

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

藤沢 仁子[†] 相原 健郎^{†‡} 神門 典子^{†‡}

[†]総合研究大学院大学 情報学専攻 〒240-0193 神奈川県三浦郡葉山町
[‡]国立情報学研究所 〒5 101-8430 東京都千代田区一ツ橋 2-1-2
E-mail: [†]satoko@grad.nii.ac.jp, [‡]{kenro.aihara, kando}@nii.ac.jp

あらまし 専門家によって書かれた文化遺産分野の解説文は、非専門家には理解が難しい専門用語を多く含む。本研究では、小学生の読解支援を目的とした文化遺産デジタルアーカイブのコンテンツに付随される解説文の言い換えの自動化の準備段階として、専門家によって書かれた一般向け、子ども向けのパラレルコーパスを手手で分析し、その結果見いだした13個の言い換えパターンについて報告する。また、文化遺産分野における言い換えの特徴について考察する。

キーワード 言い換え, 教育目的, 文化遺産

Proposal of Paraphrasing Text about Cultural Heritage for Different User Groups

Satoko FUJISAWA[†], Kenro AIHARA^{†‡} and Noriko KANDO^{†‡}

[†]The Graduate University for Advanced Studies Shonan Village, Hayama, Kanagawa 240-0193 Japan
[‡]National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan
E-mail: [†]satoko@grad.nii.ac.jp, [‡]{kenro.aihara, kando}@nii.ac.jp

Abstract The terminology and the description about cultural heritage are too technical and difficult for nonprofessional users of the domain especially for children and people from different cultures. As preparation to automatic paraphrase for reading assistance of children, and the paper reports 13 of paraphrasing patterns acquired by analyzing the parallel corpora of the cultural descriptions for adults and children.

Keyword Paraphrase, Educational purpose, Cultural Heritage.

1. はじめに

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], 文章のある目的を満たす表現への変換を目指した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、文化財のデジタル・アーカイブを教育目的で活用するためのシステムに関する研究および開発を目的とするCEAXプロジェクト[3]を推進している。プロジェクトでは、教育機関において学習者が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ているが、ここで問題となるのは、既存の文化財コンテンツ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]。その結果、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、「文が解りやすくて勉強になった」という意見が多数寄せられた。実際、操作ログを分析した結果、解説文などメタデータを参照する機能は、分類、注目、並び替えなどの機能のなかで最も多く使われており、全操作の3割近くの操作頻度であり、事後アンケートの「わからなかった言葉があるか?」との問いには66%の児童は「なかった」と回答した。

このように、受け手にあわせて記述を作成することは理解を支援するのに有意義であるが、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、文化遺産に関する解説文を子ども向けに言い換え手法を提案する。

2. 関連研究

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。

榎本ら[7]は、教育現場でのコンピュータやインターネットの活用が進むにつれて生じてくる、「漢字表記の問題」に焦点を当てている。インターネット上に提供されている情報を膨大な教育コンテンツと見做し有効活用するために、文書情報中の未習得漢字をかなに自動変換するサーバを開発した。音訓対応学年別漢字かな変換辞書を組み込むことで、学習者の学年に応じた漢字表記レベルでWWWページを表示できるようになった。筆者らは、小学校における実験で高い評価を得たと報告している。現在、漢字に読み仮名をつける機能は実用のレベルに達しており、小学生向けのポータルサイトであるYahoo!きっず[8]やキッズ goo[9]などで実際に利用されている。

乾ら[6]は聾者のコミュニケーション支援に向けたテキスト簡略化技術の研究を行っている。手話言語を母語とする先天性の聾者の中には、自然言語テキストの読解能力が不十分な人が少なくなく、これらの人の読解支援を目指した技術である。

主に、原文のどの部分に問題があり(テキスト評価)、それを解決するにはどのように言い換えれば良いか(言い換え生成)という2つの技術を柱とし、それらを切り離して設計することでより汎用な枠組みを提供しようとするものである。可読性の基準をモデル化するために聾学校の語学担当教諭を対象にアンケート調査を行い、可読性の差の見られる言い換えパターンを分析している。

佐藤ら[7]は薬の説明書などの重要な情報を伝達するためのテキストは、そのテキストに記述されている情報が読者に確実に伝わることを保証するために、読み易さに関するガイドラインまたは規格にしたがって記述されるべきであるとの観点から、日本語の平易度の規格を提案した。日本語能力試験「出題基準」を元に漢字、語彙、文法、量的複雑さの4部門に分け、3段階の平易度を定義している。この規格を実効的なものとするために、自動的に文の平易度を判断するシステムと指定された平易度を満たす文の書き換えを支援するシステムの実現を目指し、村山ら[10]による語彙と土屋ら[11]による文法の規格とその自動判定を元にしつつ、構文の規格化と言い換え支援の研究が進めら

れている。語彙に関して村山らは、規格判定の精度は95%であり、また、55%の用言に対して正しい言い換え候補を出力することができ、言い換え辞書の強化により改善が見込まれると報告している。一方、構文に関しては日本語能力試験に準拠した教科書に記載されている例文を対象として文法事項の判定をした結果、89%の適合率と51%の再現率を得ることができ、機械解釈可能な規則集合を改善することでより頑健なシステムを目指すと報告している。

3. 研究の位置づけ

ここで、言い換え技術を、(1) 文章ジャンル、(2) 目的、(3) 対象ユーザ、(4) 言い換える方向、(5) 言語の階層レベルの5つの軸を用いて分類することを考えてみる。それぞれの軸について説明を加えると次のようになる。

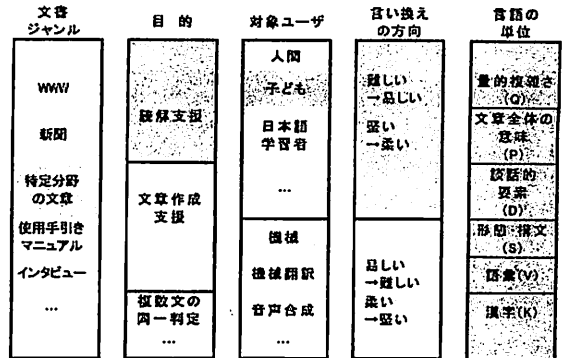


図1 言い換え研究のモデルと本研究の位置づけ

(1) 文書ジャンル

言い換えを必要とする文書の種類のこと。WWWや新聞をはじめ、説明書きやインタビューなど様々な記述や談話が対象コーパスとなり得る。

(2) 目的

何のために言い換えをするのかということ。大きく分けると人間のための言い換えの場合、読解支援と文章作成支援の2つがあるが、情報検索などのための同一文章の判定なども言い換えるの目的に含まれる。

(3) 対象ユーザ

誰のために言い換えるのかということ。子ども、外国人、障害者など読解支援を必要とする言語能力に応じたグループや、スタイルの統一された文書を書く、制限言語文書を書くなど文書作成支援を必要とするグループなど人間が対象

ユーザとなる言い換えが多く存在する。また、言い換えを必要とするのは人間だけではなく、機械翻訳や音声合成の前処理として機械のために言い換え処理を行う場合もある。

(4) 言い換への方向

より易しく解りやすい文章に言い換えるのか、もしくはより堅い専門的な文章に言い換えるのかという言い換への方向性。書き言葉特有語彙を話し言葉に言い換える場合[12]はより易しい解りやすい文に言い換えられる必要があるし、文章を要約する場合には複合名詞などに変換していくため、文章は難しく堅くなると言える。

(5) 言語の単位

言い換への対象とする言語の単位。漢字・仮名などの文字(K)をはじめ、語彙(V)や構文(S)、指示語の処理などの談話的要素(D)、そして、異なる文化背景の人に向けた補足説明など文章全体のレベル(P)まで、言い換え処理が言語のどの単位を対象としているか。また、佐藤ら[7]の指摘のとおり、文の長さや埋め込みの深さなど量的な複雑さ(Q)は文の平易度に深く関与している。これも、言語の単位に含める。

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おり、ここで示した分類に従ってその位置づけを図示してみると図1の網掛けの部分ようになる。

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ととらえ、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。専門用語などの説明に関しても、辞書へのリンクなどではなく、言い換えることにより、一連の文章の中で読みやすさを優先した理解支援を目指している。

4. パラレルコーパスの分析

本報告では、専門家によって書かれた大人向けと子ども向けのパラレルコーパスを手手で分析し、その結果見いだされた言い換えパターンについて報告を行う。

こども向けの解説記述にあつたては、歴史分野での子ども向け解説記事の経験のあるサイエンスライターに執筆を依頼し、考古学の専門家に適宜アドバイスを

受けた。ライターにより作成された文章は、読み手としては小学生6年生を想定しており、読みが難しそう漢字は習得しているかどうかに関わらず括弧で読み仮名を入れてある。また、小学校教諭の「長い文を生徒は読まない」との指摘から字数は概ね200字以内程度とした。現在、40件のオブジェクトに関するパラレルコーパスが作成されており、今後も追加されていく。

5. パラレルコーパスの分析結果

現在、作成済みの40件のパラレルコーパスの比較により表現の使い分けの分析結果を集めた。このコーパスの特徴としては、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。

5.1. 読み仮名の付与

従来の研究では、未習得漢字を仮名に置き換えたり、ふりがなを付与するツールが開発されている。しかし、文化遺産分野では漢字自体は小学校で学習する易しいものでも、組み合わせられてきた語彙は専門的で読み方が分からないものもある。

* a.: 大人向け c.: 子ども向け

(K-1) 読み仮名の付与

a.1 美豆良 → c.1 美豆良(みずら)

・漢字の学習学年:

美(3年生)、豆(3年生)、良(4年生)

・

a.2 刀子 → c.2 刀子(とうす)

・漢字の学習学年:

刀(2年生)、子(1年生)

5.2. 語彙の言い換え

語彙レベルでの言い換へのパターンで特徴的なものを列挙する。多くの専門用語を含む解説文において、辞書機能へのリンクではなく、読みやすさを考慮した簡潔な文章の中での語彙の言い換えは最も重要な役割を担う。

5.2.1. 専門用語を残した説明

専門用語のなかには、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言葉も存在する。それらの言葉を完全に平易な別の表現に言い換えてしまうのではなく、そこに説明を補完するような言い換え表現である。

(V-1) カテゴリの補足

「専門用語とよばれるカテゴリ」のように、上位概念やカテゴリを補完することで理解を支援する。

- a.1 挂甲（けいこう）に身を固め
- c.1 挂甲（けいこう）とよばれる甲（よろい）
に身を固め

- a.2 意須比を肩からかけ、
- c.2 「意須比（おすひ）」とよばれる衣をまとい、

(V-2) 用途・目的の補足

- a.1 草摺を革紐で緘してつけています。
- c.1 太ももを保護するための草摺（くさずり）
が、革ひもでとじつけられています。
- a.2 両腕には籠手（こて）をつけています
- c.2 腕を保護する籠手（こて）を両腕につけて
います。

(V-3) 形状の補足

- a.1 袴をはき、
- c.1 大きくふくらんだはかまをはき、
- a.2 鞍や障泥・輪籠をつけています。
- c.2 鞍（くら）が置かれ、どろよけや輪の形のあぶみをつけています。

5.2.2. 易しい語彙への言い換え

(V-4) 小学国語辞典の語釈文への言い換え

国語辞典の語釈文を言い換えに利用する研究は鍛冶ら[13]などによって為されているが、小学1年生～6年生までの全教科の教科書に基づいて作られている小学国語辞典は、子ども向けの語彙集として有効であると考えられる。そこで、小学国語辞典の語釈文を用いた言い換えを提案する。

- a.1 平時には華麗な装飾を施して
- c.1 普段ははなやかな飾りをつけて。
 - ・「平時：ふだんするとき」
 - 『学研 小学国語辞典 改訂第三版』
- a.2 高貴な人ばかりでなく
- c.2 身分の高い人だけでなく
 - ・「高貴：身分がたかく、とうといようです。」
 - 『学研 小学国語辞典 改訂第三版』

(V-5) 専門用語から一般名詞への言い換え

- a.1 脚結で膝のあたりを結んでいます。
- c.1 膝のあたりをひもで結んでいます。
- a.2 金銅製の飾金具がついた
- c.2 金銅製の飾りがついた

(V-6) 専門用語から身近なものへの言い換え

- a.1 ガラス製の小玉で、直径6ミリ程度です。
- c.1 ガラス製の小玉で、今でいうビーズです。
- a.2 庇は透かしが施されています。
- c.2 野球帽のようなひさしには、透かし模様

(V-7) 専門用語から一般名詞+位置・場所への言い換え

- a.1 胸繫には馬鈴が付けられている。
- c.1 首の下の革ひもには馬鈴がつけられています。
- a.2 六鈴付鏡板
- c.2 口もとの鏡板には6個の鈴がついています。

(V-8) 外来語の言い換え

- a.1 左右シンメトリーに近い装飾は安定感を感じさせます。
- c.1 左右対称に近い装飾は、安定感を感じさせます。
- a.2 屋根の部分だけをデフォルメした埴輪です。
- c.2 屋根だけを強調した埴輪です。

5.3. 形態・構文の言い換え

(S-1) 連体節の主節化

- a.1 古墳時代の一般的な家である竪穴式住居を母屋にして…
- c.1 古墳時代、ふつうの人々は竪穴式住居に住んでいました。この埴輪は、竪穴式住居を家の中心に…
- a.2 神人とお付きの人々を描いた2つの区画の間に二台の馬車が走る姿を描いた…
- c.2 神様とお付きの人々をえがいた2つの区画があります。その間には、2台の馬車をえがいた…

5.4. 談話的要素の言い換え

(D-1) 指示語、代名詞の内容の補完

指示語の指す内容がタイトルや画像から予測が付く場合も、具体的に説明を加える必要がある。

- a. その形から「蛇行状鉄器」と名付けられ
- c. へびのように曲がりくねっているところから、「蛇行状鉄器」とよばれている

5.5. 文章全体の意味レベルでの言い換え

(P-1) 補完文の追加

専門用語の説明のための、それが含まれるカテゴリの説明など、大人向け解説に明確には書かれていない事柄が、補完される場合がある。

a.1 三角板革綴式衝角付冑の埴輪です。

c.1 前面に船のへさきのような形の出っ張りがある冑(かぶと)を衝角付冑(しょうかくつきかぶと)といいます。この埴輪は、三角形の鉄板を革のひもでとじてつくった「三角板革綴式衝角付冑」とよばれるものです。

a.2 武具はおもに防御のための道具をいい、甲冑とその付属物と盾、矢を入れる筈鎌・鞞・鞞などがある。革や金属で作られている。

c.2 おもに身を守るための道具を「武具」といい、革や金属でつくられています。頭にかぶる冑(かぶと)や体につける甲(よろい)、盾のほか、矢を入れる道具である筈鎌(ころく)や鞞(ゆき)、矢を射るときに弓の弦が手首にあたるのを防ぐ鞞(とも)などがありました。攻撃するための道具は「武器」とよんで区別しています。

(P-2) 詳細な記述の削除

子ども向け解説のなかでは、補足説明がなされる一方で、デザインや文様の詳細や、それらが象徴する思想のモチーフのように、子どもにとって複雑で理解しにくい詳細な情報が省かれ簡略化される事象が見られる。

a.1 周縁に4個から10個の鈴をつけた仿製鏡(日本で作られた鏡)です。中部・関東地方の中期の古墳からの出土例が多く、鏡背文は乳文鏡(乳を主体とした文様構成の鏡)・獣形鏡(神仙思想に基づく霊獣を配した仿製鏡)・神獣鏡(神仙思想に基づく神仙像と霊獣を配した仿製鏡)が多くみられます。腰に五鈴鏡をさげた巫女の埴輪があり、祭祀に用いられたものと推測されます。

c.1 まわりに4-10個の鈴がついた鏡のことで、実際に鈴を鳴らすことができます。日本でつくられた鏡(仿製鏡(ぼうせいきょう))で、おもに関東地方や中部地方の古墳から発見されています。5個の鈴がついた五鈴鏡を腰からさげた巫女の埴輪があることから、神をまつる儀式で用いられていたと考えられます。

a.2 古墳に樹立されるために作られた土製品の総称です。円筒埴輪と形象埴輪に分けられ、円筒埴輪はいわゆる単純な円筒埴輪と朝顔形埴輪に、形象埴輪は家形埴輪・器財形埴輪・動物形埴輪・人物埴輪に分けられます。円筒埴輪の祖型は弥生時代後期の特殊器台で、朝顔形埴輪は特殊器台の上に壺形の土器を載せた形から始まったと考えられます。家形埴輪は弥生時代終末期の家形土器にさかのぼる可能性が指摘されています。武器・武具などの器財形埴輪は4世紀後半から5世紀にかけて流行し、動物埴輪は5世紀から6世紀にかけて、人物埴輪は6世紀に流行しました。

c.2 古墳の上や周囲に並べられた土製品のこと。筒型をした円筒埴輪と、家や動物・人・武器や生活の道具などをかたどった形象埴輪があります。円筒埴輪には、朝顔の花のように先が広がった朝顔型埴輪も含まれます。

6. 考察

このように、言い換えパターンを分析してみた結果、特に語彙の言い換えと文章全体の意味のレベルで文化遺産情報特有の特徴が見られた。まず、語彙では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、「美豆良(みずら)とよばれる髪形」のように、完全に平易な別の表現に言い換えてしまわずに残す。また、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。

一方、文章全体の意味レベルでは、理解するための背景となる内容の説明が補完されるかわり、複雑すぎる詳細な内容は省略される。つまり、一つのオブジェクトの説明書きには子ども向け・一般向けで共通して含まれる情報の他に子ども向けだけに含まれる補完説明と一般向けだけに含まれる詳細情報が存在する。

(図2)

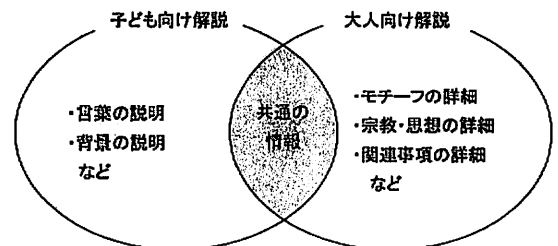


図2 文章全体のレベルでの関係

7. まとめ

本報告では人手で分析した言い換えパターンの報告にとどまったが、今後、この分析結果の妥当性の評価や自動化にむけてさらに言い換えパターンの分析を進める予定である。

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。

駒澤大学文学部講師の古庄浩明氏、サイエンスライターの財部恵子氏には解説の記述に関して多大な協力をいただいた。深く感謝する。

また、西東京市立田無小学校には、実験に際し多大な協力をいただいた。この場を借りて、感謝の意を表す。

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。

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, 清水康敏: "「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] <http://help.yahoo.co.jp/help/jp/kids/>
- [9] <http://kids.goo.ne.jp/info/kanji.html/>
- [10] 村山 賢洋, 麻岡 正洋, 土屋 雅稔, 佐藤 理史. 語彙の規格化とそれに基づく用言の言い換え支援. 言語処理学会第9回年次大会, pp85-88 (2003年3月)
- [11] 土屋 雅稔, 佐藤 理史. 文法の規格とその自動判定. 言語処理学会第9回年次大会, pp89-92 (2003年3月)
- [12] 鍛冶伸裕, 岡本雅史, 黒橋禎夫. WWWを用いた書き言葉特有語彙から話し言葉語彙への用言の言い換え. 自然言語処理, Vol. 11, No. 5, pp. 19-37, 2004.
- [13] 鍛冶伸裕, 河原大輔, 黒橋禎夫, 佐藤理史. 格フ