難しさはトレードオフであるので、どこまでを、そして何を誤りとするかを定めることは、先行研究などに基づき、判断する必要がある。

文法的誤りの共起情報およびライティング評価に寄与する文法的誤りの検討は、教室での作文指導などへの応用も可能であるので、今後もより詳細な検討がなされる必要がある分野である。

謝 辞

本研究を実施する機会を与えてくださった公益財団法人 日本英語検定協会の皆様、選考委員の先生方、助言を担当していただきました村木英治先生に心から感謝申し上げます。また本研究を実施するにあたって、専修大学附属高等学校英語科の先生方、また近藤悠介先生 (早稲田大学)、福田純也さん (名古屋大学大学院・日本学術振興会特別研究員) にはご助言、ご協力を賜りました。この場をお借りして、改めて感謝申し上げます。

参考文献 (*は引用文献)

- * Abe, M., & Tono, Y. (2005). Variations in L2 spoken and written English: investigating patterns of grammatical errors across proficiency levels. *Proceedings of Corpus Linguistics 2005 Conference*, pp.1-11.
- * Adriaans, P., & Zantinge, D. (1998). *Data Mining*. Harlow: Addison-Wesley.
- * Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of 20th International Conference on Very Large Data Bases*, pp.487-499.
- * Albert, C., Garnier, M., Rykner, A., & Saint-Dizier, P. (2009). Analyzing a corpus of documents written in English by native speakers of French: classifying and annotating lexical and grammatical errors. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.) *Proceedings of the Corpus Linguistics Conference (CL2009)*, University of Liverpool, UK.
- * Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. *AI Magazine*, 25 (3), pp.27-36.
- * Burstein, J., & Wolska, M. (2003). Toward evaluation of writing style: finding overly repetitious word use in student writing. *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, pp.35-42. Budapest, Hungary.
- Carson, J. (2001). Second language writing and second language acquisition. In T. Silva & P.K. Matsuda (Eds.), *On second language writing*, pp.191-199. Mahwah, NJ: Lawrence Erlbaum.
- Foster, J., & Andersen, O.E. (2009). GenERRate: Generating Errors for Use in Grammatical Error Detection. *Proceedings of the NAACL Workshop on Innovative Use of NLP for Building Educational Applications*, Boulder, Colorado.
- * Granger, S. (1999). Use of tenses by advanced EFL learners: evidence from an error-tagged computer corpus. In H. Hasselgard, & S. Oksefjell (Eds.), *Out of Corpora — Studies in Honour of Stig Johansson*, pp.191-202. Amsterdam: Rodopi.
- * He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 29 (1), pp.90-102.
- * Housen, A. (2002). A corpus-based study of the L2-acquisition of the English verb system. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.) *Computer learner corpora, second language acquisition and foreign language teaching*, pp.77-116. Amsterdam: John Benjamins.
- * 石田秀雄. (2002). 『わかりやすい英語冠詞講義』. 東京: 大修館書店.
- * 石井雄隆. (2014a). 「語彙機能文法を応用した日本人英語学習者の文法的誤りの自動検出」. 『2013年度 ICT 授業実践報告書 (大学英語教育学会 JACET-ICT 調査研究特別委員会)』. pp.7-19.
- * 石井雄隆. (2014b). 「日本人英語学習者の作文における誤りのアソシエーション分析」. 『外国語教育メディア学会中部支部外国語教育基礎研究部会2013年度報告論集』. pp.33-42.
- * 石井雄隆・近藤悠介. (2014). 「英語学習者を対象とした自動採点システム—課題と展望—」. 『外国語教育メディア学会 (LET) 関西支部メソドロジー研究部会2013年度報告論集』 4, pp.1-11.
- 石井雄隆・近藤悠介. (2015). 「文法的誤りの頻度情報を用いた英語学習者のライティング評価の予測」. 外国語教育メディア学会 (LET) 中部支部外国語教育基礎研究部会第2回年次例会・第1回卒論修論発表会発表資料.
- * 板津木綿子・保田幸子・大井恭子. (2013). 「学生アンケートによる日本の英語ライティング教育の実態調査—大学入学前・入学後を比較して—」. 全国英語教育学会. 北星学園大学.
- James, C. (1998). *Errors in language learning and use: Exploring error analysis*. London: Longman.
- 金明哲. (2007). 『Rによるデータサイエンス』. 東京: 森北出版.
- * Kitamura, M. (2011). Influence of Japanese EFL Learner Errors on Essay Evaluation. *Annual Review of English Language Education in Japan*, 22, pp.169-184.
- * 小林雄一郎. (2012). 「データマイニングによる誤り分析と辞書開発」. 『人文科学とコンピュータシンポジウム論文集—つながるデジタルアーカイブ』. pp.33-38.
- 小島ますみ・石井卓巳・金田拓・磐崎弘貞・赤松信彦・金澤洋子. (2014). 「ライティング評価とテキストの言語的流暢性・複雑性・正確性との関係: メタ分析による過去の研究成果の統合」. 大学英語教育学会英語語彙研究会第10回研究大会発表資料.
- 小町守. (2013). 「ウェブマイニングを用いた英語学習支援」. 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会発表資料.
- 小室俊明. (2001). 『英語ライティング論—書く能力と指導を科学する』. 東京: 河源社.
- * 甲南大学・教育測定研究所. (2011). 「The Konan-JIEM learner corpus error annotation guideline」. 言語資源協会.
- * Kondo, Y. (2010). *The development of automatic speech evaluation system for learners of English*. Unpublished doctoral dissertation, Waseda University, Tokyo, Japan.
- * Krüger, A., Merceron, A., & Wolf, B. (2010). A data model to ease analysis and mining of educational data. In M. Pechenizkiy et al. (Eds.), *Proceedings of*

- the 3rd International Conference on Educational Data Mining 2010*, pp.131-140. Eindhoven, The Netherlands: International EDM Society.
- * Mason, O., & Uzar, R. (2001). Locating the zero article: how TEFL can benefit from NLP. In G. Aston & L. Burnard (Eds.), *Corpora in the description and teaching of English*. pp.44-52.
- * Milton, J., & Freeman, R. (1996). Lexical variation in the writing of Chinese learners of English. C.E. Persy (Ed.), *Synchronic Corpus Linguistics*, pp. 121-131. Amsterdam: Rodopi.
- 宮田学. (2002). 『ここまで通じる日本人英語 新しいライティングのすすめ』. 東京: 大修館書店.
- * 水本篤. (2008). 「自由英作文における語彙の統計指標と評定者の総合的評価の関係」. 『統計数理研究所共同研究レポート215 学習者コーパスの解析に基づく客観的作文評価指標の検討』. pp.15-28.
- * 水本智也・林部祐太・小町守・永田昌明・松本裕治 (2012). 「大規模英語学習者コーパスを用いた英作文の文法誤り訂正の課題分析」. 『情報処理学会研究報告』. pp.1-8.
- * 森敏昭・吉田寿夫 (編著). (1990). 『心理学のためのデータ解析テクニカルブック』. 京都: 北大路書房.
- * Nagata, R. Whittaker, E., & Sheinman, V. (2011). Creating a manually error-tagged and shallow-parsed performance learner corpus. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp.1210-1219. Portland, OR.
- * 成田真澄. (2013). 「Konan-JIEM 学習者コーパスにおける名詞後置修飾要素の分析」. 『東京国際大学論叢—言語コミュニケーション学部編』第9号, pp.1-12.
- * Nesselhauf, N. (2005). *Collocations in Learner Corpus*. Amsterdam: John Benjamins.
- * Nicholls, D. (2003). The Cambridge Learner Corpus—error coding and analysis for lexicography and ELT. *Proceedings of Corpus Linguistics 2003*, pp.572-581.
- 沖原勝昭. (1985). 『英語のライティング』. 東京: 大修館書店.
- 大井恭子 (編著). (2008). 『パラグラフ・ライティング指導入門—中高での効果的なライティング指導のために』. 東京: 大修館書店.
- Ortega, L. (2009). Studying writing across EFL contexts: Looking back and moving forward. In R.M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research*, pp.232-255). Clevedon, UK: Multilingual Matters.
- * Peckham, T., & McCalla, G. (2012). Mining Student Behavior Patterns in Reading Comprehension Tasks. In *Proceedings of the 5th international conference on educational data mining*, pp.87-94.
- * Reid, J. (1990). Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom*, pp.191-210. New York: Cambridge University Press.
- * Romero, C., Romero, J.R., Luna, J.M., & Ventura, S. (2010). Mining Rare Association Rules from e-Learning Data. In *Proceeding of The Third International Conference of Education Data Mining*, pp.171-180. Pittsburgh, USA.
- * Romero, C., & Ventura, E. (2013). Data mining in education. *Data Mining and Knowledge Discovery*, 3(1), pp.12-27.
- * Spaan, M. (1993). The effect of prompt in essay examinations. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research*, pp.98-122. Alexandria.
- 高田康成 (編). (2004). 『21世紀に相応しい大学英語像の創出と実現へ向けた意識調査 (アンケート集計・中間報告)』. (文部科学省「英語教育に関する研究事業計画」大学英語教育に関する学際的戦略研究会) 東京大学大学院総合文化研究科英語研究室.
- * Tono, Y. (1996). Using learner corpora for lexicography. *LEXIKOS*, 6, pp.116-132.
- * 投野由紀夫. (2013). 「CEFR をどのように日本で適応すべきか——CEFR-J の具体的な試み」. 立教大学異文化コミュニケーション研究科言語科学専攻主催講演会.
- * 投野由紀夫・望月源. (2012). 「編集距離を用いた英文自動エラータグ付与ツールの開発と評価」. 『コーパスに基づく言語学教育研究報告』9, pp.71-92.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag.
- Yasuda, S. (2011). Genre-based tasks in foreign language writing: Developing writers' genre awareness, linguistic knowledge, and writing competence. *Journal of Second Language Writing*, 20, pp.111-133.
- * 吉田弘子, Johnston, S., & Comwell, S. (2010). 「大学ライティングセンターに関する考察—その役割と目的」. 『大阪経大論集』, 61, pp.99-109.
- * 吉島茂・大橋理枝 (訳・編). (2004). 『外国語教育Ⅱ 外国語の学習, 教授, 評価のためのヨーロッパ共通参照枠』. 東京: 朝日出版社.

資料

資料 1 : 学習者のエッセイの例 (University Life)

My university life is very interesting. Because I <v_lxc crr="do">act</v_lxc> many things <prp crr="">since</prp> now. First I <uk crr="am a member of">join</uk> <at crr="a"></at> circle. I feel <ord crr="very good about this"></ord><prp crr="about"></prp> this very good</ord>. <uk crr="I kill time by">My killing time is</uk> writing <n_num crr="novels">novel</n_num> or drawing <n_num crr="pictures">picture</n_num>. <uk>This has many people like me</uk>. So I concentrate <prp crr="on"></prp> this. Second is summer vacation. I <v_tns crr="did"><v_lxc crr="do">act</v_lxc></v_tns> many <n_num crr="things">thing</n_num> in <pn crr="my"></pn> summer vacation. My best memory is <at crr="the"></at> seminar on the sea. I went to Ho-chi-min and Singapore. I got many friends <prp crr="from">around</prp> Hyogo university. And I <av crr="sometimes">sometime</av> meet <pn crr="them">friends</pn>. Last I have many friends <prp crr="from">in</prp> high school, junior high school and <aj crr="other">etc</aj> <n_num crr="groups">group</n_num>. We always talk about each <n_o crr="other's">other</n_o> <n_num crr="lives">life</n_num> <prp crr="by">in</prp> e-mail or internet. And We play <prp crr="">in</prp> inside or outside home. We play funny <n_num crr="games">game</n_num>. For example, <n_lxc crr="one of us">a friend</n_lxc> <v_agr crr="calls">call</v_agr> <prp crr="">in</prp> Macdonald <con crr="and"></con> <v_lxc crr="says"></v_lxc> "Please give me <at crr="a"></at> hundred <n_num crr="hamburgers">hamburger</n_num>." And others look <prp crr="at"></prp> him and laugh. I have many friends, so my university life is very interesting.

資料 2 : エラータグの例 (甲南大学・教育測定研究所, 2011)

タグ	例文
n_num	This is the only one <n_num crr="thing"> things</n_num> you have to do.
n_lxc	She listened to his <n_lxc crr="speech">speech</n_lxc>.
n_o	I went to <n_o crr="Nihonbashi in Osaka">Osaka Nihonbashi</n_o>.
pn	I took Martin and a frien of<pn crr="his">him</pn> to the park.
v_agr	The number of students who work part-time after school <v_agr crr="has been increasing"> have been increasing</v_agr>.
v_tns	I'll make researvations for the ferry as soon as I <v_tns crr="find">will find</v_tns> out the schedule.
v_lxc	He wanted to <v_lxc crr="conceal">cancel</v_lxc> his guilt.
v_o	If it <v_o crr="is forgotten"><v_agr crr="forgets">forget</v_agr></v_o>, plants are going to die.
mo	"The phone is ringing." "I <mo crr="will">'m going to </mo>answer it."
aj	It was a <aj crr="genuine">genius</aj>diamond.
av	He worked <av crr="hard">hardly</av>today.
prp	He took full advantage<prp crr="of">with</prp>his position.
at	She is active in <at crr="the">a</at> development of low cost water pumps.
con	Clint hit a home run, <con crr="but">and </con> I didn't.
rel	I phoned all his friends, none of <rel crr="whom">who</rel> could tell me where he was.
itr	<itr crr="Which">What</itr> would you like to eat, Japanese or Chinese food?
o_lxc	He <o_lxc crr="made an attempt">had an attempt</o_lxc> at the conquest of the peak.
ord	When did you buy that <ord crr="large old brown wooden">old brown large wooden</ord>table?
uk	<uk crr="X">...</uk> In case of the UK tag, correction (crr="X") is not annotated in the tag.
f	<f>The last day,</f>