

日本語作文推敲支援システム「ナツメグ」における 学習者評価実験から見られる課題

八木 豊 (株式会社ピコラボ)
ホドシチェク・ボル (明治大学)
阿辺川 武 (国立情報学研究所)
仁科 喜久子 (東京工業大学)

Problems Found in a Learner Evaluation Experiment Using Japanese Composition Supporting System "Nutmeg"

Yutaka YAGI (Picolab Co., Ltd.)
Bor HODOŠČEK (Meiji University)
Takeshi ABEKAWA (National Institute of Informatics)
Kikuko NISHINA (Tokyo Institute of Technology)

要旨

我々は BCCWJ に科学技術論文を加えたコーパスを使用してレジスター誤り検出を行う日本語作文推敲支援システム「ナツメグ」を開発した。このシステムを使用した、日本語学習者によるレポート形式作文の推敲支援を評価する実験と分析結果について報告する。

実験は 4 つの課題ごとにテーマとプロンプトを示し、学習者が作文する際に一度だけレジスター誤り検出機能による指摘を受けて書き直せる形で実施し、36 名の学習者から 144 作文を収集した。

レポートに近い比較的硬い文章から学習した言語モデルによる評価では、いずれの課題においてもシステムによる指摘後にパーブレキシティが減少しており、全体としてレポートに適した文章に修正できたことが確認された。一方、個別のレジスター誤りのうち、特に副詞については、指摘を受けると削除してしまう学習者が多く見られた。これらの事例に対しては修正候補および例文の提示法による改善を検討、提案する。

1 はじめに

国立国語研究所による「現代日本語書き言葉均衡コーパス」(以後 BCCWJ)に加えて、近年、日本語学習者の作文に含まれている誤用に対してタグ付けを施した学習者作文コーパスが公開されてきた。それに伴い、日本語教育の分野では、それらのコーパスおよび自然言語処理の技術を利用して作文に含まれる誤用を自動的に検出し、学習者の作文を支援するための研究が行われている(今村(2012)、水本(2013))。

我々も、レジスターの誤り検出を中心に、日本語作文推敲支援システム「ナツメグ」の開発を進めてきた(八木(2014))。レジスターとは、「社会的な拘束力をもつ言語学上の規範」における言語使用域の変異のことであり、書き手と読み手がどのような関係で、どのようなコンテキストのもとで言語表現を使用するかによって、異なる語彙や文法項目で記述されることを示すものである。学習者作文においては、論文や授業で提出するレポートの中で話し言葉を使用しているなど、場にそぐわない表現がレジスターの誤りに該当する。

本稿では、日本語作文推敲支援システム「ナツメグ」におけるレジスター誤り検出機能について概説し、このシステムを使用した、日本語学習者によるレポート形式作文の推敲

支援を評価する実験と分析結果について報告する。

2 「ナツメグ」におけるレジスター誤り検出機能

ホドシチェク (2011) は、学習者が作文の目的とするレジスターを想定し、目的のレジスターに近いコーパスを準正用データ、目的のレジスターから遠いコーパスを準誤用データと設定したうえで、準正用・準誤用それぞれのデータに含まれる形態素および共起表現の頻度について統計処理を行い、準誤用データの頻度が有意に多い場合に、その表現は目的のレジスターの下ではふさわしくない、即ち、誤用であると判定する手法を提案した。

日本語作文推敲支援システム「ナツメグ」では、この手法を使用してレジスター誤り検出を行う。作文の目的とするレジスターには、アカデミック・ライティングの中でも論文・レポート、申請書などの硬い文書を想定し、科学技術論文、白書、法律を準正用データ、Yahoo!知恵袋、Yahoo!ブログ、国会会議録を準誤用データとして利用する。

レジスター誤り検出の精度を確認するため、日本語教師が学習者作文に対してレジスター誤りなど様々な誤用タグを付与した学習者作文コーパス「なたね」¹のデータの一部を利用して予備調査を実施した。その結果、日本語教師が指摘したレジスター誤り箇所の再現率が 78.0%、システムの検出結果全体の精度が 77.6%であった (八木 (2014))。

3 学習者評価実験

3.1 実験手順

本稿における評価実験は以下の手順で実施した。

- (1) 母語や日本語の学習時間など、背景調査のアンケートに回答
- (2) J-CAT (Japanese Computerized Adaptive Test) を受験
- (3) システムを利用して作文を入力
- (4) システムに対するアンケートに回答
- (5) 日本語教師によるコメントを送付

まず、学習者の言語的な背景および実験開始時点の日本語能力を確認するため、背景調査のアンケートへの回答と J-CAT の受験を必須とした。次に、システムを利用した作文入力では、課題ごとにテーマとプロンプトを示し、学習者がそれに関する作文を一通り書き終えてから、一回だけレジスター誤り検出機能による指摘を受けて作文を書き直せるようにした。一回に制限した理由は、レジスター誤り検出の精度が 100%でないシステムに対して、誤用の指摘がなくなるまで修正を繰り返し試行し、過度にシステムに依存した表現になってしまうことを避けるためである。学習者には以下に挙げる 4 つの課題を一つずつ順番に提示し、それぞれ 400 字以上、各課題に取り組む間隔は最低 3 日間となるよう設定した。

課題 1 : 日本人について理解できないこと

課題 2 : 原子力発電の可否

課題 3 : 日本のアニメやゲームソフトはなぜ人気があるか

課題 4 : インターネット社会の功罪

最後に、システムに対するアンケートを回収し、至らぬシステムのアフターケアとして、学習者が書いた全ての作文に対する日本語教師によるコメントを学習者に向けて送付した。

¹ 学習者作文コーパス「なたね」 <http://hinoki-project.org/natane>

表1 学習者の習熟度

習熟度 (J-CAT)	人数
母語話者相当	1
上級	7
上級前半	17
中級後半	7
中級	4
計	36

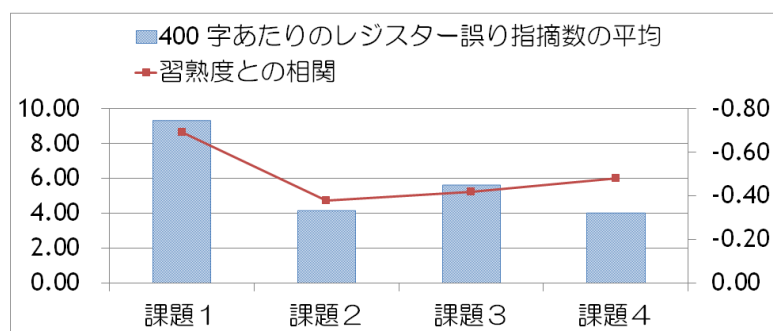


図1 400字あたりのレジスター誤り指摘数

3.2 実験結果

学習者評価実験は5つの調査地で実施し、36名の学習者から144作文を収集した。J-CATの得点に基づいた学習者の習熟度の分布を表1に示す。

図1は、学習者が書いた作文(400字あたり)に対するシステムのレジスター誤り指摘数の平均および、学習者の習熟度との相関を課題に示したものである。課題1と課題2を比較すると、システムによるレジスター誤り指摘数が大幅に減少していることがわかる。これは、課題1の作文時に、デスマス体による記述や「ちょっと」、「ちゃんと」のような副詞がレジスター誤りに該当するという指摘を受け、課題2の作文時には、そういった表現を使用する学習者が減少したことによるところが大きい。一方、テーマを少し柔らかいものに設定した課題3では、課題2と比較してレジスター誤り指摘数が増加しており、課題の内容によりレジスター誤り指摘数に増減があることがわかった。

また、課題1のレジスター誤り指摘数は学習者の習熟度とやや強い負の相関(-0.69)があり、最初の課題では、やはり習熟度の高い学習者のほうがレジスター誤りを犯しにくいことがわかる。ただし、課題2以降は、レジスター誤り指摘数と学習者の習熟度の相関は弱くなっており、中級以上の学習者は、システムが指摘することでレジスター誤りを学習し、自身で訂正できることを示している。

準正用データとして利用したコーパスから言語モデルを作成し、学習者の作文の訂正前後のパープレキシティの変化を確認したところ、微量ではあるが、いずれの課題においてもシステムによる指摘後にパープレキシティが減少しており、全体としてレポートに適した文章に修正できたことが確認された。パープレキシティが大きく変化しなかった要因の一つとして、今回の実験では、レジスター誤りの可能性を指摘するのみで、代替候補など訂正方針を示すことをしないため、訂正後の表現もレジスター誤りに該当してしまったことが挙げられる。

表2は、今回の実験においてシステムが指摘した全てのレジスター誤りについて、その指摘が妥当なものか否か日本語教師による判定を実施した結果である。全体の精度は83.05%と、予備調査よりも良い結果を得られた。品詞別では、動詞の精度が著しく悪いが、これは、「驚く」、「言える」、「教える」、「探す」のような一般的な動詞にも関わらず、準正用データに出現しないものが多く見られたことによるものである。

表3は、システムが指摘したレジスター誤りに対して学習者がどのような対応をしたか品詞ごとにまとめたものである。何の対応もせずそのまま残した場合を「未対応」、削除以外の何らかの変更をした場合を「変更」、該当する表現を削除してしまった場合を「削除」としてカウントした。助動詞の削除割合が高いのは、「です」や「ます」を削除してデスマ

表2 システムによる指摘の妥当性

品詞	適切	不適切	計	精度
助動詞	342	0	342	100.00%
副詞	158	29	187	84.49%
動詞	86	97	183	46.99%
ナ形容詞	88	11	99	88.89%
イ形容詞	83	12	95	87.37%
助詞	55	21	76	72.37%
名詞	42	4	46	91.30%
補助動詞	10	0	10	100.00%
連体詞	10	0	10	100.00%
感動詞	3	5	8	37.50%
計	877	179	1056	83.05%

表3 指摘箇所に対する学習者の対応

品詞	未対応	変更	削除	削除割合
助動詞	258	33	51	14.91%
副詞	107	51	29	15.51%
動詞	138	39	6	3.28%
ナ形容詞	69	23	7	7.07%
イ形容詞	62	31	2	2.11%
助詞	51	21	4	5.26%
名詞	33	11	2	4.35%
補助動詞	4	1	5	50.00%
連体詞	7	3	0	0.00%
感動詞	6	0	2	25.00%
計	735	213	108	10.23%

システムによる記述を修正するという正しい対応によるものである。一方、副詞の削除割合が高いのは、学習者の語彙が少なく、代替表現が思いつかないために削除してしまっているように見受けられた。また、1,056件の指摘箇所のうち735件が未対応のまま残されているが、このうちのおよそ8割は、前述の日本語教師による妥当性判定で妥当であると判断されたものである。こういった箇所については、学習者自身が代替表現を考えたり、調べたりして、訂正を試みてほしいところではあるが、実験後に回収したシステムに対するアンケートでは、「誤り指摘箇所に対してどのように訂正したらよいか分からない」、「訂正のヒントとなるような情報をもっと提示できないか」という意見が多く寄せられており、今後は、誤用コーパスに頻出する典型的な誤用の場合はその修正候補を、その他に検出した誤用に対しては訂正の参考となるような例文の提示を目指していきたい。

謝辞

本研究は、文部科学省科学研究費補助金基盤研究(C)「日本語作文支援システムで考慮すべき学習者属性情報と提示項目の分析研究」(研究代表者:阿辺川武、研究期間:2012年4月~2015年3月)による助成を得て実施しています。

文献

- 今村賢治、齋藤邦子、貞光九月、西川仁(2012)「小規模誤りデータからの日本語学習者作文の助詞誤り訂正」自然言語処理, vol.19, no.5, pp.381-400.
- 水本智也、小町守、永田昌明、松本裕治(2013)「日本語学習者の作文自動誤り訂正のための語学学習 SNS の添削ログからの知識獲得」人工知能学会論文誌, vol.28, no.5, pp.420-432.
- ホドシチェク・ボル、仁科喜久子(2011)「作文支援システムにおけるレジスターの扱い」世界日本語教育研究大会 異文化コミュニケーションのための日本語教育 2, pp.522-523.
- 八木豊、ホドシチェク・ボル、阿辺川武、仁科喜久子(2014)「日本語作文推敲支援システム「ナツメグ」における誤用検出手法の評価」第5回コーパス日本語学ワークショップ予稿集, pp.167-170.