

専門用語のユーザに合わせた言い換え支援システムの評価方法

藤沢仁子[†] 神門 典子^{†‡} 相原 健郎^{†‡} 安達 淳[‡]

[†]総合研究大学院大学 情報学専攻 [‡]国立情報学研究所

E-mail: [†]satoko@grad.nii.ac.jp, [‡]{kando,kenro.aihara,adachi}@nii.ac.jp

専門家によって書かれた解説文は、非専門家には理解が難しい専門用語を多く含む。我々は、文化遺産デジタルアーカイブのコンテンツに付随する解説文の言い換えを支援する、用語の言い換えに着目した言い換えデータベースを開発している。提案システムは、書き手が既存の解説文を易しい表現に書き換えるための支援を目的とし、カテゴリなど用語を定義する「用語テーブル」と言い換えテンプレートを集めた「言い換えパターンテーブル」の2つのテーブルから言い換え候補を生成し提示する。このシステムの評価について、出力結果が適切な言い換えであるかの判定だけではなく、書き手の支援となっているか、また、対象読者の読解支援に役立っているか、という3つの側面からの評価方法について考察する。

1. まえがき

近年、電子化された文書を利用者や利用形態に適した形に自動編集する技術の必要性が説かれており[1][2]、ある目的を満たす表現へ文章を変換する言い換え処理の研究が多数されてきている。

筆者らが所属する研究グループでは、教育機関において学習者が文化財コンテンツを利用して知識や多様な知見を発見することを支援するソフトウェアの開発を目指した CEAX プロジェクト[3]を推進している。ここで問題となるのは、既存の文化財の解説記述には専門用語が多数使われているため、記述内容を理解するには専門知識を必要とすることである。そこで、1つのオブジェクトの説明にも「受け手」に応じた多様性を考慮した記述の必要性に着目し、従来の専門家向けの記述に加え、小学生や非専門家にも理解し易い平易な表現など複数のメタデータや記述、関連するコンテンツを柔軟に管理できる仕組みを提案している[4]。そして、提案システムのプロトタイプを用いた西東京市立田無小学校6年生89名を対象とした研究授業を実施したところ、受け手にあわせて記述を作成することで理解を支援できることが示唆された。

しかし、現在、子ども向けの記述がされているものは大変限られており、全てを人手で作成するのは困難である。そこで本研究では、文化財に関する解説文を子ども向けに言い換えることを支援するシステムの構築を目指した、言い換えデータベースを提案する。

2. 専門用語の言い換への分析

既存研究のなかで、読解支援を目的とした言い換えの研究の例を見てみると、小学生を対象とした榎本ら[5]、福祉的利用を目的とした乾[6]、日本語学習者を対象とした佐藤ら[7]の研究などが上げられる。

言い換え研究の中での位置づけとして、本研究は、文化遺産分野の記述を子どもの読解支援のために易しく解りやすい文章に、漢字から文章全体にわたり、簡潔な表現の中で言い換えることを目指している。特徴としては、漢字かな変換だけでなく語彙・構文・文章全体を考慮した子どものため（教育目的）の言い換え研究である点、量的複雑さは難易度のひとつの要素にとらえ、説明的になることで文章が冗長になることを避け簡潔な文章で言い換えることを目指す点などが挙げられる。

筆者らは今までに、40組の大人向け・子ども向けの文化遺産分野の解説文の比較により表現の使い分けの分析をした結果、13個の言い換えパターンを見出した[8][表1]。

言語レベル	言い換えパターン
文章全体の意味	(P-1) 補完文の追加 (P-2) 詳細な記述の削除
談話的要素	(D-1) 指示語・代名詞の内容の補完
形態・構文	(S-1) 連体節の主節化
用語	専門用語を残した言い換え (L-1) カテゴリの補足 (L-2) 用途・目的の補足 (L-3) 形状の補足 (L-4) 位置の補足 易しい用語への言い換え (L-5) 一般名詞への言い換え (L-6) 身近なものへの言い換え (L-7) 一般名詞に位置/場所を補足 (L-8) 小学国語辞典の語釈文と同等表現への言い換え
漢字	(K-1) 読み仮名の付与

表1 検出された言い換えパターン

特に注目したいのは、上記(L-1) (L-2) (L-3)

(L-4)の専門用語を残した説明である。その時代や分野を語るうえで重要であり、知識として学習させたい専門用語は、完全に平易な別の表現に言い換えてしまわずに残す。そして、辞書へのリンク機能などによりその言葉を説明するのではなく、画像に付随する説明書きとしての読みやすさを考慮し、その専門用語が含まれるカテゴリや用途・目的などを文章のなかに補足することで理解を支援する点である。

今回はこの点に着目し、例えば「美豆良」を「美豆良とよばれる髪形」と自動的に言い換えるための言い換えデータベースの構築を提案する。

3. システム概要

図1は、提案する言い換え支援システムの概要である。乾[6]の指摘にもあるようにテキスト評価と言い換え生成の2つのモジュールがシステムの中核となる。

おおまかな仕組みとしては、言い換える必要とする文章を入力すると、ある評価基準に基づいてテキストを評価し、言い換える必要のある箇所を提示する。国語辞典や小学国語辞典などの辞典、教科書や指導要領、日本語能力試験の出題基準などを、評価基準として用いることができると考えられる。

そして、指摘された用語が既に言い換えデータベースに登録されている場合は、提示された生成された言い換え候補リストの中から適当なものを選択するか、リストの表現を参考に人手で書き起こす。言い換えた結果はデータベースにフィードバックされる。また、指摘された用語が未登録の

場合はリンクされている美術用語辞典を参考に人手または半自動で用語テーブルへの入力を行う。用語辞典としては、美術の専門用語を扱った美術用語辞典や日本史用語辞典、百科辞典などが使用可能である。

4. 言い換えデータベース

我々が対象とする文章の特徴としては、画像に付けられた説明書きであり、用語辞典のようにすべてを言葉によって説明するのではなく、画像から判断できる情報がある点、200字程度の短い文章という制約の中で簡潔に表現されている点が挙げられる。各解説文は独立しているため、同じ専門用語の説明が各解説文の中で繰り返し必要となる。また、教師がWWWなど外部からデータを追加する際に説明文を書換える場合には、表現が統一されることが有用であると考えられる。

そこで、言い換え文例をデータベース化し、ユーザに対して言い換え例を提示する為の言い換えデータベースの構築を提案する。類似のシステムとしては翻訳メモリが挙げられるが、翻訳メモリが対訳文を原文とともに保管し、入力文に対するマッチングと置き換えを行う[9]のに対し、提案システムでは用語と言い換えパターンとの2つのテーブルから言い換え例を生成する。

用語テーブルでは、<カテゴリ><用途/目的><位置/場所><形状>など各用語の属性を定義する。これらの属性は、先の人手による解説文の分析の中で見出されたものだが、AATやCDWAにはこれらを記述するためのファセットや項目が定義されており、ドキュメンテーションにおいても重要であるといえる。しかし、カテゴリなどその

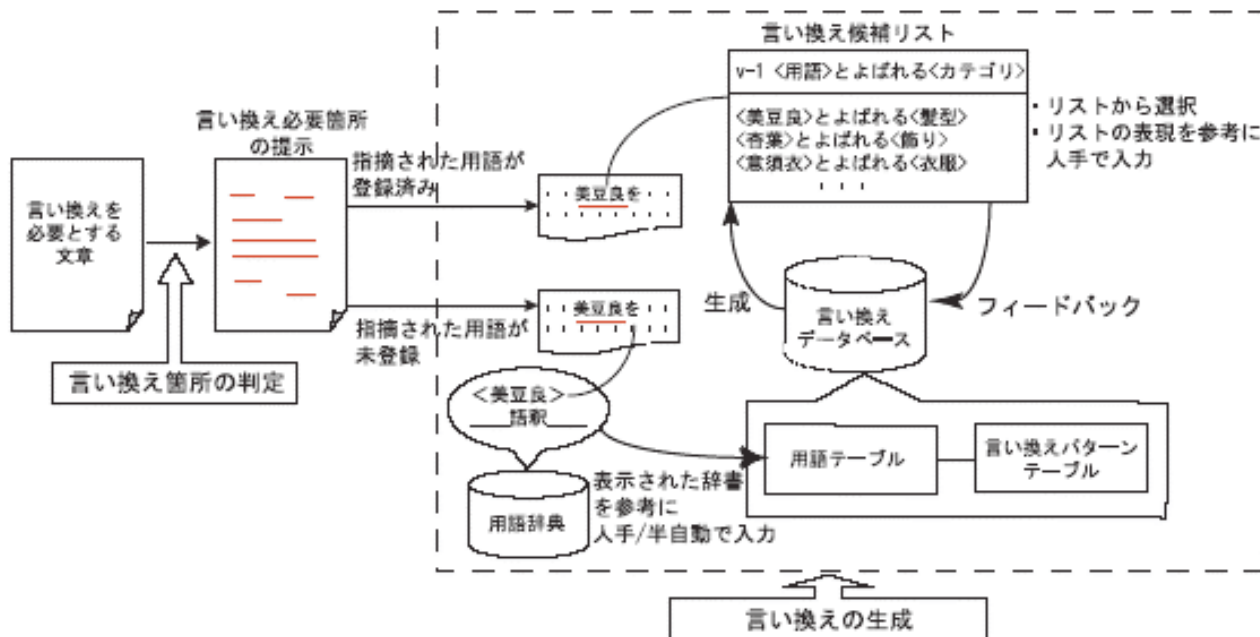


図1 システム概要図

用語が示すもの全てに当てはまる属性と、形状や位置のように一部の例でのみ有効な属性がある。それらをどのように管理していくかは今後の課題の一つである。一方、言い換えパターンテーブルは「<専門用語>とよばれる<カテゴリ>」「<用途/目的>ための<専門用語>」など、言い換えるのテンプレートを集めたテーブルである。

言い換えを必要とする用語に対し、まず用語テーブルを検索し属性を取り出す。その属性を含む言い換えパターンを取り出し、用語テーブルのデータをテンプレートに当てはめて言い換え例を提示する。また、ユーザが実際に言い換えた結果を記録しておき、そのデータを基に頻度情報を作成し、同じ文例を提示する場合のランキングに使用する。

用語	カテゴリ	用途/目的	位置	形状
歩揺	飾り	飾りの		スパンコールのような
短甲	甲	体を保護する	胸部分の	丈の短い
美豆良	髪型		頭部の	肩まで下がった
脚結	束帯	正装の	膝のあたりの	
...	

図3 用語テーブル

Ph_id	pattern
L-1	<用語>とよばれる<カテゴリ>
L-2	<用途/目的>ための<用語>
L-3	<形状><カテゴリ>
L-4	<位置><用語>

図4 言い換えパターンテーブル

<ul style="list-style-type: none"> ・<スパンコールのような><歩揺> (L-3) ・<体を保護する>ための<短甲> (L-2) ・<美豆良>とよばれる<髪形> (L-1) ・<膝のあたりの><脚結> (L-4)

図5 生成された言い換えの例

5. 評価方法

本システムの評価にあたっては、単に言い換え結果が適格であるだけでなく、より多面的な評価判定が必要と考える。そこで、出力結果が適格な言い換えであるか、書き手の支援となっているか、対象読者の読解支援に役立っているか、という3つの側面からの評価方法について考察する。

5.1 適格な言い換えであるか

まず、当然ながら言い換え結果が適格な日本語としての「絶対的な尤もらしさ」[10]（藤田，乾，松本 2004）を判定しなくてはならない。

筆者らは今までに、この2つのテーブルから機械的に生成された言い換えが、一文内で適切な表現になっているかを判定した。[11] しかし、「文脈にあっているか」「該当の解説文において真実であるか」など、文全体としての適格性を評価する必要がある。例えば、分析した解説文からは「美豆良」の形状として「肩まで下がった」という属性を検出したが、これは美豆良のなかでも「下げ美豆良」を表している「上げ美豆良」に関しては当てはまらない。また、「縦割りの」と「スカート状の」はともに「草擦」の形状を表しているが、それぞれ適する文脈は異なるだろう。

正解判定の手法としては、さしあたり人手で行っていくが、解説文間の類似度や言い換えるの規則性を利用して自動化もしくは半自動化する方法も検討したい。

5.2 書き手の言い換え支援になっているか

本システムは書き手の支援を目指しているのだから、実際に書き手の作業を軽減しているかを評価する必要がある。類似システムである翻訳メモリの評価方法によると、システムを利用した書き手へのアンケートが有効だといえる。アンケート項目の例として以下が考えられる[9][12]。

- ・提示される言い換え案の書き手からみた妥当性
- ・ユーザへのアンケート
- ・翻訳に要する時間は減ったか？
- ・翻訳の質は向上したか？
- ・手順は変化したか？
- ・自分のメモ帳や共有の用語リストに対訳を追加する頻度は変化したか？
- ・検索結果から必要な部分を抜き出すのに掛る時間は支障のない程度か？

他にも、言い換え支援の観点から独自の評価項目を検討し加える必要がある。

5.3 読者の読解支援になっているか

言い換えるの目的は読者の読解支援であるので、言い換えた結果が読者の理解支援になったかも評価しなくてはいけない。主にアンケートやインタビューなどで読者の意見を集める方法、操作ログを解析する方法、読者の観察などの方法が考えられる。

アンケートの項目としては、「元の文より分かりやすくなったか？」「専門家が自分で書き換え

たものとどちらがより分かりやすいか？」などを尋ねる他、解説文中の難解な部分を指摘してもらい、複数の解説文を読んだ後に理解度のチェックテストを行い結果を比較するのも有効である。

筆者らが所属するグループが、西東京市の小学校で行った発見学習支援ソフトの実証実験の際に、操作ログ収集、操作画面キャプチャ、児童の発話を記録する音声録音、実験授業風景のビデオ・写真撮影などの記録を行い、現在、解析を行っている。この時に用いたのは、人手で書き換えた小学生向けの解説記述であるが、言い換えシステムを使用して生成した解説文においてもこの手法を用いることができる。また、読者のアイ・トラッキングからも評価に有効なデータが得られると考えられる。

6. まとめ

本研究では、書き手が既存の解説文を易しい表現に書き換えるための支援を目的とした言い換えデータベースの評価方法として、適格な言い換えであるか、書き手の支援となっているか、対象読者の読解支援に役立っているか、という3つの側面から考察した。

本論文では、評価方法の考察にとどまっているが、各側面についてより詳細な評価項目を検討し、評価実験を行っていく予定である。

また、得られた評価結果からシステムのどの部分がどんなタスクに実際に役立ったのかを分析し、システムの改良に役立てたい。

謝辞

本研究の実施にあたり、独立行政法人国立博物館東京国立博物館が所蔵する文化財の画像およびメタデータの使用させていただいた。

本研究の一部は、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」の支援を受けて実施した。

参考文献

- [1] 乾健太郎, 藤田篤. 言い換え技術に関する研究動向. 自然言語処理, Vol. 11, No. 5, pp. 151-198, 2004.
- [2] 山本和英. 換言処理の現状と課題. [言語処理ワークショップ, 2001], pp. 93-96, 2001.
- [3] <http://ceax.ex.nii.ac.jp/ceax/>
- [4] 山田太造, 相原健郎, 藤沢仁子, 神門典子, 上原祐介, 馬場孝之, 長田茂美, 安達 淳, "学校教育における文化財コンテンツ利活用のための教育支援システム", 日本教育工学会 研究報告集 JSET06-3, pp. 23-30 (2006年5月)
- [5] 榎本聡, 室田真男, 清水康敬: "「音訓の読み方」と「ふりがな表記」に対応した漢字かな自動変換サーバの開発", 教育システム情報学会誌, Vol.17, No.3, pp.275-284 (2000年10月)

[6] 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.

[7] 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)

[8] 藤沢仁子, 相原,健郎, 神門,典子. 文化遺産に関する説明文の対象ユーザに合わせた言い換えの提案. 情報処理学会研究報告, 2006(82), pp.7-12, 2006

[9] 熊野正, 後藤功雄, 田中英輝, 浦谷則好, 江原暉将. 翻訳用例提示システムの設計・開発・運用. 電子情報通信学会論文誌. J84-D-II(6), pp.1175-1184, 2001.

[10] 藤田篤, 乾健太郎, 松本裕治. 自動生成された言い換え文における不適格な動詞格構造の検出. 情報処理学会論文誌, Vol. 45, No. 4, pp. 1176-1187, 2004

[11] 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, 京都, 2006

[12] 高橋 博之, 川崎立八, 牧田 光晴, 樋口 重人, 藤井 敦, 石川 徹也. 日英特許公報を用いた対訳辞書および翻訳メモリの構築. 情報処理学会研究報告, 2003-NL-155, pp.39-46,2003