

# 専門用語の言い換えに着目した解説文リライト支援 - WEB を介した大規模な読み比べ評価実験システム -

藤沢仁子<sup>†</sup> 神門 典子<sup>†‡</sup> 相原 健郎<sup>†‡</sup> 安達 淳<sup>‡</sup>

<sup>†</sup>総合研究大学院大学 情報学専攻 <sup>‡</sup>国立情報学研究所

E-mail: {satoko, kando, kenro.aihara, adachi}@nii.ac.jp

筆者らは、背景知識を持たない読み手の理解を促進するための言い換えに着目した、解説文リライト支援システムの研究を行っている。そこで、提案手法によってリライトされた解説文が、実際に読み手の理解支援に繋がっているのか、解説文の読み手は解説文中のどこの部分を難しいと感じているのか、インターネットを介して様々な年齢、職業の不特定多数の人が評価するための「WEB アンケートシステム」を開発した。

## 1. まえがき

文章の難易度やリーダビリティの研究が多数なされているが[1,2,3,4,5]、その評価実験では、工学的な精度の評価が中心であり、人の理解度にふみこんだ調査はほとんど実施されていない。

スタンドアロン型のシステムを用いて評価実験を行う場合は、実験参加者をアルバイトとして、実験会場に召集する必要があり、大学生未満の未成年者や、教師、学校関係者などの公務員をアルバイトとして参加を要請することは困難である。よって、実験参加者は、アルバイトとしてリクルートしやすい大学生などに限定されてしま。

そこで、本研究では、インターネットを介して様々な年齢、職業の不特定多数の人が、複数の文章の読み比べ評価実験に参加できる「WEB アンケートシステム」を開発した。

## 2. 先行研究

酒井[6]は、ヘルスリテラシーと可読性の研究において、一般市民にわかりやすい健康医学分野の日本語テキストの特徴を探るため、可読性や語彙の異なる3種類の「慢性化膿性中耳炎」を説明する文書を用意し、大学生34名を対象に、読みやすさと理解を問うテストを実施した。

医師の執筆した患者向けテキストと、可読性と内容理解の観点で改善したテキスト2種の作成し、改善テキストが実際に読みやすくなったか、内容の理解を助けているかを検証するテストを行った。

改善点は、①可読性に関わる構文的要素、②語彙的要素、③情報構造で、改善テキストAは①と②のみ、改善テキストBは①～③すべてについて改善を行った。また、②語彙的要素は、改善テキストAでは医学用語や難解と思われる漢字に読み仮名を

カッコ内に補足、改善テキストBでは病気の内容理解に必ずしも関係ない解剖学用語を削除し、医学用語は「耳漏」を「耳だれ」というように本文中で言い換えた。

実験の結果、改善テキストA、Bではオリジナルテキストより読みにくい点の指摘が減り、内容理解の正答率が高くなった。構文的要素の改善、語彙の言い換えや情報構造の改善が読みやすさや内容の理解につながっているといえる。しかし、改善テキストAでは読みの所要時間が余分にかかり、改善テキストBでも内容理解の正答率が劇的に改善したとはいえない。可読性を向上するとされている要素や語彙の言い換えは、必ずしも読みやすさには直結しないこと、内容理解を促進するものとしてでないものがあることがわかった。

## 3. 解説文リライト支援システム

筆者らは、博物館、美術館やデジタルミュージアムの専門的な内容を含む解説文を、子どもなど背景知識の違う解説文の読み手に分かりやすくリライトする際に、言い換え箇所と言い換え候補表現を提示することで、書き手を支援するシステムを提案している[7]。

そして、今までに、解説文をリライトする書き手による評価実験として、4名の日本史の専門知識をもつ実験参加者が、提案手法の支援がある場合とない場合、2通りのリライト作業を行い、支援の有用性を評価をした。そして、支援内容が、書き手の方針と必ずしも一致しなくても、判断基準や参考例として役に立つ、という結果を得た [8]。

さらに、提案するリライト支援で作成した解説文が読者の理解支援になっているかを調べるために、「WEB アンケートシステム」を開発した。

## 4. 解説文読み比べ評価システム

「WEB アンケートシステム」は、1) アンケートモジュールと、2) 管理モジュール、の2つのモジュールからなる。

### 4.1 アンケートモジュール

実験参加者が、「解説文の読み比べアンケート」を行うモジュールである。

「解説文の読み比べアンケート」では、ある遺物やオブジェクトについての、複数枚の画像と3通りの解説文が表示され、参加者は3通りの解説文のなかから「もっとも分かりやすい」と思ったものを選択する。「表示順」による影響を排除するため、問題の表示の順序と、各設問ごとの解説文の表示の順序は、参加者ごとにランダムになっている。一つのオブジェクトに関する複数の画像は、参加者が任意に選択、拡大縮小ができる。また、参加者は、解説文中でとくに分かりにくい文字列を選択するよう求められている。

システムの構成は以下の通りである。

- (1) アンケートサイトの説明
- (2) 読み比べタスク
- (3) 参加者についてのアンケート

#### 4.1.1 アンケートサイトの説明

アンケートサイトの説明は、「研究の説明」「アンケートサイトの説明」「アンケート開始」の3ページから成る。

アンケート調査の目的や、研究の概要、実験参加者が行う基本的な動作の説明を行う。

3ページ目の「アンケート開始」で、参加者は参加規約に同意したうえで、メールアドレスでログインして、アンケート調査に参加する。

#### 4.1.2 読み比べタスク

読み比べタスクでは、1件のオブジェクトにつき、複数の画像と3通りの解説文が表示され、参加者は実際に解説文の読み比べを行う。

解説文と画像は、管理モジュールのデータベースに保存されている。画像は、メインの大きな画像の下に、複数のサムネイル画像が表示され、選択するとメイン画像が入れ替わる。メイン画像は、拡大、縮小することができる。また、画像の著作権保護のために、右クリックの禁止、透かし文字などの処理をした。

上述のように、表示順による影響を排除するため、オブジェクトを表示する順番と、各解説文を表示する順番は参加者ごと、オブジェクトごとにシャッフルされる。

参加者は、画像を見ながら解説文を読み比べ、もっとも分かりやすいと思う解説文のボタンをクリックする。分かりやすさに差が見られない場合は「どれも変わらない」を選択する。また、解説文中のどこを難しいと感じているのかを調査する

ために、解説文中の文字列をハイライトする機能を備えている。参加者は、解説文中の難解箇所を任意に選択することができる。

回答が済んだら、「次に進む」ボタンをクリックして次のオブジェクトの読み比べに進む。各ページには、「これまでの回答を送信して終了する」ボタンがあり、参加者は随時、アンケートを終了することができる。

未回答でこれらのボタンをクリックすると、「分かりやすい解説文が選択されていません」「分かりにくい表現が選択されていません」という警告文が表示され、参加者は「選択しないで次に進む」と「次に進む」の、いずれかを選択する。「これまでの回答を送信して終了する」、または、最後のオブジェクトについての回答を終了し「次に進む」をクリックすると、参加者のプロフィールについてのアンケートのページに移動する。尚、ログイン時と最後の問題の送信時の時刻を採取し、各参加者が回答に要した時間を計測する。[図1 「読み比べタスク」スクリーンショット]

#### 4.1.3 参加者についてのアンケート

参加者についてのアンケートは、参加者自身についての情報を取得するためのページである。性別、年齢(20代、30代など)、職業の基本情報のほかに、理解度に影響のある背景知識との関連で、好きな分野・教科/大学時の専攻、博物館や美術館への来館頻度についても訊く。また、当アンケートサイトや、博物館などについて、自由な意見を訊くためのテキストボックスも用意した。

表 参加者についてのアンケートの質問

	質問	回答方法
1	性別	選択肢
2	年齢	選択肢
3	職業 大分類 / 小分類	選択肢 /自由記述
4	好きな分野・教科 / 大学時の専攻	自由記述
5	博物館、美術館などへの 来館頻度	選択肢
6	当サイトや博物館などへの 意見や感想を自由にお書き ください。	自由記述

### 4.2 管理モジュール

管理モジュールは、(1)評価タスクのデータの管理と、(2)回答の管理、の2つの部分から成る。

#### 4.2.1 評価タスク管理

読み比べの対象となる解説文と、それに付随する画像のデータを管理するための機能。管理者は、この画面を通じて、解説文と画像の入力と削除、

解説文の編集を行う。入力されたデータは SQL でデータベースに保存される。

#### 4.2.2 アンケート 回答管理

実験参加者の回答の集計を管理する機能。

アンケート結果一覧は、実験参加者ごとの ID、ログイン名と参加者についてのアンケートの結果、投稿日時を一覧で表示する。各参加者ごとに、「詳細」へのリンクがあり、クリックすると、参加者ごとの読み比べタスク、参加者についてのアンケート回答が表示される。

CSV 形式でダウンロードされ、ダウンロードした日付がファイル名に記される。

### 5. 公開前の小規模実験

#### 5.1 実験概要

「WEB アンケートシステム」の公開前の準備として、研究所に 10 名の実験参加者を招いて小規模な評価実験を行った[9]。

実験には、日本考古の分野のデータを用いた。全部で 5 件の文化財を選び、各文化財について 1 つ以上の画像と 3 通りの解説文を用意した。

一人につき 5 組の、(1)「一般向け」解説文、(2)「子ども向け(支援なし)」(サイエンスライターと内容の専門家の共同作業により作成)、(3)「子ども向け(提案手法)」の 3 通りの解説文を読み比べて、最も解りやすい解説文を選択した。また、解説文中の解りにくい部分にマークした。その後、アンケートとグループインタビューを行った。今回の実験のための (3)「子ども向け(提案手法)」作成作業には MSWord を用いた。画像と一般向けの解説文を提示し、その子ども向け解説文を記入欄にフリーで作成するよう求めた。

#### 5.2 実験結果

アンケートシステムの操作としては、ほぼ問題なく操作が行えた。

「もっとも分かりやすい解説文」の選択では、(1)6%、(2)42%、(3)48%となり、提案手法の支援を受けてリライトした解説文が、サイエンスライターが内容の専門家と共同でリライトした解説文に匹敵する評価を得た。

また、難解部分の選択では、用語が難しさの原因になっていることが明らかになり、アンケートとグループインタビューから、本研究が提案する専門用語の言い換え手法が、読み手のニーズに合っているという結果を得た。

### 6. まとめ

インターネットを介して不特定多数の人が、複数の文章の読み比べ評価実験に参加できる WEB アンケートシステム」を開発した。

現在、小規模実験における実験参加者の感想を基に修正をし、画像データの提供元である博物館と公開の調整を行っている。近日、約 1 ヶ月の期間で WEB 公開する予定である。

### 謝辞

本研究は独立行政法人国立博物館東京国立博物館との共同研究として行われた。

本研究は、科学研究費補助金 萌芽研究「利用者とおブジェクトの特性に応じた文化財コンテンツ解説文の言い換え支援」の支援を受けて実施した。また実施にあたり、文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」によって作成したデータを一部、使用した。

### 参考文献

1. 乾健太郎. コミュニケーション支援のための言い換え. 言語処理学会第 7 回年次大会併設ワークショップ, 2001.
2. 佐藤 理史, 土屋 雅稔, 村山 賢洋, 麻岡 正洋, 王 晴晴. 日本語文の規格化. 情報処理学会 研究報告 NL-153, pp.133-140 (2003 年 1 月)
3. 川村よしこ. 語彙チェッカーを用いた読解テキストの分析. 講座日本語教育. Vol.34, pp.1-22, 1999
4. 近藤陽介, 松吉俊, 佐藤理史. 教科書コーパスを用いた日本語テキストの難易度推定. 言語処理学会第 14 回年次大会発表論文集, pp.1113-1116, 2008.
5. 柴崎秀子. 日本語コーパスを応用した文章の難易測定の研究. 特定領域研究日本語コーパス平成 19 年度公開ワークショップ研究成果報告会予稿集. 文科省科学研究費特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築: 21 世紀の日本語研究の基盤整備」総括班. pp.125-130, 2008
6. 酒井由紀子. ヘルスリテラシー研究における日本語の可読性研究と内容理解. 三田図書館・情報学会研究大会発表論文集 2009 年度, 21~24, 2009
7. 藤沢 仁子, 神門 典子, 相原 健郎, 安達 淳: 専門用語のユーザに合わせた言い換え支援システムの構築—言い換えデータベースの提案—, 人文科学とコンピュータシンポジウム (じんもんこん :-), 2006), 京都 (2006)
8. 藤沢仁子, 神門 典子, 相原 健郎, 安達 淳: 言い換え箇所と言い換え候補の提示による解説文リライト支援の書き手の評価実験, 情報処理学会第 190 回自然言語処理研究会, 東京 (2009 年 3 月)
9. 藤沢 仁子, 神門 典子, 相原 健郎, 安達 淳: 専門用語のユーザに合わせた言い換え支援システムの構築—言い換えデータベースの提案—, 人文科学とコンピュータシンポジウム (じんもんこん :-), 2009), 滋賀 (2009)



文化財解説の読み比べアンケート

この研究は、東京国立博物館と国立情報学研究所の共同研究の一環として行っております。

線の引き方の例・ヘルプ

文化遺産の写真について解説が3つあります。一番分かりやすい解説のラジオボタン○を選択してください。  
また、解説文中の解りにくいと感じる表現に線を引いてください。

「一般向け」を提案手法で支援してさらにわかりやすく書き換えた解説文

「一般向け」をサイエンスライターが書きなおした子供向けの解説文

どれも変わらない○

次に進む  
もう止める

次に進む  
もう止める

解説文A ○

解説文B ○

解説文C ○

画像の拡大(+), 縮小(-) ボタン

メイン画像 (透かし文字入り)

拡大表示している部分を示すガイド

同じ遺物の他の画像 (選択するとメイン画像に入れ替わる)

一般向けの解説文

子供向けの解説文

椅子(いす)に座った巫女(みこ)の埴輪(はにわ)です。江戸時代特有の髪型(「島田髷(しまだまげ)」)に似ているとされる右前合わせの衣(ころも)を着ています。両肩には禪(たすき)を掛けて後差させており、腰には模様のある帯(おび)をしています。頸(くび)・手首・足首に玉(ぎょく)をつけ、耳には耳飾りをして盛装(せいそう)しています。椅子(いす)は古墳(こふん)に副葬(ふくそう)された石製品(せきせいひん)や椅子(いす)の埴輪(はにわ)に似ていることから、身分の高い方が腰掛けるものとされ、左腰には祭祀(さいぎ) (おまつり) に用いたと思われる五鈴鏡(ごりんぎょう) (鏡のふちに五つの鈴が付いている)と香袋(こうぶくろ)を下げていることから、神聖(しんせい)な巫女(みこ)の姿を表したものと考えられています。

椅子(いす)に座った巫女(みこ)の埴輪(はにわ)です。島田髷(しまだまげ)は女性特有の髪型で、意須比(おすひ)と呼ばれる右前合わせの衣(ころも)を着ています。両肩には後ろで交差させた禪(たすき)を掛けており、腰には模様のある帯(おび)をしています。頸(くび)・手首・足首に玉(ぎょく)をつけて盛装(せいそう)しています。椅子(いす)は古墳(こふん)に副葬(ふくそう)された石製品(せきせいひん)や椅子(いす)の埴輪(はにわ)に似ていることから、身分の高い方が腰掛けるものとされ、左腰には儀式(ぎしぎ) で使ったと思われる五鈴鏡(ごりんぎょう) (鏡のふちに五つの鈴が付いた鏡)と香袋(こうぶくろ)を下げていることから、神聖(しんせい)な巫女(みこ)をかたどったものと考えられています。

椅子(いす)に座った巫女(みこ)の埴輪(はにわ)です。島田髷(しまだまげ)と呼ばれる髪型に結び、意須比(おすひ)と呼ばれる右前合わせの衣(ころも)を着ています。両肩にはたすきをかけた後差させた禪(たすき)を掛けており、腰には帯(おび)をしめています。首飾り(くびかざり)と耳飾り(みみかざり)をつけ、手首(てのこ)・足首(あしのこ)にも玉(ぎょく)をつけて着飾っています。身分の高い人が腰掛ける椅子(いす)に座り、左の腰に儀式(ぎしぎ) で使ったと思われる五鈴鏡(ごりんぎょう) (鏡のふちに五つの鈴が付いた鏡)と香袋(こうぶくろ)を下げていることから、神聖(しんせい)な巫女(みこ)をかたどったものと考えられています。

copyright 2009 TNM-Nii joint research. All Rights Reserved.

図 1 WEB アンケートシステム「読み比べタスク」のスクリーンショット